

Compensation Functions for Shifts of Finite Type and a Phase Transition in the
 p -Dini Functions

by

John Antonioli

B.Sc., Montana State University - Bozeman, 2006

M.Sc., Montana State University - Bozeman, 2008

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

DOCTOR OF PHILOSOPHY

in the Department of Mathematics and Statistics

© John Antonioli, 2013
University of Victoria

All rights reserved. This dissertation may not be reproduced in whole or in part, by
photocopying or other means, without the permission of the author.

Compensation Functions for Shifts of Finite Type and a Phase Transition in the
 p -Dini Functions

by

John Antonioli

B.Sc., Montana State University - Bozeman, 2006

M.Sc., Montana State University - Bozeman, 2008

Supervisory Committee

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

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

Dr. P. Kovtun, Outside Member
(Department of Physics and Astronomy)

Supervisory Committee

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

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

Dr. P. Kovtun, Outside Member
(Department of Physics and Astronomy)

ABSTRACT

We study compensation functions for an infinite-to-one factor code $\pi : X \rightarrow Y$ where X is a shift of finite type. The p -Dini condition is given as a way of measuring the smoothness of a continuous function, with 1-Dini corresponding to functions with summable variation. Two types of compensation functions are defined in terms of this condition. Given a fully-supported invariant measure ν on Y , we show that the relative equilibrium states of a 1-Dini function f over ν are themselves fully supported, and have positive relative entropy. We then show that there exists a compensation function which is p -Dini for all $p > 1$ which has relative equilibrium states supported on a finite-to-one subfactor.

Contents

Supervisory Committee	ii
Abstract	iii
Table of Contents	iv
List of Figures	vi
List of Figures	vi
Acknowledgements	vii
1 Introduction	1
1.1 Symbolic dynamics and thermodynamic formalism	1
1.2 Factor maps and compensation functions	3
1.3 Structure	6
2 Preliminaries	8
2.1 Subshifts	8
2.2 Information	10
2.3 Entropy	11
2.4 Some results from ergodic theory	16
2.5 Pressure	17
2.5.1 Pressure for Shifts of Finite Type	18
2.5.2 The Variational Principle	22
2.5.3 Pressure of a Function of Two Coordinates	29
3 Relative Pressure, Relative Equilibrium States and Compensation Functions	35
3.1 Relative Entropy and Relative Pressure	35

3.2	Maximal Relative Pressure	39
3.3	Compensation Functions	45
3.3.1	Existence of a Compensation Function	47
3.4	Walters functions and Compensation Functions	53
3.4.1	Constructing compensation functions in $\mathcal{W}(X)$	56
4	Structure of Compensation Functions and Relative Equilibrium States	66
4.1	Two Types of Compensation Functions	66
4.2	1-Dini functions and oracle-type compensation functions	69
4.2.1	Preparation for the proof of Theorem 4.2.1	70
4.2.2	Proof of Theorem 4.2.1 part 1	72
4.2.3	Proof of Theorem 4.2.1 part 2	76
4.3	A Walters-type compensation function	78
4.3.1	The Marcus-Petersen-Williams subshift and induced subsystems	79
4.3.2	Clothespinning sequences	84
4.3.3	Proof of Theorem 4.3.1	88
5	Outlook	92
6	Bibliography	94

List of Figures

Figure 3.1 The split golden mean shift	62
Figure 4.1 Between adjacent clothespins x is minimal, but if we pull out n_{i+1} we see the word w hangs lower than $x_{n_i}^{n_i+2}$	85

ACKNOWLEDGEMENTS

First I would like to express my sincerest thanks to my advisor Anthony Quas. Your patience and confidence were a great reassurance whenever my spirits flagged. I think you provide inspiration to all of your students, and I know that working closely with you over the last five years has given me the desire to be able to approach mathematical problems the way you do. It is my hope that we can remain both professional peers and friends for many years to come.

I would also like to thank Chris Bose for providing me with thoughtful questions, great professional advice and a few laughs.

I thank Pavel Kovtun for both reading and providing thoughtful comments on this thesis.

My gratitude also goes out to Karl Petersen for providing me with some very interesting context for, and potential extensions of this work.

I have made a lot of good friends in Victoria over the past few years, both in the UVic math department and elsewhere. To any and all of you who read this, please know that you are always welcome in my home, wherever that may be.

Medellee, if it wasn't for you I never would have believed in myself enough to accomplish this. You make me want to do everything I can to live life to the fullest and make you proud of me. I'm excited for wherever life takes us, so long as we get there together.

Mom and Dad, you never seemed to expect any less of me, which must be part of how I was able to do it. Kathleen and Clare, you guys are my inspiration! I love you all so much, and I am so grateful for your support.

Chapter 1

Introduction

1.1 Symbolic dynamics and thermodynamic formalism

At its core, this thesis is concerned with the thermodynamic formalism of symbolic dynamical systems. Symbolic dynamics originally grew out of a simple yet profound observation. Suppose we are observing a space which is being acted upon by some transformation. By discretizing the space, generally by considering some finite partition, we can encode the trajectory of a point under the action into the sequence of partition elements it visits over time. The space of all such sequences is called a *shift space* because the progress of time in the original system becomes a simple shifting of the encoded sequences.

In many cases a sufficient amount of information is encoded to allow properties proved for the shift to have significance for the original dynamical system. In this way the flows of differential equations, geodesics, the iteration of maps on topological spaces and the action of continuous groups can all be simplified and better understood through symbolic dynamics. In time shift spaces came to be appreciated as interesting objects for their own sake. They provide a wealth of examples of spaces which are simple to construct and analyze, but which can exhibit a broad range of complex behaviours. The advent of the computer also created a demand for a rigorous theory of communication channels, algorithms, memory storage, language and other essentially digital phenomena which can be modelled by symbolic dynamical systems.

Thermodynamic formalism is a relatively young branch of mathematics (although what isn't when compared to, for example, number theory) which can be considered a subfield of ergodic theory. Ergodic theory is the study of the statistical properties of a dynamical system which are invariant in time. In particular it deals with the asymptotic properties of points which are “average” with respect to some measure. Thermodynamic formalism grew chiefly out of the works of Sinai, Bowen, Ruelle and others to identify invariant measures which are particularly pertinent to a wide range of applications. The measures they identified are desirable because they maximize certain averages of the system.

Examples of averages which one would like to maximize in applications are the integral of a potential function on the system and the *entropy*, which is interpreted as the average amount of information contained in a partitioning of the space. Thermodynamic formalism is called so because it works by analogy with statistical thermodynamics in physics. In both fields, macroscopic invariants are obtained by averaging over the set of states in the system. In physics, these quantities include the temperature, pressure and entropy of a statistical ensemble. Accordingly, mathematicians name the invariants so obtained the topological entropy, topological pressure, etc.

Although the concepts of thermodynamic formalism can be defined and applied in many different topological settings, symbolic dynamics constitutes one of the largest and best understood areas of its application. Shift spaces tend to have the types of topological properties mathematicians dream of: they are compact, metrizable, have a basis of clopen sets and simultaneously seem discrete and continuous. The states of the system are usually a finite alphabet, or perhaps the words which can be written in that alphabet. This makes the statement of many theorems in thermodynamic formalism and the calculation of its principle quantities considerably easier. And, as is often the case with symbolic dynamics, we lose very little of the generality of the theory by restricting ourselves to this setting. Many of the thermodynamic properties which can be proved for shift spaces are readily transferred to other settings through encoding.

In this work, we will concentrate on a special class of shift spaces which are simplest from a combinatorial point of view. These spaces are the *shifts of finite type* (SFT's). They are also sometimes called topological Markov chains, because they are representable as bi-infinite sequences of vertices visited by a walk on a directed graph. Dynamical systems which are hyperbolic, in the sense that locally they have expand-

ing and contracting directions, can often be encoded to shifts of finite type. They are also one of the standard settings for information theory, and more recently have been studied for their connection to cellular automata. The combinatorial properties of SFT's can be exploited to prove very strong theorems about their topological and measure theoretic properties. For this reason, the amount of literature which deals exclusively with SFT's is large and well developed. This provides us with a lot of powerful tools to work with, but also presents the challenge of saying something new in a field about which much has already been said.

1.2 Factor maps and compensation functions

In any mathematical setting it is natural to want to understand maps between spaces which preserve the essential characteristics of the setting, the so called morphisms of those spaces. In symbolic dynamics, these morphisms are defined by maps which observe a fixed, finite size window of a sequence of states and re-encode it to a new symbol. By moving along a point and recording the new sequence of recoded symbols, these *sliding block codes* translate points in one shift space into points in another. If such a code is surjective it is called a *factor code*.

In ergodic theory, one nice interpretation of factor codes comes from information theory. Here, a factor $\pi : X \rightarrow Y$ is viewed as a communication channel with X as the space of inputs and Y the space of outputs. We observe some probability ν on the space of outputs, and would like to understand what can be said about the possible distributions μ on the inputs which could give rise to ν . The two distributions may differ because information may be lost in the transmission. For examples of papers where this point of view is taken see [11] and [13]. Here, one role filled by thermodynamic formalism is to identify those measures μ which have maximal entropy given that they produced the observed output distribution ν . The entropy of μ then gives the maximal transmission rate of the channel. Occasionally, in addition to SFTs we will also work with *sofic shifts*. A sofic shift is simply the image of an SFT under a block code.

So far we have identified two types of averages on a space which we can investigate through thermodynamic formalism: the entropy of a partition and the integral of a continuous potential. A related quantity is the *topological pressure* of a potential function. Consider a shift space X . Given some $f \in C(X)$, the pressure seeks to find

a measure which maximizes the sum of its entropy with the integral of f .

The sum of the entropy and the integral is sometimes called the *free energy* of f . In order to maximize the integral, the pressure wants us to concentrate measure on the points in X where f is largest. This is generally handled by putting the measure on or near the orbit of a periodic point. The entropy has quite the opposite effect on the pressure. Entropy is maximized by equally distributing the measure on all words of the same length in X . The net effect is that the pressure is maximized by a measure, called an *equilibrium state* of f , which tends to give equal measure to all words of the same length which give the same contribution to the integral. In this way, even if a word makes a significant contribution to the integral, its measure can be small if it occurs with very low frequency in words of greater length.

In the presence of a block code we would like to understand the thermodynamic measures which are maximal relative to some measure in the factor. This is the case with the transmission rate discussed earlier. The μ we seek does not maximize the entropy of X , it only does so among all measures which also push forward to ν . In a similar fashion we can seek measures which maximize the free energy of f among all measures which live over a common ν in the factor. These measures will be called *relative equilibrium states*. In some sense these measures answer the question: what is a natural way for the input space to behave under the influence of a potential, given that we observe ν in the output space? When our block code is finite-to-one, in the sense that the preimage of every point in the factor is finite, questions of this nature are more readily handled. If the block code is infinite-to-one, understanding the thermodynamics of the system becomes trickier.

Another question we can ask is: given a potential on the outputs, how is its equilibrium state related to its lift to the inputs? Information is only ever lost through a factor, never gained, so the entropy of the factor can only be less than or equal to that of its cover. In fact, if the factor is infinite-to-one, this loss of information is guaranteed. Similarly, the pressure of a function must be less than or equal to that of its lift. An amazing observation, first made in [3] and later refined in [22], is that the difference between these two pressures is entirely made up of topological entropy being generated in the fibres and because of this it can be accounted for by a single function. Such a function, for which a more precise definition will be given later, is called a *compensation function*. Most concrete examples of compensation functions which have been seen up to this point achieve this equality by cancelling out some

kind of relative entropy.

Let us call a collection (X, Y, π) with X and Y shift spaces and π a block code a *factor triple*. Compensation functions arise when studying the thermodynamic measures which have particular significance for both X and Y , such as the relative equilibrium states discussed previously. One general strategy for employing compensation functions is to prove that one with some structure exists for a factor triple, and that this structure enforces certain conditions on the relative equilibrium states which arise. Saturated compensation functions are those which can be written as $g \circ \pi$ for some $g \in C(Y)$. They have been studied in [19] and [23] for their relation to weighed entropy and Hausdorff measure. In [22] it was shown that factors with a compensation function in the Walters class of continuous functions have nice lifting properties for equilibrium states.

In the original survey of compensation functions by Walters [22], it is asserted that every subshift factor triple possesses at least a continuous compensation function. However, the type of function proved to exist does not subtract off relative entropy of any kind. Walters' compensation functions achieve their task by forcing relative equilibrium states to live on a subset of X where no relative entropy can be generated. This makes the pressure of a function and that of its lift equal in a more trivial way. In general no such subfactor need exist for a subshift factor triple, which has caused this statement to be disputed. In the case where X is at least sofic, the existence of a finite-to-one subfactor is guaranteed by [11].

A compensation function of this type must be sharply negative off the subfactor to scare equilibrium states away from trying to generate relative entropy. A natural question to ask is how sharp they must be. It is known that functions which satisfy certain smoothness conditions have equilibrium states which live everywhere on X [20]. It would be sensible to conjecture that a high degree of smoothness is sufficient to ensure the same for relative equilibrium states. In addition, some continuity classes which allow sharper functions have been shown contain functions with badly behaved equilibrium states. In this work, we will introduce the *p-Dini condition* as a candidate for recognizing both types of behaviour for relative equilibrium states.

This condition is not pulled out of the blue. The 1-Dini functions will be the familiar class of summable variation. In [6], Hofbauer used a function which is *p-Dini* for all $p > 1$ to prove that functions can have non-unique equilibrium states. A similar phenomenon was seen in [5]. There, as the modulus of continuity of a

piecewise expanding map of the interval is relaxed from one which is summable to one which is summable when raised to a power p which is strictly greater than 1 we see fundamentally different types of equilibrium states. In fact, we will observe this type of sharp boundary between the 1-Dini and p -Dini functions for $p > 1$. To continue borrowing from the terminology of thermodynamics, this shift in the types of relative equilibrium states observed constitute a sort of phase transition in the class of p -Dini functions. By explicitly constructing compensation functions on both sides of this boundary, we propose a sensible definition for two very different types of compensation functions.

1.3 Structure

The structure of this thesis is divided into three primary chapters, with the ultimate goal being the statement and proof of Theorems 4.2.1 and 4.3.1. These two results, which we will describe in a moment, constitute the primary original contributions of this work to the current theory. The first three chapters present a collection of relevant results about thermodynamic formalism in a factor setting and compensation functions. Most of the material from these sections appears elsewhere in the literature, although perhaps not together in one place.

Background material from symbolic dynamics, ergodic theory and thermodynamic formalism are reviewed in Chapter 2. For the most part, these topics are developed within the setting of shifts of finite type, to prepare for their application to the main theorems. This also greatly simplifies the elucidation of many facts which have much more general statements than the ones given here. In section 2.5.3 we give an explicit formula for the pressure of a function of two coordinates which motivates a number of later examples and results. For a more thorough account of the theory of symbolic dynamics, see [10]. Good general references for ergodic theory are [21] and [14]. A nice introduction to thermodynamic formalism, in particular equilibrium states, is [8].

In Chapter 3 we introduce the relative setting for thermodynamic formalism. This material is largely based on [22] and [9]. The concepts of relative pressure, relative equilibrium states and compensation functions are defined, and a number of their most important properties are discussed. In section 3.3.1 we first encounter compensation functions which are the difference of two information cocycles. This is the type of

compensation function which was first constructed in [3]. Theorem 3.4.3 gives an explicit construction of this type of compensation function for a broad class of factor triples.

One of the main purposes of the thesis is to explore the different types of compensation functions which exist for shifts of finite type. Chapter 4 comprises the primary investigation of this topic. In Definitions 4.1.1 and 4.1.2 we propose two different types of compensation functions whose relative equilibrium states have very different properties. The type constructed previously fall into the first definition. The p -Dini condition is proposed as a means of exploring both types of compensation functions, and more generally the structure of relative equilibrium states.

The first main result of the thesis, Theorem 4.2.1, proves that the relative equilibrium states of a 1-Dini function are fully supported, and have positive relative entropy. In preparation for the proof we recall several facts from the theory of joinings and the \bar{d} metric and give a number of useful lemmas from [24]. The proof of the theorem is split between 4.2.2 and 4.2.3.

The second main result, Theorem 4.3.1, explicitly constructs a compensation function which is p -Dini for all $p > 1$, and whose relative equilibrium states are not fully supported. Induced subsystems, a common topic in ergodic theory, and the construction of the finite-to-one subfactor discussed above are reviewed. In section 4.3.2, a finite extension based on this subfactor is constructed. This space of “clothespinned” sequences is a key feature of the proof of Theorem 4.3.1. Finally, the proof is given in section 4.3.3.

Chapter 2

Preliminaries

2.1 Subshifts

Let \mathcal{A} be a finite set of k elements, say $\{1, 2, \dots, k\}$, which we will think of as an alphabet. Let $\Omega = \mathcal{A}^{\mathbb{Z}}$. We can make Ω into a metric space by letting $k(x, y) = \min\{|i| \mid x_i \neq y_i\}$ and defining the metric $d(x, y) = 2^{-k(x, y)}$. A *word* w in \mathcal{A} is a finite sequence of symbols $x_0 x_1 \dots x_n$. By $[w]_i^j$ we denote the set of all sequences that have w in the i through j positions. In other words,

$$[w]_i^j = \{x \in \Omega \mid x_i x_{i+1} \dots x_j = w\}$$

If i and j are not specified, we assume that the word begins at the zeroth position, so $[w] = [w]_0^{|w|-1}$. The collection of $[w]_i^j$ for all words in the alphabet will be called the *cylinder sets* of Ω . They are both closed and open, and they form a basis for the topology on Ω . The set W_n is the collection of all words of length $|n|$ in Ω . The *state partition* of Ω is the partition formed by the collection of sets $\{[a]_0 \mid a \in \mathcal{A}\}$.

The *shift map* on Ω is defined by $(T(x))_i = x_{i+1}$ for all $i \in \mathbb{Z}$. The map T is continuous and bijective. Together, Ω and T form a dynamical system which we will call the *full k -shift*. A closed, nonempty, T -invariant subset of Ω is called a *subshift* of Ω .

A *shift of finite type (SFT)* is a particular type of subshift which has a finite set of forbidden words which define it. They define X in the sense that, if \mathcal{F} is the set

of forbidden words, then

$$X = \{x \in \Omega \mid x_i x_{i+1} \dots x_j \notin \mathcal{F} \forall i, j \in \mathbb{Z}\}$$

If k is the length of the longest forbidden word in \mathcal{F} then the collection \mathcal{F}' consisting of every word of length k which contains one of the forbidden words from \mathcal{F} defines the same SFT. In this way we can always assume all of the words of \mathcal{F} are of the same length. A *k-step shift of finite type* is an SFT where each of the forbidden words is of length $k + 1$. Unless noted otherwise we will generally be working in the context of shifts of finite type.

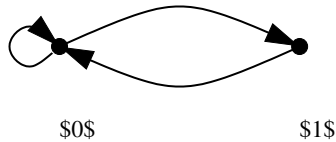
An *n-block code* between two subshifts (X, T) and (Y, S) is a continuous shift-commuting surjection π such that each word of length n in X maps to a single symbol in Y . In other words $(\pi x)_i = F(x_{i+a} \dots x_{i+a+n-1})$, a function of n -coordinates. If there exists an invertible block code between X and Y then we will say they are *conjugate*. It is an elementary result in symbolic dynamics that every n -block code from X to Y can be thought of as a 1-block code from a subshift \bar{X} to Y , where \bar{X} is conjugate to X .

A particularly nice type of SFT is a 1-step SFT. This is because we can create a matrix A , called the *transition matrix* of the SFT, which has the form $A(i, j) = 1$ if $ij \notin \mathcal{F}$ and $A(i, j) = 0$ if $ij \in \mathcal{F}$. This matrix completely describes the SFT, and we will frequently exploit its properties in proofs. Luckily, every k -step SFT is conjugate to a 1-step SFT by a recoding of words of length k to unique single symbols. This allows us to safely treat the SFT's encountered in this work as 1-step, without losing any of the generality of our statements.

A subshift which is the image of an SFT under a block code is called a *sofic shift*. Every SFT is sofic, because it is its own image under the identity. However, not every sofic shift is an SFT. The next example demonstrates this fact.

Example 2.1.1. Consider the subshift defined by the set of forbidden words $\mathcal{F} = \{10^{2n+1}1 \mid n \in \mathbb{Z}^+\}$. This shift is called the even shift, because there is always an even number of 0's between any two 1's. No finite list of forbidden words can equivalently represent this shift, so it is not an SFT. Now consider the SFT defined by $\mathcal{F} = \{11\}$. This shift, called the golden mean shift, can be represented by the following directed graph:

The block code given by $\pi(00) = 1$, $\pi(01) = \pi(10) = 0$ maps the golden mean



shift onto the even shift. Thus, the even shift is sofic, but not an SFT.

2.2 Information

Suppose \mathcal{B} is a σ -algebra on X , and μ a T -invariant measure, i.e. $\mu(B) = \mu(T^{-1}(B))$ for $B \in \mathcal{B}$. Furthermore, let ξ be a countable \mathcal{B} -measurable partition of X and \mathcal{C} a sub- σ -algebra of \mathcal{B} . The *information* of ξ is given by

$$I(\xi) = - \sum_{A \in \xi} \chi_A \log \mu(A)$$

where χ_A is the characteristic function of A .

We can interpret $I(\xi)(x)$ as the amount of information encoded in knowing which element of ξ x is contained in. The *conditional information* of ξ given \mathcal{C} is defined as

$$I(\xi | \mathcal{C}) = - \sum_{A \in \xi} \chi_A \log(\mu(A | \mathcal{C}))$$

For two partitions ξ and η we define their *join* by

$$\xi \vee \eta = \{A \cap B \mid A \in \xi, B \in \eta\}$$

The following fact about the information function will be useful in numerous calculations.

Lemma 2.2.1. *If ξ , η and ζ are countable, measurable partitions of X , then*

$$I(\xi \vee \eta | \zeta) = I(\xi | \zeta) + I(\eta | \xi \vee \zeta)$$

Proof. Expanding from the definition of I we see that

$$\begin{aligned}
& I(\xi \mid \zeta) + I(\eta \mid \xi \vee \zeta) \\
&= - \sum_{a \in \xi} \chi_A \log(\mu(A \mid \zeta)) - \sum_{B \in \eta} \chi_B \log(\mu(B \mid \xi \vee \zeta)) \\
&= - \sum_{A \in \xi, B \in \eta} \chi_{A \cap B} \log(\mu(A \mid \zeta)) - \sum_{A \in \xi, B \in \eta} \chi_{A \cap B} \log(\mu(B \mid \xi \vee \zeta)) \\
&= - \sum_{A \in \xi, B \in \eta} \chi_{A \cap B} \log(\mu(A \mid \zeta) \mu(B \mid \xi \vee \zeta)) \\
&= - \sum_{A \in \xi, B \in \eta, C \in \zeta} \chi_{A \cap B \cap C} \log \left(\frac{\mu(A \cap C)}{\mu(C)} \frac{\mu(A \cap B \cap C)}{\mu(A \cap C)} \right) \\
&= - \sum_{A \in \xi, B \in \eta, C \in \zeta} \chi_{A \cap B \cap C} \log \left(\frac{\mu(A \cap B \cap C)}{\mu(C)} \right) \\
&= - \sum_{A \in \xi, B \in \eta} \chi_{A \cap B} \log \left(\sum_{C \in \zeta} \chi_C \frac{\mu(A \cap B \cap C)}{\mu(C)} \right) \\
&= - \sum_{A \in \xi, B \in \eta} \chi_{A \cap B} \log(\mu(A \cap B \mid \zeta)) \\
&= I(\xi \vee \eta \mid \zeta)
\end{aligned}$$

Thus the identity is shown. □

2.3 Entropy

Using the same conventions as in the definition of information, we can define the *entropy* of a partition ξ as

$$H(\xi) = - \sum_{A \in \xi} \mu(A) \log \mu(A)$$

Similarly, we define the *conditional entropy* of ξ with respect to \mathcal{C} as

$$H(\xi \mid \mathcal{C}) = - \sum_{A \in \xi} \mu(A \mid \mathcal{C}) \log \mu(A \mid \mathcal{C})$$

We can see that information is related to entropy by the following formula

$$H(\xi | \mathcal{C}) = \int I(\xi | \mathcal{C}) d\mu$$

Thus entropy tells us the average amount of information in the elements of the partition. Integrating the expression in Lemma 2.2.1 gives us the following corollary.

Corollary 2.3.1. *If ξ , η and ζ are countable, measurable partitions of X , then*

$$H(\xi \vee \eta | \chi) = H(\xi | \chi) + H(\eta | \xi \vee \chi)$$

For two partitions η and ζ , let us say $\eta \leq \zeta$ if each element of η is a union of elements of ζ . We will also sometimes say η is *coarser* than ζ . The following Lemma about conditional entropy makes use of this definition.

Lemma 2.3.2. *If ξ , η and ζ are countable, measurable partitions of X , then*

- i. If $\eta \leq \zeta$ then $H(\xi | \eta) \geq H(\xi | \zeta)$.*
- ii. $H(\xi \vee \eta | \zeta) \leq H(\xi | \zeta) + H(\eta | \zeta)$*

Proof. i. First note that, because $\eta \leq \zeta$, for $A \in \xi$ and $C \in \eta$ we have

$$\sum_{D \in \zeta} \frac{\mu(D \cap C)}{\mu(C)} \frac{\mu(A \cap D)}{\mu(D)} = \frac{\mu(A \cap C)}{\mu(C)}$$

Let $\phi(x) = -x \log(x)$. Then ϕ is concave and by Jensen's inequality

$$\begin{aligned} \phi\left(\frac{\mu(A \cap C)}{\mu(C)}\right) &= \phi\left(\sum_{D \in \zeta} \frac{\mu(D \cap C)}{\mu(C)} \frac{\mu(A \cap D)}{\mu(D)}\right) \\ &\geq \sum_{D \in \zeta} \frac{\mu(D \cap C)}{\mu(C)} \phi\left(\frac{\mu(A \cap D)}{\mu(D)}\right) \end{aligned}$$

Multiplying both sides by $\mu(C)$ and summing over C and A gives

$$\begin{aligned} \sum_{A \in \xi, C \in \eta} -\mu(A \cap C) \log \left(\frac{\mu(A \cap C)}{\mu(C)} \right) \\ \geq \sum_{A \in \xi, C \in \eta, D \in \zeta} -\mu(D \cap C) \frac{\mu(A \cap D)}{\mu(D)} \log \left(\frac{\mu(A \cap D)}{\mu(D)} \right) \\ = \sum_{A \in \xi, D \in \zeta} \mu(D) \frac{\mu(A \cap D)}{\mu(D)} \log \left(\frac{\mu(A \cap D)}{\mu(D)} \right) \end{aligned}$$

Thus $H(\xi | \eta) \geq H(\xi | \zeta)$.

ii. Use (i) and Cor 2.3.1.

□

We now define some notation which will prove useful in dealing with partitions. Let (X, T) and (Y, S) be subshifts, and $\pi : X \rightarrow Y$ be a 1-block code. For any partition ξ of X we will define

$$\xi_i^j = \bigvee_{k=i}^j T^k(\xi) = T^i(\xi) \vee \dots \vee T^j(\xi)$$

The state partitions of X and Y will be called \mathcal{P}_X and \mathcal{P}_Y respectively. We will also define $\mathcal{Q} = \pi^{-1}(\mathcal{P}_Y)$.

We can now define the entropy of a measure μ on (X, T) . Let ξ be a partition which generates the sigma algebra on X (i.e. $\xi_{-\infty}^{\infty} = \mathcal{B}$). The state partition is one such partition. Then the *entropy of μ* is

$$h_T(\mu) = \lim_{n \rightarrow \infty} \frac{1}{n} H(\xi_{-n+1}^0)$$

To see that this limit exists, we will show that $a_n = H(\xi_{-n+1}^0)$ is a subadditive sequence. From this and the fact that $a_n \geq 0$ we know that $\lim_{n \rightarrow \infty} a_n/n$ exists, and is equal to $\inf a_n/n$. To see that a_n is subadditive, note that

$$\begin{aligned} H(\xi_{-n-m+1}^0) &\leq H(\xi_{-n+1}^0) + H(\xi_{-n-m+1}^{-n}) \text{ by Lemma 2.3.2} \\ &= H(\xi_{-n+1}^0) + H(\xi_{-m+1}^0) \text{ because } T \text{ is measure preserving} \end{aligned}$$

The standard definition of entropy is more involved than this, but is equivalent to the one given here in the context of subshifts.

A homeomorphism T is said to be *expansive* if $\exists \delta > 0$ such that if $x \neq y$ then $\exists n \in \mathbb{Z}$ with $d(T^n(x), T^n(y)) \geq \delta$. For any subshift, we can show that the shift map T is expansive. Suppose that $x \neq y$. Let i be the smallest number (in magnitude) such that $x_i \neq y_i$. Then $(T^i(x))_0 \neq (T^i(y))_0$. So $d(T^i(x), T^i(y)) = 2^0 = 1$. So T is expansive with $\delta = 1$. The following theorem is true in general, but we will only prove it in the case of the shift map.

For a subshift (X, T) , the space of all T -invariant measures on X will be called $\mathcal{M}(X, T)$, or sometimes just $\mathcal{M}(X)$.

Before stating the theorem, we should note that the topology we will use on $\mathcal{M}(X)$ will be the weak-* topology given by thinking of $\mathcal{M}(X)$ as $(C(X))^*$. This is the smallest topology for which each map $\mu \rightarrow \int f d\mu$ is continuous. Thus $\mu_n \rightarrow \mu$ if and only if for every $f \in C(X)$, $\int f d\mu_n \rightarrow \int f d\mu$.

Theorem 2.3.3. *If T is an expansive homeomorphism of a compact metric space, then the entropy map $h : \mathcal{M}(X) \rightarrow \mathbb{R}$ is upper semi-continuous.*

Proof. As stated, we will prove this for the shift map T , which we already know to be expansive. Let $\mu \in \mathcal{M}(X)$ and $\epsilon > 0$. We are trying to show that for some neighbourhood U of μ , $h(m) < h(\mu) + \epsilon$ for all $m \in U$. First, choose n so that

$$\frac{1}{n} H_\mu((\mathcal{P}_X)_{-n+1}^0) < h(\mu) + \frac{\epsilon}{2}$$

Let W_n denote the set of all words of length n . By the continuity of $t \log t$, for each $w \in W_n$ we can pick ϵ_w such that for all t with $|\mu[w] - t| < \epsilon_w$, we have

$$|\mu[w] \log(\mu[w]) - t \log t| < \frac{n\epsilon}{2|W_n|}$$

Taking $\epsilon_0 = \min\{\epsilon_w\}$ we can see that, for any set $\{t_w \mid |\mu[w] - t_w| < \epsilon_0\}$

$$\begin{aligned} & \left| \frac{1}{n} \sum_{w \in W_n} \mu[w] \log(\mu[w]) - \frac{1}{n} \sum_{w \in W_n} t_w \log(t_w) \right| \\ & \leq \frac{1}{n} \sum_{w \in W_n} |\mu[w] \log(\mu[w]) - t_w \log(t_w)| \\ & \leq \frac{1}{n} \frac{n\epsilon}{2|W_n|} |W_n| \\ & \leq \frac{\epsilon}{2} \end{aligned}$$

Now we let U be a neighborhood of μ given by

$$U = \{m \in \mathcal{M}(X) \mid |m[w] - \mu[w]| < \epsilon_0, \forall w \in W_n\}$$

Then by the observation above, we can see that

$$\left| \frac{1}{n} H_m((\mathcal{P}_X)_{-n+1}^0) - \frac{1}{n} H_\mu((\mathcal{P}_X)_{-n+1}^0) \right| < \frac{\epsilon}{2}$$

Thus, using the fact that $H_m((\mathcal{P}_X)_{-n+1}^0)$ is decreasing to $h(m)$,

$$h(m) < \frac{1}{n} H_m((\mathcal{P}_X)_{-n+1}^0) < \frac{1}{n} H_\mu((\mathcal{P}_X)_{-n+1}^0) + \frac{\epsilon}{2} < h(\mu) + \epsilon$$

□

Another type of entropy which we will need is the *topological entropy* of a shift map T . For a subshift (X, T) the topological entropy is defined as

$$h_{\text{top}} = \lim_{n \rightarrow \infty} \frac{\log |W_n|}{n}$$

Where W_n is again the number of words of length n . To see this limit exists, note that $|W_{n+m}| \leq |W_n| |W_m|$ so the sequence $\log |W_n|$ is subadditive.

For any finite partition $\xi = \{A_1, \dots, A_k\}$, we have that

$$\begin{aligned} \frac{1}{k}H(\xi) &= \sum_{i=1}^k -\frac{1}{k}\mu(A_i) \log \mu(A_i) \\ &= \sum_{i=1}^k -\frac{1}{k}\mu(A_i) \log \mu(A_i) \\ &\leq -\left(\sum_{i=1}^k \frac{1}{k}\mu(A_i)\right) \log \left(\sum_{i=1}^k \frac{1}{k}\mu(A_i)\right) \\ &= -\frac{1}{k} \log \frac{1}{k} \end{aligned}$$

Where the inequality is obtained by applying Jensen's inequality to $-x \log x$ with weights $\frac{1}{k}$. This shows that $H(\xi) \leq \log k$. Note that $|(\mathcal{P}_X)_{-n+1}^0| = |W_n|$, so $H((\mathcal{P}_X)_{-n+1}^0) \leq \log |W_n|$. This tells us that $h(\mu) \leq h_{\text{top}}$. We will later see that $h_{\text{top}} = \sup_{\mu \in \mathcal{M}(X)} h(\mu)$.

2.4 Some results from ergodic theory

In this section we will recall three of the foundational theorems of ergodic theory. Proofs of these theorems can be found in [21].

The Poincaré recurrence theorem actually predates ergodic theory, and exists in many contexts outside of invariant measures. Still, one of its most common formulations is essentially about the asymptotic properties of a statistically significant set of points, so in this sense it is a kindred spirit.

Theorem 2.4.1 (Poincaré recurrence theorem). *Let (X, \mathcal{B}, T) be a subshift and $\mu \in \mathcal{M}(T)$. For any $A \in \mathcal{B}$ and μ -a.e. $x \in X$ there exists a sequence $n_k \rightarrow \infty$ such that $T^{n_k}x \in A$.*

If (X, \mathcal{B}, T) is a subshift then a measure $\mu \in \mathcal{M}(X)$ is *ergodic* if for every $E \in \mathcal{B}$ such that $T^{-1}(E) = E$ we have $\mu(E) = 0$ or $\mu(E) = 1$. So an ergodic measure is one for which every T -invariant set has either full or zero measure.

The Birkhoff ergodic theorem (BET) is probably the most well known result in ergodic theory, being both one of the foundational theorems of the field and one of its most frequently cited results. It has given rise to many generalizations and inspired

numerous other types of ergodic theorems, but we give here a rather simple version. The essence of the BET is that for almost every point in our space, the integral of continuous function is equal to the average of that function along the orbit of a point.

Theorem 2.4.2 (Birkhoff ergodic theorem). *Let (X, \mathcal{B}, T) be a subshift and $\mu \in \mathcal{M}(X)$ an ergodic measure. Then for every $f \in C(X)$*

$$\int f d\mu = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} f \circ T^i(x)$$

for μ -a.e. $x \in X$.

The Shannon-McMillan-Breiman theorem is an invaluable tool for understanding the relationship between the entropy of a partition with respect to an ergodic measure and the weight that measure gives to words in the partition.

Theorem 2.4.3 (Shannon-McMillan-Breiman theorem). *Let (X, \mathcal{B}, T) be a subshift and $\mu \in \mathcal{M}(X)$ an ergodic measure. For any partition α of X such that $H_\mu(\alpha) < \infty$*

$$\lim_{n \rightarrow \infty} \frac{1}{n} I(\alpha_{-n+1}^0)(x) = H(\alpha \mid \alpha_{-\infty}^{-1})$$

for μ -a.e. $x \in X$.

If α is a generating partition for X then the entropy in this theorem will be the measure entropy of μ . By taking α to be the state partition, we have the following interpretation of Theorem 2.4.3. As we have noted before, α_{-n+1}^0 is the set of words of length n in X . So for very large n and μ -a.e. $x \in X$ we have

$$\mu([x_0 \dots x_{n-1}]) \sim e^{-nh(\mu)}$$

2.5 Pressure

Let (X, T, \mathcal{B}) be a shift of finite type and $f \in C(X)$. In the following sections we will define a map $\mathcal{P} : C(X) \rightarrow \mathbb{R} \cup \{\infty\}$ called the pressure of f . The pressure is a refinement of entropy in the sense that $\mathcal{P}(0) = h_{\text{top}}$. Pressure takes both the action of T and the weight of a potential function f into account. It also has many nice properties which can be exploited to find measures in $\mathcal{M}(X)$ which are naturally related to f .

2.5.1 Pressure for Shifts of Finite Type

Pressure can be defined very generally, and has numerous equivalent definitions, but for shifts of finite type the following definition will be sufficient.

Definition 2.5.1. Let (X, T, \mathcal{B}) be a shift of finite type. Given a continuous real-valued function $f \in C(X)$ we define the pressure of f to be

$$\mathcal{P}(f) = \lim_{n \rightarrow \infty} \frac{1}{n} \log \sum_{w \in W_n} e^{(S_n f)[w]}$$

where

$$(S_n f)[w] = \sup_{x \in [w]} \sum_{i=0}^{n-1} f(T^i(x))$$

When actually computing pressure, we can simply let $(S_n f)[w] = (S_n f)(x)$ for some $x \in [w]$. To see this, first let w be a word of length n and $x, y \in [w]$. Because f is continuous, for every ϵ there is a δ such that $d(x, y) < \delta$ implies $|f(x) - f(y)| < \epsilon$. Because x and y are in $[w]$, we see that

$$d(T^i(x), T^i(y)) = \max(2^{-(i+1)}, 2^{-(n-i)}) \text{ for } 0 \leq i \leq n-1$$

So granted that n is large enough, if

$$-\log \delta - 1 < i < \log \delta + n$$

then $d(T^i(x), T^i(y)) < \delta$. So at most $2 \lceil -\log \delta \rceil$ terms do not satisfy the inequality. Let G be the set of i such that $d(T^i(x), T^i(y)) < \delta$. We have

$$\begin{aligned} & |(S_n f)(x) - (S_n f)(y)| \\ & \leq \sum_{i=0}^{n-1} |f(T^i(x)) - f(T^i(y))| \\ & = \sum_{i \in G} |f(T^i(x)) - f(T^i(y))| + \sum_{i \notin G} |f(T^i(x)) - f(T^i(y))| \\ & \leq (n\epsilon + 4\|f\| \lceil -\log \delta \rceil) \end{aligned}$$

This implies that, up to choice of x , the most $(S_n f)[w]$ can vary is $n\epsilon + c$ where c is

a constant with respect to n . If for each w we choose $x(w) \in [w]$, we have

$$\frac{1}{n} \log \sum_{w \in W_n} e^{(S_n f)(x(w)) - (n\epsilon + c)} \leq \frac{1}{n} \log \sum_{w \in W_n} e^{(S_n f)[w]} \leq \frac{1}{n} \log \sum_{w \in W_n} e^{(S_n f)(x(w)) + (n\epsilon + c)}$$

Thus

$$\frac{1}{n} \sum_{w \in W_n} e^{(S_n f)(x(w))} - \frac{c}{n} - \epsilon \leq \frac{1}{n} \sum_{w \in W_n} e^{(S_n f)[w]} \leq \frac{1}{n} \sum_{w \in W_n} e^{(S_n f)(x(w))} + \frac{c}{n} + \epsilon$$

Taking limits in n establishes the claim.

One relation between pressure and entropy is immediately obvious. As stated above, we can see that:

$$\begin{aligned} \mathcal{P}(0) &= \lim_{n \rightarrow \infty} \frac{1}{n} \log \sum_{w \in W_n} 1 \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} \log |W_n| \\ &= h_{\text{top}} \end{aligned}$$

Example 2.5.1. Let f be the function $f(x) = x_0$ and (X, T) the full shift on $\{0, 1\}^{\mathbb{N}}$. Then if a word w of length n has k 1's, $(S_n f)(x) = k$. There are $\binom{n}{k}$ such words, so:

$$\begin{aligned} \mathcal{P}(f) &= \lim_{n \rightarrow \infty} \frac{1}{n} \log \sum_{k=0}^n \binom{n}{k} e^k \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} \log (1 + e)^n \\ &= \log (1 + e) \end{aligned}$$

We will now summarize some basic properties of pressure.

Theorem 2.5.1. Let (X, T, \mathcal{B}) be a shift of finite type, $f, g \in C(X)$ and $c \in \mathbb{R}$. Then

- i. If $f \leq g$ then $\mathcal{P}(f) \leq \mathcal{P}(g)$
- ii. $h_{\text{top}} + \inf f \leq \mathcal{P}(f) \leq h_{\text{top}} + \sup f$
- iii. $\mathcal{P}(f + c) = \mathcal{P}(f) + c$

iv. $\mathcal{P}(\cdot)$ is convex.

v. $\mathcal{P}(f + g \circ T - g) = \mathcal{P}(f)$

vi. $\mathcal{P}(f + g) \leq \mathcal{P}(f) + \mathcal{P}(g)$

vii. If $c \geq 1$ then $\mathcal{P}(cf) \leq c\mathcal{P}(f)$

Proof. (i) This is clear from the definition of pressure.

(ii) Let $L(n)$ be the number of words in X of length n . For all words w , $n \inf f \leq (S_n f)[w] \leq n \sup f$ so

$$\frac{1}{n} \log \sum_{w \in W_n} e^{n \inf f} \leq \frac{1}{n} \log \sum_{w \in W_n} e^{S_n f[w]} \leq \frac{1}{n} \log \sum_{w \in W_n} e^{n \sup f}$$

Note that

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{n} \log \sum_{w \in W_n} e^{n \inf f} &= \lim_{n \rightarrow \infty} \frac{1}{n} \log (L(n) e^{n \inf f}) \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} (\log(L(n)) + n \inf f) \\ &= h_{\text{top}} + \inf f \end{aligned}$$

We can similarly show that the right side of the inequality is equal to $h_{\text{top}} + \sup f$ in the limit.

(iii) $S_n(f + c)[w] = (S_n f)[w] + nc$, so we have

$$\begin{aligned} \log \sum_{w \in W_n} e^{S_n(f+c)[w]} &= \log \sum_{w \in W_n} e^{(S_n f)[w] + nc} \\ &= \log e^{nc} \sum_{w \in W_n} e^{(S_n f)[w]} \\ &= nc + \log \sum_{w \in W_n} e^{(S_n f)[w]} \end{aligned}$$

After taking limits we arrive at the desired equality.

(iv) Let $0 \leq p \leq 1$. Then $\frac{1}{1/p} + \frac{1}{1/(1-p)} = 1$. So by Hölder's inequality, we know that

$$\begin{aligned} & \sum_{w \in W_n} e^{p(S_n f)[w] + (1-p)(S_n g)[w]} \\ & \leq \left(\sum_{w \in W_n} (e^{(p(S_n f)[w])})^{\frac{1}{p}} \right)^p \left(\sum_{w \in W_n} (e^{((1-p)(S_n g)[w])})^{\frac{1}{1-p}} \right)^{1-p} \\ & = \left(\sum_{w \in W_n} (e^{(S_n f)[w]}) \right)^p \left(\sum_{w \in W_n} (e^{(S_n g)[w]}) \right)^{1-p} \end{aligned}$$

Taking logarithms and limits we arrive at

$$\mathcal{P}(pf + (1-p)g) \leq p\mathcal{P}(f) + (1-p)\mathcal{P}(g)$$

(v) It is easy to show that $S_n(f + g \circ T - g)[w] = (S_n f)[w] + g \circ T^n(x) - g(x)$, for some x such that $[x]_0^{n-1} = w$, the choice of which will not matter in the limit. Then

$$e^{-2\|g\|} \sum_{w \in W_n} e^{(S_n f)[w]} \leq \sum_{w \in W_n} e^{(S_n f)[w] + g \circ T^n(x) - g(x)} \leq e^{2\|g\|} \sum_{w \in W_n} e^{(S_n f)[w]}$$

From this, the result is clear.

(vi) We can see that

$$\sum_{w \in W_n} e^{S_n(f+g)[w]} \leq \sum_{w \in W_n} e^{(S_n f)[w]} \sum_{w \in W_n} e^{(S_n g)[w]}$$

from which the inequality is clear.

(vii) Let $A_n = \sum_{w \in W_n} e^{(S_n f)[w]}$. Then

$$\sum_{w \in W_n} \frac{e^{(S_n f)[w]}}{A_n} = 1$$

So because $c \geq 1$

$$\sum_{w \in W_n} \left(\frac{e^{(S_n f)[w]}}{A_n} \right)^c \leq 1$$

and thus

$$\sum_{w \in W_n} e^{c(s_n f)[w]} \leq \left(\sum_{w \in W_n} e^{(s_n f)[w]} \right)^c$$

The result follows from this. □

2.5.2 The Variational Principle

One of the most important properties of pressure is that it satisfies the Variational Principle:

Theorem 2.5.2. *Let (X, \mathcal{B}, T) be a shift of finite type and $f \in C(X)$. Then*

$$\mathcal{P}(f) = \sup_{\mu \in \mathcal{M}(X)} \left\{ h(\mu) + \int f d\mu \right\}$$

In fact, the Variational Principle is so fundamental that some authors define pressure as this quantity to begin with.

We have already seen that $\mathcal{P}(0) = h_{\text{top}}$. We can now also see that $\mathcal{P}(0) = \sup_{\mu \in \mathcal{M}(X)} h(\mu)$, thus verifying our earlier claim.

A closely related idea is that of equilibrium states.

Definition 2.5.2. A measure $\mu \in \mathcal{M}(X)$ is called an equilibrium state if

$$\mathcal{P}(f) = h(\mu) + \int f d\mu$$

One natural question is under what conditions equilibrium states exist. If T is an expansive homeomorphism, we have already seen that entropy is upper semi-continuous. Thus $h(\mu) + \int f d\mu$ is upper semi-continuous as well. The invariant measures on X form a compact set, and an upper semi-continuous function on a compact set attains its maximum, so we know that for any $f \in C(X)$ some measure

will attain the supremum in

$$\sup_{\mu \in \mathcal{M}(X)} \left\{ h(\mu) + \int f d\mu \right\}$$

So in this case, every continuous function has an equilibrium state. Later, we will answer the question of when there is a unique equilibrium state, for a particular type of f .

Upper semi-continuity of h also allows us to prove the following corollary of the variational principle. This fact can be found in, for example, [18][Theorem 6.14].

Corollary 2.5.3. *For $\mu_0 \in \mathcal{M}(X)$,*

$$h(\mu_0) = \inf_{f \in C(X)} \left\{ \mathcal{P}(f) - \int f d\mu_0 \right\}$$

Before we proceed, let us prove a corollary to Jensen's inequality which will be useful in the discussion to follow, as well as in later proofs.

Lemma 2.5.4. *Let q_1, \dots, q_n be a finite sequence of positive numbers and let \mathbb{P}_n denote the space of probability vectors of length n . Then*

$$\max_{p \in \mathbb{P}_n} - \sum p_i \log(p_i/q_i) = \log \left(\sum q_i \right)$$

With equality being obtained only if $p_i = \frac{q_i}{\sum q_i}$.

Proof. Let $Q = \sum q_i$ and $s_i = q_i/Q$. Then for any probability vector p :

$$\begin{aligned} - \sum p_i \log(p_i/q_i) &= - \sum p_i \log \left(\frac{p_i/Q}{q_i/Q} \right) \\ &= - \sum (p_i \log(p_i/s_i) - p_i \log(Q)) \\ &= \sum (p_i \log(s_i/p_i)) + \log Q \end{aligned}$$

By Jensen's inequality we know that $\sum p_i \log(s_i/p_i) \leq \log(\sum p_i(s_i/p_i)) = 0$ with equality exactly when $s_i/p_i = \sum p_i(s_i/p_i) = 1$. In other words, equality occurs when $p_i = s_i = \frac{q_i}{\sum q_i}$. So continuing the calculation above we have

$$- \sum p_i \log(p_i/q_i) \leq \sum (p_i \log(s_i/p_i)) + \log Q \leq \log Q$$

□

Corollary 2.5.5. *If a_1, \dots, a_k are real numbers then letting $q_i = e^{a_i}$ in the lemma gives us that*

$$\sum_{i=0}^k p_i(a_i - \log p_i) \leq \log \sum_{i=0}^k e^{a_i}$$

We will now sketch a proof of the variational principle. We will be able to prove that for any $\mu \in \mathcal{M}(X)$, $\mathcal{P}(f) \geq h(\mu) + \int f d\mu$. We will then hint at how the opposite inequality is proved.

Proof. For the first part of the proof of the variational principle, we let $\mu \in \mathcal{M}(X)$. Let W_n denote the words in X of length n and $\alpha(w) = \sup\{(S_n f)(x) \mid x \in [w]\}$. The first key observation is that

$$\begin{aligned} H_\mu((\mathcal{P}_X)_{-n+1}^0) + \int (S_n f) d\mu &\leq \sum_{w \in W_n} \mu([w])[-\log \mu([w]) + \alpha([w])] \\ &\leq \log \left(\sum_{w \in W_n} e^{\alpha(w)} \right) \end{aligned}$$

This last inequality is due to Corollary 2.5.5. Now for each $w \in W_n$ we are able to choose an $x(w) \in [w]$ such that $\alpha(w) = (S_n f)(x(w))$. So we have

$$\sum_{w \in W_n} e^{\alpha(w)} \leq \sum_{w \in W_n} e^{(S_n f)(x(w))}$$

Taking logarithms gives us

$$\log \left(\sum_{w \in W_n} e^{\alpha(w)} \right) \leq \log \left(\sum_{w \in W_n} e^{(S_n f)(x(w))} \right)$$

and thus

$$\begin{aligned} \frac{1}{n} H_\mu ((\mathcal{P}_X)^0_{-n+1}) + \int f d\mu &= \frac{1}{n} H_\mu ((\mathcal{P}_X)^0_{-n+1}) + \frac{1}{n} \int (S_n f) d\mu \\ &\leq \frac{1}{n} \log \left(\sum_{w \in W_n} e^{(S_n f)(x(w))} \right) \end{aligned}$$

We can see that, after taking limits, this simplifies to the desired inequality:

$$h(\mu) + \int f d\mu \leq \mathcal{P}(f)$$

For the second part of the proof, we produce a sequence of measures which are almost equilibrium states for f . First, for each word of length n choose some $x(w) \in [w]$ and set

$$\sigma_n = \frac{\sum_{w \in W_n} e^{(S_n f)(x(w))} \delta_{x(w)}}{\sum_{w \in W_n} e^{(S_n f)(x(w))}}$$

Let us justify this choice of measure. Note that

$$\begin{aligned} H_{\sigma_n} ((\mathcal{P}_X)^0_{-n+1}) + \int (S_n f) d\sigma_n &= \sum_{w \in W_n} (-\sigma_n([w]) \log \sigma_n([w]) + \sigma_n([w])(S_n f)[w]) \\ &= \log \left(\sum_{w \in W_n} e^{(S_n f)[w]} \right) \end{aligned}$$

Again, the last equality is from Lemmas 2.5.4 and 2.5.5. Also, the last expression is exactly the term that occurs in the definition of pressure, thereby justifying this choice of measure.

However, σ_n is not in general a T -invariant measure. To deal with this, we define $\mu_n = \frac{1}{n} \sum_{i=0}^{n-1} \sigma_n \circ T^{-i}$. A similar calculation to that which was carried out with σ_n shows that μ_n is a near equilibrium state. We know that $C(X)^*$ is the set of finite signed Borel measures on X , and by Alaoglu's Theorem the unit ball of $C(X)^*$ is compact in the weak-* topology. The Borel probability measures form a closed subset of this compact set. Thus μ_n has a convergent subsequence. The measure that this converges to, μ , will have the desired property, namely that $\mathcal{P}(f) \leq h(\mu) +$

$\int f d\mu$. Furthermore, μ is T -invariant. To show this, it is sufficient to show that $\int (g - g \circ T) d\mu = 0$ for all $g \in C(X)$. Because μ is the weak-* limit of a subsequence of μ_n , we know that $\int (g - g \circ T) d\mu_n \rightarrow \int (g - g \circ T) d\mu$. Furthermore

$$\begin{aligned} \int (g - g \circ T) d\mu_n &= \frac{1}{n} \int \sum_{i=0}^{n-1} (g \circ T^i - g \circ T^{i+1}) d\sigma \\ &= \frac{1}{n} \int (g - g \circ T) d\sigma \\ &\leq \frac{2\|g\|}{n} \end{aligned}$$

So we have established that $\mu \in \mathcal{M}(X)$. □

In order to get a better feel for the variational principle, we will show how some parts of 2.5.1 could be proven using it.

Proof. (iv) Let $f, g \in C(X)$ and $0 \leq p \leq 1$. Then

$$\begin{aligned} &\mathcal{P}(pf + (1-p)g) \\ &= \sup_{\mu \in \mathcal{M}(X)} \left\{ h(\mu) + \int (pf + (1-p)g) d\mu \right\} \\ &= \sup_{\mu \in \mathcal{M}(X)} \left\{ ph(\mu) + (1-p)h(\mu) + p \int f d\mu + (1-p) \int g d\mu \right\} \\ &\leq p \left[\sup_{\mu \in \mathcal{M}(X,T)} \left\{ h(\mu) + \int f d\mu \right\} \right] + (1-p) \left[\sup_{\mu \in \mathcal{M}(X)} \left\{ h(\mu) + \int g d\mu \right\} \right] \\ &= p\mathcal{P}(f) + (1-p)\mathcal{P}(g) \end{aligned}$$

Thus pressure is convex.

(vi) Keeping in mind that $h(\mu) \geq 0$ we see that

$$\begin{aligned} \mathcal{P}(f+g) &= \sup_{\mu \in \mathcal{M}(X)} \left\{ h(\mu) + \int (f+g) d\mu \right\} \\ &\leq \sup_{\mu \in \mathcal{M}(X)} \left\{ h(\mu) + \int f d\mu \right\} + \sup_{\mu \in \mathcal{M}(X)} \left\{ h(\mu) + \int g d\mu \right\} \\ &= \mathcal{P}(f) + \mathcal{P}(g) \end{aligned}$$

(vii) Let $c > 1$. Then

$$\begin{aligned} \mathcal{P}(cf) &= \sup_{\mu \in \mathcal{M}(X)} \left\{ h(\mu) + \int cf d\mu \right\} \\ &\leq \sup_{\mu \in \mathcal{M}(X)} \left\{ ch(\mu) + c \int f d\mu \right\} \\ &= c \sup_{\mu \in \mathcal{M}(X)} \left\{ h(\mu) + \int f d\mu \right\} \\ &= c\mathcal{P}(f) \end{aligned}$$

□

Example 2.5.2. In Example 2.5.1 we found that for $f(x) = x_0$ on the full two-shift, $\mathcal{P}(f) = \log(1+e)$. One could also arrive at this using the variational principle. First let us note that $\mathcal{P}(f) = \sup_v \sup_{\mu[1]=v} \{h(\mu) + \int f d\mu\}$. So we begin by assuming that $\mu[1] = v$ and trying to maximize the inner expression. Because f only depends on the first coordinate, this fixes $\int f d\mu$. Thus, we focus on maximizing $h(\mu)$ first.

Let \mathcal{P}_X be the state partition. Then $h(\mu) = \lim_{n \rightarrow \infty} \frac{1}{n} H((\mathcal{P}_X)^0_{-n+1})$. To maximize the term on the right hand side, we note that

$$H((\mathcal{P}_X)^0_{-n+1}) = H(\mathcal{P}_X) + H(\mathcal{P}_X | T^{-1}(\mathcal{P}_X)) + \dots + H(\mathcal{P}_X | (\mathcal{P}_X)^{-1}_{-n+1})$$

Because we have fixed $\mu[1] = v$ we know that $H(\mathcal{P}_X) = -v \log v - (1-v) \log(1-v)$. We also know that

$$H(\mathcal{P}_X) \geq H(\mathcal{P}_X | T^{-1}(\mathcal{P}_X)) \geq \dots \geq H(\mathcal{P}_X | (\mathcal{P}_X)^{-1}_{-n+1})$$

so by combining these observations we have

$$\begin{aligned} H(\mathcal{P}_X) + H(\mathcal{P}_X | T^{-1}(\mathcal{P}_X)) + \dots + H(\mathcal{P}_X | (\mathcal{P}_X)_{-n+1}^{-1}) \\ \leq n(-v \log v - (1-v) \log(1-v)) \end{aligned}$$

Thus we can say that this expression is maximized when all the terms in the sum are equal to $H(\mathcal{P}_X)$. So for all n , $H(\mathcal{P}_X) = H(\mathcal{P}_X | (\mathcal{P}_X)_{-n+1}^{-1})$. By the increasing Martingale convergence theorem, we have that $H(\mathcal{P}_X) = H(\mathcal{P}_X | (\mathcal{P}_X)_{-\infty}^{-1})$. So \mathcal{P}_X and $(\mathcal{P}_X)_{-\infty}^{-1}$ are independent. This tells us that the μ which maximizes the expression must be a Bernoulli measure.

Now we find v which maximizes the outer term. Again the variational principle tells us that:

$$\mathcal{P}(f) = \sup_{\mu} \left\{ h(\mu) + \int f d\mu \right\} = \max_v (-v \log v - (1-v) \log(1-v) + v)$$

To find the maximum, we take the derivative of this equation and find critical points.

$$\begin{aligned} 1 - \log(v) - 1 + \log(1-v) + 1 &= 0 \\ \log\left(\frac{v}{1-v}\right) &= 1 \\ \frac{v}{1-v} &= e \\ v(1+e) &= e \\ v &= \frac{e}{1+e} \end{aligned}$$

This also gives us that $1-v = 1/(1+e)$. So

$$\mathcal{P}(f) = -\frac{e}{1+e} \log\left(\frac{e}{1+e}\right) - \frac{1}{1+e} \log\left(\frac{1}{1+e}\right) + \frac{e}{1+e} = \log(1+e)$$

It is important to note that, while this calculation was more complex than Example 2.5.1, we not only determined the pressure of f , but we also identified the equilibrium state as a Bernoulli measure. This extra information about equilibrium states is often obtained via the variational principle, and is one of its main advantages in calculating pressure.

2.5.3 Pressure of a Function of Two Coordinates

Definition 2.5.3. If M is an irreducible $k \times k$ matrix with largest positive eigenvalue β and left eigenvector l then the *left stochasticization* of M is the matrix P given by

$$P(i, j) = \frac{M(i, j)l(i)}{\beta l(j)}$$

To see that the matrix P is left stochastic, we first note that, because β is an eigenvalue and l a left eigenvector

$$\beta l(j) = \sum_{i=1}^k M(i, j)l(i)$$

thus

$$1 = \sum_{i=1}^k \frac{M(i, j)l(i)}{\beta l(j)}$$

Definition 2.5.4. Let (X, \mathcal{B}, μ, T) be a shift of finite type. The information cocycle of μ is

$$I_\mu = I(\mathcal{P}_X | (\mathcal{P}_X)_{-\infty}^{-1})$$

In general, I_μ is only defined μ -a.e.. This creates an issue when trying to integrate I_μ against measures other than μ . However, when μ is a Markov measure, I_μ has a continuous version which we will describe in a moment. This allows us to integrate I_μ against any measure without ambiguity.

Let us first show that a left stochastic matrix P together with a right eigenvector p (so that $Pp = p$) defines an invariant measure μ . Let $\mu([x_0 \dots x_n]) = P(x_0, x_1) \dots P(x_{n-1}, x_n)p(x_n)$. Then $\mu([x_{-1}x_0 \dots x_n]) = P(x_{-1}, x_0)\mu([x_0 \dots x_n])$. To see this is invariant, we note that

$$\sum_{x_{-1} \in \mathcal{A}} P(x_{-1}, x_0)P(x_0, x_1) \dots P(x_{n-1}, x_n)p(x_n) = P(x_0, x_1) \dots P(x_{n-1}, x_n)p(x_n)$$

To check that this is a measure, we see that

$$\sum_{x_{n+1} \in \mathcal{A}} P(x_0, x_1) \dots P(x_n, x_{n+1})p(x_{n+1}) = P(x_0, x_1) \dots P(x_{n-1}, x_n)p(x_n)$$

By the Daniell-Kolmogorov consistency theorem, this is sufficient to show that μ is

an invariant measure [21]. For a measure μ defined as above, we can see that for $x_0 \in \mathcal{A}$:

$$\begin{aligned} \mu([x_0] | (\mathcal{P}_X)_{-n+1}^{-1}) &= \sum_{P \in (\mathcal{P}_X)_{-n+1}^{-1}} \chi_P \frac{\mu([x_0] \cap P)}{\mu(P)} \\ &= \sum_{P = [x_1 x_2 \dots x_{n-1}]_1^{n-1}} \chi_P \left(\frac{P(x_0, x_1) \dots P(x_{n-2}, x_{n-1}) p(x_n)}{P(x_1, x_2) \dots P(x_{n-2}, x_{n-1}) p(x_n)} \right) \\ &= P(x_0, x_1) \text{ on } [x_1 \dots x_{n-1}]_1^{n-1} \end{aligned}$$

Note that this is independent of n , so we can say that $\mu([x_0] | (\mathcal{P}_X)_{-\infty}^{-1})(x) = P(x_0, x_1)$. Thus

$$\begin{aligned} I_\mu(x) &= I(\mathcal{P}_X | (\mathcal{P}_X)_{-\infty}^{-1})(x) \\ &= -\log \mu([x_0] | (\mathcal{P}_X)_{-\infty}^{-1}) \\ &= -\log(P(x_0, x_1)) \end{aligned}$$

The following theorem shows us that the equilibrium state of a function that depends only on the first two coordinates of x is a Markov measure. This fact will be useful in later proofs, and for calculating explicit examples of equilibrium states.

Theorem 2.5.6. *Let (X, \mathcal{B}, μ, T) be a shift of finite type with left transition matrix A . If $f \in C(X)$ depends on only two coordinates, i.e. $f(x) = f(x_0, x_1)$, then the unique equilibrium state of f is a 1-step Markov measure μ given by the left stochastization of the matrix*

$$M(i, j) = A(i, j)e^{f(i, j)}$$

The pressure of f is given by $\log(\beta)$ where β is the largest positive eigenvalue of M .

Before we begin the proof of this theorem, let us note that it answers a question that was raised earlier. This theorem identifies the *unique* equilibrium state of a function of two coordinates, which is a Markov measure.

Proof. The proof of this theorem will come in two stages. First we will show that μ is an equilibrium state of the function. Then we will show that μ is the unique

equilibrium state.

To show that μ is an equilibrium state, let us consider M^n , where M is the matrix given above. In this case,

$$M_{i,j}^n = \sum_{ix_1 \dots x_{n-2} j \in W_n} e^{f(i,x_0) + \dots + f(x_{n-2},j)}$$

Thus

$$\begin{aligned} & \sum_{w \in W_n} \exp \left(\sup_{x \in [w]} (S_n f)(x) \right) \\ & \leq \sum_{i,j} \sum_{ix_1 \dots x_{n-2} j \in W_n} \exp(f(i, x_0) + \dots + f(x_{n-2}, j)) \\ & \leq |\mathcal{A}(X)| \sum_{w \in W_n} \exp \left(\sup_{x \in [w]} (S_n f)(x) \right) \end{aligned}$$

This shows us that

$$\begin{aligned} & \frac{1}{|\mathcal{A}(X)|} \sum_{i,j} \sum_{ix_1 \dots x_{n-2} j \in W_n} \exp(f(i, x_0) + \dots + f(x_{n-2}, j)) \\ & \leq \sum_{w \in W_n} \exp \left(\sup_{x \in [w]} (S_n f)(x) \right) \\ & \leq \sum_{i,j} \sum_{ix_1 \dots x_{n-2} j \in W_n} \exp(f(i, x_0) + \dots + f(x_{n-2}, j)) \end{aligned}$$

So we see that we can calculate pressure using $\sum_{i,j} M_{i,j}^n$ instead of $\sum_{w \in W_n} e^{(S_n f)[w]}$.

Substituting this into the expression for pressure gives us

$$\begin{aligned}
\mathcal{P}(f) &= \lim_{n \rightarrow \infty} \frac{1}{n} \log \sum_{w \in W_n} e^{(S_n f)(w)} \\
&= \lim_{n \rightarrow \infty} \frac{1}{n} \log \sum_{i,j} M_{i,j}^n \\
&= \lim_{n \rightarrow \infty} \frac{1}{n} \log \left(\begin{array}{cccc} 1 & & & \\ & 1 & & \\ & & \dots & \\ & & & 1 \end{array} \right) M^n \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}
\end{aligned}$$

From its construction, we know that M is non-negative and irreducible. So by the Perron-Frobenius theorem [4] it has an eigenvalue $\beta \in \mathbb{R}^+$ such that for any other eigenvalue λ , $|\lambda| < \beta$. In addition, the eigenvector corresponding to β has non-negative real entries. Let l_1 and r_1 be left and right eigenvectors corresponding to β and $l_2, \dots, l_k, r_2, \dots, r_k$ correspond to the other eigenvalues of M . We can choose these eigenvectors so that $l_i \cdot r_i = 1$. We also know that $l_i \cdot r_j = 0$ for $i \neq j$. Writing

$$\begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} = \sum_{i=1}^k a_i r_i \text{ we have that}$$

$$l_1 \cdot \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} = \sum_{i=1}^k a_i l_1 \cdot r_i = a_1$$

Thus l_1 having non-negative entries tells us $a_1 > 0$. Now we see that

$$\begin{aligned} \begin{pmatrix} 1 & 1 & \dots & 1 \end{pmatrix} M^n \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} &= \begin{pmatrix} 1 & 1 & \dots & 1 \end{pmatrix} \left(a_1 \beta^n r_1 + \sum_{i=2}^k a_i \lambda_i^n r_i \right) \\ &= a_1 \beta^n \begin{pmatrix} 1 & 1 & \dots & 1 \end{pmatrix} r_1 + o(\beta^n) \end{aligned}$$

So we have shown that $\begin{pmatrix} 1 & 1 & \dots & 1 \end{pmatrix} M^n \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$ has β^n growth rate. Thus we have

$$\begin{aligned} \mathcal{P}(f) &= \lim_{n \rightarrow \infty} \frac{1}{n} \log k \beta^n (1 + o(1)) \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} (n \log \beta + \log [k (1 + o(1))]) \\ &= \log \beta \end{aligned}$$

We now show that $h(\mu) + \int f d\mu = \log \beta$. Recall from the statement of the theorem that the measure μ is given by the matrix P with $P(i, j) = \frac{M(i, j) l(i)}{\beta l(j)}$. If $A(x_0, x_1) = 1$ then

$$I_\mu = -\log(P(x_0, x_1)) = -f(x_0, x_1) + \log(\beta) + \log(l(x_1)) - \log(l(x_0))$$

Taking $g(x) = \log(l(x_0))$, we have

$$I_\mu(x) + f(x) = \log(\beta) - g(x) + g \circ T(x)$$

So we have that $\int (I_\mu + f) d\mu = \log(\beta)$. This shows that μ is an equilibrium state of f .

Next we let $m \in \mathcal{M}(X)$ and show

$$h(\mu) + \int f d\mu = \int (I_\mu + f) d\mu \geq \int (I_m + f) dm = h(m) + \int f dm$$

We have already shown that there is a function $g(x)$ such that $I_\mu(x) + f(x) = \log(\beta) - g(x) + g \circ T(x)$. Thus, $\int (I_\mu + f) d\mu = \log \beta$ as well. We now have

$$\int (I_\mu + f) d\mu = \log \beta = \int (I_\mu + f) dm$$

If we can show that $\int I_\mu dm \geq \int I_m dm$ then we will have the desired inequality.

From the discussion above, $I_\mu(x) = -\log P(x_0, x_1) = -\sum_{i \in \mathcal{A}} \chi_{[i]} \log P(i, x_1)$. Conditioning this on $(\mathcal{P}_X)_{-\infty}^{-1}$ gives us

$$\begin{aligned} \mathbb{E}_m (\log P(x_0, x_1) \mid (\mathcal{P}_X)_{-\infty}^{-1}) &= \sum_{i \in \mathcal{A}} \mathbb{E}_m (\chi_{[i]} \mid (\mathcal{P}_X)_{-\infty}^{-1}) \log P(i, x_1) \\ &= \sum_{i \in \mathcal{A}} m([i] \mid (\mathcal{P}_X)_{-\infty}^{-1}) \log \mu([i] \mid (\mathcal{P}_X)_{-\infty}^{-1}) \end{aligned}$$

Recalling Lemma 2.5.4, we now have that

$$\begin{aligned} \int (I_m - I_\mu) dm &= \int \left(\sum_{i \in \mathcal{A}} m([i] \mid (\mathcal{P}_X)_{-\infty}^{-1}) \log \frac{\mu([i] \mid (\mathcal{P}_X)_{-\infty}^{-1})}{m([i] \mid (\mathcal{P}_X)_{-\infty}^{-1})} \right) dm \\ &\leq \int \left(\log \sum_{i \in \mathcal{A}} \mu([i] \mid (\mathcal{P}_X)_{-\infty}^{-1}) \right) dm \\ &= 0 \end{aligned}$$

Also, by the same corollary, equality will occur if and only if $m([i] \mid (\mathcal{P}_X)_{-\infty}^{-1}) = \mu([i] \mid (\mathcal{P}_X)_{-\infty}^{-1}) = -\log P(x_0, x_1)$. The right hand side tells us that m must be a Markov measure. But then m is a Markov measure that has the same conditional probability as μ . Also, μ is strongly irreducible. Thus $\mu = m$. This establishes μ as the unique equilibrium state of f . \square

Chapter 3

Relative Pressure, Relative Equilibrium States and Compensation Functions

3.1 Relative Entropy and Relative Pressure

In this section, we study the case where (X, T) is a shift of finite type and (Y, S) is a sofic factor of X by the code $\pi : X \rightarrow Y$. We will sometimes refer to such a collection $F = (X, Y, \pi)$ as a *factor triple*. Our goal is to define a function on X which shares many of the useful properties of pressure, such as satisfying a variational principle and convexity, but which bears some meaningful relationship to the code π . To do this, we must first define the concept of relative entropy.

Definition 3.1.1. Let (X, T) be an SFT and (Y, S) a factor of X via the block code π . Recall that \mathcal{Q} is the partition of X induced by $\pi^{-1}(\mathcal{P}_Y)$. Then the relative entropy of a measure $\mu \in \mathcal{M}(X)$ is

$$h_\pi(\mu) = \lim_{n \rightarrow \infty} \frac{1}{n} H((\mathcal{P}_X)_0^{n-1} \mid \mathcal{Q}_{-\infty}^\infty)$$

To see that this limit exists, we again show that $H((\mathcal{P}_X)_0^{n-1} \mid \mathcal{Q}_{-\infty}^\infty)$ is a sub-

additive sequence. We see that

$$\begin{aligned} H((\mathcal{P}_X)_{-n-m+1}^0 \mid \mathcal{Q}_{-\infty}^\infty) &= H((\mathcal{P}_X)_{-n+1}^0 \vee (\mathcal{P}_X)_{-n-m+1}^{-n} \mid \mathcal{Q}_{-\infty}^\infty) \\ &= H((\mathcal{P}_X)_{-n+1}^0 \mid \mathcal{Q}_{-\infty}^\infty \vee (\mathcal{P}_X)_{-n-m+1}^{-n}) + H((\mathcal{P}_X)_{-n-m+1}^{-n} \mid \mathcal{Q}_{-\infty}^\infty) \\ &\leq H((\mathcal{P}_X)_{-n+1}^0 \mid \mathcal{Q}_{-\infty}^\infty) + H((\mathcal{P}_X)_{-m+1}^0 \mid \mathcal{Q}_{-\infty}^\infty) \end{aligned}$$

So the sequence is sub-additive and

$$h_\pi(\mu) = \inf_n H((\mathcal{P}_X)_{-n+1}^0 \mid \mathcal{Q}_{-\infty}^\infty)$$

The next theorem gives a relationship between the relative entropy and the regular entropy of a measure.

Theorem 3.1.1. *Let $\mu \in \mathcal{M}(X)$. We have the following identity.*

$$h_T(\mu) = h_S(\mu \circ \pi^{-1}) + h_\pi(\mu)$$

Proof. Observe that

$$H((\mathcal{P}_X)_{-n+1}^0) = H(\mathcal{Q}_{-n+1}^0) + H((\mathcal{P}_X)_{-n+1}^0 \mid \mathcal{Q}_{-n+1}^0)$$

From this we see that

$$H((\mathcal{P}_X)_{-n+1}^0) \geq H(\mathcal{Q}_{-n+1}^0) + H((\mathcal{P}_X)_{-n+1}^0 \mid \mathcal{Q}_{-\infty}^\infty)$$

Thus, after dividing by n and taking limits, we see that

$$h_T(\mu) \geq h_S(\mu \circ \pi^{-1}) + h_\pi(\mu)$$

To show the opposite inequality, we first note that

$$H((\mathcal{P}_X)_{-nm+1}^0) = H\left(\bigvee_{i=0}^{n-1} T^{-im}(\mathcal{P}_X)_{-m+1}^0\right) \quad (3.1.1)$$

Next, we need to show that

$$\lim_{j,k \rightarrow \infty} H((\mathcal{P}_X)_{-m+1}^0 \mid \mathcal{Q}_{-k}^j) = H((\mathcal{P}_X)_{-nm+1}^0) \quad (3.1.2)$$

Let $\epsilon > 0$ be given. By the increasing Martingale convergence theorem, we know that there exists j_0 such that for all $j > j_0$

$$H((\mathcal{P}_X)_{-m+1}^0 \mid \mathcal{Q}_{-\infty}^j) < H((\mathcal{P}_X)_{-m+1}^0 \mid \mathcal{Q}_{-\infty}^\infty) + \frac{\epsilon}{2}$$

Similarly there is a k_0 such that for $k > k_0$

$$H((\mathcal{P}_X)_{-m+1}^0 \mid \mathcal{Q}_{-k}^j) < H((\mathcal{P}_X)_{-m+1}^0 \mid \mathcal{Q}_{-\infty}^j) + \frac{\epsilon}{2}$$

Combining these proves claim (3.1.2).

We are now ready to prove the inequality. From (3.1.1) we can say

$$\begin{aligned} H((\mathcal{P}_X)_{-nm+1}^0) &= H(\mathcal{Q}_{-nm+1}^0) + H\left(\bigvee_{i=0}^{n-1} T^{-im}(\mathcal{P}_X)_{-m+1}^0 \mid \mathcal{Q}_{-nm+1}^0\right) \\ &\leq H(\mathcal{Q}_{-nm+1}^0) + H((\mathcal{P}_X)_{-nm+1}^0 \mid \mathcal{Q}_{-nm+1}^0) + H(T^{-m}(\mathcal{P}_X)_{-m+1}^0 \mid \mathcal{Q}_{-nm+1}^0) \\ &\quad + \dots + H(T^{-(n-1)m}(\mathcal{P}_X)_{-m+1}^0 \mid \mathcal{Q}_{-nm+1}^0) \\ &= H(\mathcal{Q}_{-nm+1}^0) + \sum_{i=0}^{n-1} H(T^{-im}(\mathcal{P}_X)_{-m+1}^0 \mid \mathcal{Q}_{-nm+1}^0) \\ &= H(\mathcal{Q}_{-nm+1}^0) + \sum_{i=0}^{n-1} H((\mathcal{P}_X)_{-m+1}^0 \mid \mathcal{Q}_{-(n-i)m+1}^{im}) \end{aligned}$$

From (3.1.2) we know that there is a K such that $j, k > K$ implies that

$$H((\mathcal{P}_X)_{-m+1}^0 \mid \mathcal{Q}_{-k}^j) < H((\mathcal{P}_X)_{-m+1}^0 \mid \mathcal{Q}_{-\infty}^\infty) + \epsilon$$

If $m > K + 1$ then $im > K$ and $(n-i)m - 1 > K$ for all $0 \leq i \leq n-1$. So continuing the expression above we have

$$\begin{aligned} H(\mathcal{Q}_{-nm+1}^0) &+ \sum_{i=0}^{n-1} H((\mathcal{P}_X)_{-m+1}^0 \mid \mathcal{Q}_{-(n-i)m+1}^{im}) \\ &\leq H(\mathcal{Q}_{-nm+1}^0) + \sum_{i=0}^{n-1} H((\mathcal{P}_X)_{-m+1}^0 \mid \mathcal{Q}_{-\infty}^\infty) + n\epsilon \\ &= H(\mathcal{Q}_{-nm+1}^0) + nH((\mathcal{P}_X)_{-m+1}^0 \mid \mathcal{Q}_{-\infty}^\infty) + n\epsilon \end{aligned}$$

Dividing by nm and taking the limit as $n \rightarrow \infty$ gives us

$$h_T(\mu) \leq h_S(\mu \circ \pi^{-1}) + \frac{1}{m} H((\mathcal{P}_X)_{-m+1}^0 \mid \mathcal{Q}_{-\infty}^\infty)$$

Now taking limits in m establishes the inequality. Thus the identity is shown. \square

For SFT's, the three quantities in Theorem 3.1.1 will be finite, so we will usually take

$$h_\pi(\mu) = h_T(\mu) - h_S(\mu \circ \pi^{-1})$$

Following a proof similar to the one given in section 2.3 for Theorem 2.3.3, one can show that the relative entropy map is upper semi-continuous for shifts of finite type.

Theorem 3.1.2. *For X and SFT and Y a factor of X by a block code π , the relative entropy map $h_\pi(\mu)$ is upper semi-continuous.*

Proof. Let μ_k be a sequence in $\mathcal{M}(X)$ such that $\mu_k \rightarrow \mu$. Then for any k and $n \geq 1$ we have

$$h_\pi(\mu_k) \leq \frac{1}{n} H_{\mu_k}((\mathcal{P}_X)_{-n+1}^0 \mid \mathcal{Q}_{-\infty}^\infty)$$

Next we see that, for $i > 0$ fixed

$$H_{\mu_k}((\mathcal{P}_X)_{-n+1}^0 \mid \mathcal{Q}_{-\infty}^\infty) \leq H_{\mu_k}((\mathcal{P}_X)_{-n+1}^0 \mid \mathcal{Q}_{-i}^i)$$

Which allows us to say

$$\begin{aligned} \lim_{k \rightarrow \infty} H_{\mu_k}((\mathcal{P}_X)_{-n+1}^0 \mid \mathcal{Q}_{-\infty}^\infty) &\leq \lim_{k \rightarrow \infty} H_{\mu_k}((\mathcal{P}_X)_{-n+1}^0 \mid \mathcal{Q}_{-i}^i) \\ &= H_\mu((\mathcal{P}_X)_{-n+1}^0 \mid \mathcal{Q}_{-i}^i) \end{aligned}$$

The last equality is justified by the continuity of $-x \log x$ and the fact that both $(\mathcal{P}_X)_{-n+1}^0$ and \mathcal{Q}_{-i}^i are finite. Now taking the limit as $i \rightarrow \infty$ gives us

$$\lim_{k \rightarrow \infty} H_{\mu_k}((\mathcal{P}_X)_0^{n-1} \mid \mathcal{Q}_{-\infty}^\infty) \leq H_\mu((\mathcal{P}_X)_0^{n-1} \mid \mathcal{Q}_{-\infty}^\infty)$$

Combining these two expressions and taking lim sup gives us

$$\begin{aligned} \limsup_{k \rightarrow \infty} h_\pi(\mu_k) &\leq \limsup_{k \rightarrow \infty} H_{\mu_k} \left((\mathcal{P}_X)^0_{-n+1} \mid \mathcal{Q}_{-\infty}^\infty \right) \\ &\leq \frac{1}{n} H_\mu \left((\mathcal{P}_X)^0_{-n+1} \mid \mathcal{Q}_{-\infty}^\infty \right) \end{aligned}$$

Finally, taking limits as $n \rightarrow \infty$ yields $\limsup_{k \rightarrow \infty} h_\pi(\mu_k) \leq h_\pi(\mu)$. \square

We are now ready to define the relative pressure. This function will be similar to the pressure we have already defined, but will be restricted to a single fiber over Y .

Definition 3.1.2. The relative pressure of a function f over a point $y \in Y$ is defined as

$$\mathcal{P}_\pi(f)(y) = \limsup_{n \rightarrow \infty} \frac{1}{n} \log \left(\sum_{\substack{w \in W_n \\ [w] \cap \pi^{-1}(y) \neq \emptyset}} e^{(S_n f)[w]} \right)$$

where

$$(S_n f)[w] = \sup_{\substack{x \in \pi^{-1}(y) \\ x \in [w]}} \sum_{i=0}^{n-1} f(T^i x)$$

In [9] it was shown that $\mathcal{P}_\pi(f)$ satisfies the following relative variational principle.

Theorem 3.1.3.

$$\int \mathcal{P}_\pi(f) d\nu = \sup_{\substack{\mu \in \mathcal{M}(X) \\ \mu \circ \pi^{-1} = \nu}} \left\{ h_\pi(\mu) + \int f d\mu \right\}$$

It is important to note that both relative entropy and pressure are generalizations of their non-relative versions. In both cases, taking Y to be the space consisting of a single point gives the non-relative version of the function.

3.2 Maximal Relative Pressure

We now define a maximal average relative pressure, which will be useful in the discussion of compensation functions. This function will be defined in terms of the relative variational principle.

Definition 3.2.1. For any $f \in C(X)$, we define the maximal relative pressure of f , with respect to π , as

$$W(f, \pi) = \sup_{\nu \in \mathcal{M}(Y)} \left\{ \int \mathcal{P}_\pi(f) d\nu \right\}$$

When the map π is clear, we will often drop it from the notation. By the relative variational principle, we see that

$$W(f, \pi) = \sup_{\mu \in \mathcal{M}(X)} \left\{ h_\pi(\mu) + \int f d\mu \right\}$$

As before, the upper semi-continuity of $h_\pi(\mu)$ tells us that there is at least one measure which attains this supremum.

We have defined the maximal relative pressure in terms of a variational principle, so it clearly has this in common with our earlier pressure function. We now show that it shares a number of other properties with \mathcal{P} .

Lemma 3.2.1. *i. If $f \leq g$ then $W(f) \leq W(g)$.*

ii. W is continuous.

iii. W is convex.

iv. $W(f + c) = W(f) + c \forall x \in \mathbb{R}$

v. $W(f + g \circ T - g) = W(f)$

Proof. (i) We have seen that there is at least one measure μ_0 which attains the supremum in $W(f)$. For such a measure,

$$\begin{aligned} W(f) &= h_\pi(\mu_0) + \int f d\mu_0 \\ &\leq h_\pi(\mu_0) + \int g d\mu_0 \\ &\leq W(g) \end{aligned}$$

Thus $W(f) \leq W(g)$.

(ii) Let μ_0 be as in (i). Then we have

$$\begin{aligned} W(f) - W(g) &< h_\pi(\mu_0) + \int f d\mu_0 - \left(h_\pi(\mu_0) + \int g d\mu_0 \right) \\ &\leq \int (f - g) d\mu_0 \end{aligned}$$

Repeating this for $W(g) - W(f)$ gives us

$$W(g) - W(f) \geq \int (g - f) d\mu'_0$$

Combining these gives us $|W(f) - W(g)| \leq \|f - g\|$, which implies continuity of W .

(iii) The proof of this fact follows the same structure as the one given for 2.5.1

(iv) using the variational principle.

(iv) and (v) are clear from the definition of W .

□

In Theorem 3.1.2, we saw that relative entropy was upper semi-continuous. This fact allows us to rearrange the relative variational principle in a fashion similar to Theorem 2.5.3.

Theorem 3.2.2. *Let $\mu \in \mathcal{M}(X)$. Then*

$$h_\pi(\mu) = \inf_{g \in C(X)} \left\{ W(g) - \int g d\mu \right\}$$

The next theorem will allow us to relate the relative pressure to the non relative version.

Theorem 3.2.3. *Suppose that Y has the property that every ergodic member of $\mathcal{M}(Y)$ is an equilibrium state for some member of $C(Y)$. Then for each $f \in C(X)$*

$$W(f) = \sup_{\phi \in C(Y)} \{ \mathcal{P}_X(f + \phi \circ \pi) - \mathcal{P}_Y(\phi) \}$$

Proof. Let $Q(f) = \sup_{\phi \in C(Y)} \{ \mathcal{P}_X(f + \phi \circ \pi) - \mathcal{P}_Y(\phi) \}$. Let $\epsilon > 0$ and choose $\phi_\epsilon \in C(Y)$ such that

$$\mathcal{P}_X(f + \phi_\epsilon \circ \pi) - \mathcal{P}_Y(\phi_\epsilon) > Q(f) - \epsilon$$

Now choose μ_ϵ such that

$$h_X(\mu_\epsilon) + \int (f + \phi_\epsilon \circ \pi) d\mu_\epsilon > \mathcal{P}_X(f + \phi_\epsilon \circ \pi)$$

Defining $m_\epsilon = \mu_\epsilon \circ \pi^{-1}$, we have that

$$\mathcal{P}_Y(\phi_\epsilon) \geq h_Y(m_\epsilon) + \int \phi_\epsilon dm_\epsilon$$

Combining these gives us

$$\begin{aligned} \mathcal{P}_X(f + \phi_\epsilon \circ \pi) - \mathcal{P}_Y(\phi_\epsilon) &< h_X(\mu_\epsilon) + \int (f + \phi_\epsilon \circ \pi) d\mu_\epsilon - \left(h_Y(m_\epsilon) + \int \phi_\epsilon dm_\epsilon \right) \\ &= h_\pi(\mu_\epsilon) + \int f d\mu_\epsilon \end{aligned}$$

So $Q(f) < h_\pi(\mu_\epsilon) + \int f d\mu_\epsilon + \epsilon$. Thus $Q(f) \leq W(f)$.

We now show that $Q(f) \geq W(f)$. By the variational principle, we see that

$$\begin{aligned} \mathcal{P}_X(f + \phi \circ \pi) &= \sup_{\nu \in \mathcal{M}(Y)} \left[\sup_{\mu \circ \pi^{-1} = \nu} \left(h_X(\mu) + \int f d\mu + \int \phi d\nu \right) \right] \\ &= \sup_{\nu \in \mathcal{M}(Y)} \left[\sup_{\mu \circ \pi^{-1} = \nu} \left(h_X(\mu) + \int f d\mu \right) + \int \phi d\nu - h_Y(\nu) + h_Y(\nu) \right] \\ &= \sup_{\nu \in \mathcal{M}(Y)} \left[\sup_{\mu \circ \pi^{-1} = \nu} \left(h_\pi(\mu) + \int f d\mu \right) + \int \phi d\nu + h_Y(\nu) \right] \\ &= \sup_{\nu \in \mathcal{M}(Y)} \left[\int \mathcal{P}_\pi(f) d\nu + \int \phi d\nu + h_Y(\nu) \right] \end{aligned}$$

Here, the last equality is given by the relative variational principle.

We know that for every ϕ has an ergodic equilibrium state ν_ϕ . Furthermore, we have assumed that every ergodic measure is an equilibrium state ν_ϕ for some ϕ . Thus

$$\begin{aligned}
Q(f) &= \\
&\sup_{\phi \in C(Y)} \left(\sup_{\nu \in \mathcal{M}(Y)} \left[\int \mathcal{P}_\pi(f) d\nu + \int \phi d\nu + h_Y(\nu) \right] - \sup_{\nu \in \mathcal{M}(Y)} \left[h_Y(\nu) + \int \phi d\nu \right] \right) \\
&\geq \sup_{\phi \in C(Y)} \left(\int \mathcal{P}_\pi(f) d\nu_\phi + \int \phi d\nu_\phi + h_Y(\nu_\phi) - h_Y(\nu_\phi) + \int \phi d\nu_\phi \right) \\
&= \sup_{\phi \in C(Y)} \left(\int \mathcal{P}_\pi(f) d\nu_\phi \right) \\
&= \sup_{\nu \text{ ergodic}} \left(\int \mathcal{P}_\pi(f) d\nu \right) \\
&= W(f)
\end{aligned}$$

□

In order to better understand Theorem 3.2.3, let us observe under what conditions the inequality $Q(f) \leq W(f)$ is sharp. Suppose that ν is a measure which maximizes $W(f)$. Such a measure will be ergodic, and thus there exists a ϕ_0 such that $\nu = \nu_{\phi_0}$. Clearly we have that $Q(f) \geq \mathcal{P}_X(f + \phi_0 \circ \pi) - \mathcal{P}_Y(\phi_0)$. In addition we know that

$$\mathcal{P}_X(f + \phi_0 \circ \pi) \geq \int (\mathcal{P}_\pi(f) + \phi_0) d\nu_{\phi_0} + h(\nu_{\phi_0})$$

Thus

$$\begin{aligned}
Q(f) &\leq W(f) \\
&= \int \mathcal{P}_\pi(f) d\nu_{\phi_0} \\
&= \int \mathcal{P}_\pi(f) d\nu_{\phi_0} + \int \phi_0 d\nu_{\phi_0} + h(\nu_{\phi_0}) - \int \phi_0 d\nu_{\phi_0} - h(\nu_{\phi_0}) \\
&= \int \mathcal{P}_\pi(f) d\nu_{\phi_0} + \int \phi_0 d\nu_{\phi_0} + h(\nu_{\phi_0}) - \mathcal{P}_Y(\phi_0) \\
&\leq \mathcal{P}_X(f + \phi_0 \circ \pi) - \mathcal{P}_Y(\phi_0)
\end{aligned}$$

This shows us that ϕ which maximize $Q(f)$ have equilibrium states which maximize $W(f)$, and vice versa.

In [7] it is shown that X being a subshift is sufficient to ensure that the hypotheses of Theorem 3.2.3 are satisfied. Specifically, it is shown that every ergodic $\nu \in \mathcal{M}(Y)$ appears as the equilibrium state of some function $\phi \in C(Y)$.

The following definition is analogous to definition 2.5.2.

Definition 3.2.2. Let $f \in C(X)$. If $\mu_0 \in \mathcal{M}(X)$ is a measure such that

$$W(f) = h_\pi(\mu_0) + \int f d\mu_0$$

then μ_0 is called a relative equilibrium state of f . Furthermore, if $\mu_0 \circ \pi = \nu_0$ and

$$h_\pi(\mu_0) + \int f d\mu_0 = \sup_{\mu \circ \pi^{-1} = \nu_0} \left\{ h_\pi(\mu) + \int f d\mu \right\}$$

then μ_0 is called a relative equilibrium state of f over ν_0 .

A tangent functional to W at f is a signed Borel measure μ such that $W(f+g) - W(f) \geq \int g d\mu$. The following theorem will show that the tangent functionals to W at f are exactly the relative equilibrium states of f .

Theorem 3.2.4. *Let $f \in C(X)$. Then μ is a tangent functional to W at f iff $\mu \in \mathcal{M}(X)$ and $W(f) = h_\pi(\mu) + \int f d\mu$.*

Proof. We will use several of the facts from Theorem 3.2.1 in this proof. Let μ be a tangent functional. Let $g \in C(X)$ with $g \geq 0$. Then

$$\begin{aligned} - \int g d\mu &= \int (-g) d\mu \\ &\leq W(f-g) - W(f) \\ &\leq (W(f) - \inf g) - W(f) \\ &\leq 0 \end{aligned}$$

So μ is a nonnegative measure. For any real t , $\int t d\mu \leq W(f+t) - W(f) = t$. Thus $\mu(X) \leq 1$ and $-\mu(X) \leq -1$ so μ is a probability measure. Also

$$t \left(\int (g \circ T - g) d\mu \right) \leq W(f + (tg \circ T - tg)) - W(f) = 0$$

Thus $\int g \circ T d\mu = \int g d\mu$. Together, these show that $\mu \in \mathcal{M}(X)$.

We will now show that $h_\pi(\mu) + \int f d\mu = W(f)$. We know μ is a tangent functional so for any $g \in C(X)$

$$W(f + g) - W(f) \geq \int g d\mu + \int (f - f) d\mu = \int (g + f) d\mu - \int f d\mu$$

so

$$W(f + g) - \int (g + f) d\mu \geq W(f) - \int f d\mu$$

Setting $g = h - f$ we see that for any $h \in C(X)$

$$W(h) - \int h d\mu \geq W(f) - \int f d\mu$$

By Theorem 3.2.2 we have that $h_\pi(\mu) \geq W(f) - \int f d\mu$. Thus $h_\pi(\mu) + \int f \geq W(f)$. Clearly $W(f) \geq h_\pi(\mu) + \int f d\mu$ so equality is shown.

Now assume that μ is a relative equilibrium state of f . Then

$$W(f + g) - W(f) \geq h_\pi(\mu) + \int (f + g) d\mu - h_\pi(\mu) - \int f d\mu = \int g d\mu$$

So μ is a tangent functional to W at f . □

3.3 Compensation Functions

When dealing with a block code $\pi : X \rightarrow Y$ it is natural to wonder if there is a relationship between $\mathcal{P}_X(\phi \circ \pi)$ and $\mathcal{P}_Y(\phi)$ for a function $\phi \in C(Y)$. In one sense this question is easily answered. We know that $h(\mu) \geq h(\nu)$ for any $\mu \in \mathcal{M}(X)$ and $\nu \in \mathcal{M}(Y)$ such that μ pushes forward to ν . In fact, for some infinite-to-one codes this inequality will be strict. In addition, we know that $\int \phi \circ \pi d\mu = \int \phi d\nu$. So we

can say that

$$\begin{aligned}
\mathcal{P}_Y(\phi) &= \sup_{\nu \in \mathcal{M}(Y)} \left(h(\nu) + \int \phi d\nu \right) \\
&\leq \sup_{\nu \in \mathcal{M}(Y)} \left(\sup_{\mu \circ \pi^{-1} = \nu} (h(\mu) - h(\nu)) + h(\nu) + \int \phi d\nu \right) \\
&= \sup_{\nu \in \mathcal{M}(Y)} \sup_{\mu \circ \pi^{-1} = \nu} \left(h(\mu) + \int \phi \circ \pi d\mu \right) \\
&= \mathcal{P}_X(\phi \circ \pi)
\end{aligned}$$

This calculation reveals an important concept. Intuitively it seems like the difference between the pressures comes from the extra term introduced in the third line, namely $\sup_{\mu \circ \pi^{-1} = \nu} (h(\mu) - h(\nu))$. This term measures the amount of extra entropy that can occur in a fiber over ν . On both sides of the equation, the ν which attains the supremum is dependent on the ϕ we have chosen. However, we will soon see that the extra entropy term is not dependent on ϕ .

By the relative variational principle, we know that

$$\sup_{\mu \circ \pi^{-1} = \nu} (h(\mu) - h(\nu)) = \int \mathcal{P}_\pi(0) d\nu$$

From this we can show that

$$\sup_{\nu \in \mathcal{M}(Y)} \left(h(\nu) + \int (\phi + \mathcal{P}_\pi(0)) d\nu \right) = \sup_{\nu \in \mathcal{M}(Y)} \left(\sup_{\mu \circ \pi^{-1} = \nu} (h(\mu)) + \int \phi d\nu \right)$$

The function $\mathcal{P}_\pi(0)$ is not necessarily continuous, so applying the variational principle is not strictly valid. Still, if we could apply it to this expression we would arrive at

$$\mathcal{P}_X(\phi \circ \pi - \mathcal{P}_\pi(0) \circ \pi) = \mathcal{P}_Y(\phi)$$

So, if some continuous version of $-\mathcal{P}_\pi(0) \circ \pi$ exists then it is a candidate for a compensation function. This should convince us that if we add a function which exactly cancels out the extra entropy term, it will “compensate” for the difference in pressure. It seems incredible that a single function can perform this compensation simultaneously for all functions ϕ . In studying equilibrium states we saw that there is a close

relationship between ϕ and the measures that occur in pressure calculations. The reason this compensation can be accomplished is that the entropy we are cancelling out is a sort of topological entropy occurring in fibers over points in Y . In fact, for every invariant measure the following equality holds for a.e. $y \in Y$ [16].

$$\mathcal{P}_\pi(0)(y) = \limsup_{n \rightarrow \infty} \frac{1}{n} \log \sum_{\substack{w \in W_n \\ [w] \cap \pi^{-1}(y) \neq \emptyset}} 1 = \limsup_{n \rightarrow \infty} \frac{1}{n} \log |\pi^{-1}[y_0 y_1 \dots y_{n-1}]|$$

This shows us what Petersen and Boyle mean in [2] when they say that functions like $\mathcal{P}_\pi(0)$, which we will soon define to be compensation functions, are “a kind of oracle for how entropy can appear in a fiber.”

3.3.1 Existence of a Compensation Function

Definition 3.3.1. Let (X, T) and (Y, S) be subshifts and $\pi : X \rightarrow Y$ be a block code. A function $f \in C(X)$ is called a compensation function if for all $\phi \in C(Y)$

$$\mathcal{P}_X(f + \phi \circ \pi) = \mathcal{P}_Y(\phi)$$

We have already shown that $\mathcal{P}_\pi(0)$ may behave like a compensation function, but its usefulness is somewhat limited. In particular it is generally not a continuous function. In [22] Walters claims that for X and Y subshifts with a block code $\pi : X \rightarrow Y$, a compensation function exists in $C(X)$. There are two difficulties with Walters proof. The first is that it is nonconstructive, so although it guarantees the existence of a compensation function it does not provide us with a candidate. The other problem is that his construction relies on the existence of a structure which does not exist for all subshift factor triples. We will return to the discussion of this second issue later. For now, our next step will be to construct an entire family of compensation functions which are continuous and have a familiar structure, thus somewhat addressing the first problem.

Theorem 3.3.1. *Let (X, Y, π) be a factor triple with X and Y both subshifts. Let $\mu_0 \in \mathcal{M}(X)$ and $\nu_0 \in \mathcal{M}(Y)$ such that $\mu_0 \circ \pi^{-1} = \nu_0$. Assume I_{μ_0} and I_{ν_0} have continuous versions, and that $\mu_0(Q | (\mathcal{P}_X)_{-\infty}^{-1}) = \mu_0(Q | \mathcal{Q}_{-\infty}^{-1})$ for all $Q \in \mathcal{Q}$. Then $f = I_{\nu_0} \circ \pi - I_{\mu_0}$ is a compensation function for (X, Y, π) .*

To prove this theorem we will need another corollary to Jensen's inequality, which is analogous to Lemma 2.5.4.

Lemma 3.3.2. *Let $\mu \in \mathcal{M}(X)$ and $g \in C(X)$ such that $g \geq 0$. Let \mathcal{C} be a sub- σ -algebra of \mathcal{B} such that there exist a sequence of sub- σ -algebras $\mathcal{C}_n \rightarrow \mathcal{C}$. Then for any $h \in C(X)$ such that $\mathbb{E}_\mu(h \mid \mathcal{C}) = 1$,*

$$-\mathbb{E}_\mu(h \log(h/g) \mid \mathcal{C}) \leq \log \mathbb{E}_\mu(g \mid \mathcal{C})$$

Proof. Assume that \mathcal{C} is a finite partition of X , say $\mathcal{C} = \{C_1, \dots, C_n\}$. Then $\mathbb{E}_\mu(h \mid \mathcal{C}) = 1$ implies that

$$\int_{C_i} h d\mu = \mu(C_i)$$

We then have that, on C_i

$$\begin{aligned} -\mathbb{E}_\mu(h \log(h/g) \mid \mathcal{C}) &= -\frac{1}{\mu(C_i)} \int_{C_i} h \log(h/g) d\mu \\ &= \int_{C_i} (h/\mu(C_i)) \log(g/h) d\mu \end{aligned}$$

Using the concavity of \log and Jensen's inequality gives us

$$\int_{C_i} (h/\mu(C_i)) \log(g/h) d\mu \leq \log \left(\frac{1}{\mu(C_i)} \int_{C_i} g d\mu \right) \quad (3.3.1)$$

With equality exactly when g/h is a \mathcal{C} -measurable function. To finish, we note that the right hand side of the above equation is equal to $\log \mathbb{E}_\mu(g \mid \mathcal{C})$ μ -a.e.

Now let \mathcal{C}_n be a sequence of finite sub- σ -algebras which converge to \mathcal{C} . If h is a function such that $\mathbb{E}_\mu(h \mid \mathcal{C}) = 1$ then we also have that $\mathbb{E}_\mu(h \mid \mathcal{C}_n) = 1$ for all n . From (3.3.1) we have that for every n ,

$$-\mathbb{E}_\mu(h \log(h/g) \mid \mathcal{C}_n) \leq \log \mathbb{E}_\mu(g \mid \mathcal{C}_n)$$

Also, by the increasing Martingale convergence theorem, the left and right hand sides of this converge to $-\mathbb{E}_\mu(h \log(h/g) \mid \mathcal{C})$ and $\log \mathbb{E}_\mu(g \mid \mathcal{C}_n)$. So the corollary is proved in general. \square

We may now prove that f as defined in Theorem 3.3.1 is a compensation function.

Proof of Theorem 3.3.1. Fix $\nu \in \mathcal{M}(Y)$. We see that

$$\sup_{\mu \circ \pi^{-1} = \nu} \left(h_X(\mu) + \int (I_{\nu_0} \circ \pi - I_{\mu_0} + \phi \circ \pi) d\mu \right) \quad (3.3.2)$$

$$= \int (I_{\nu_0} + \phi) d\nu + \sup_{\mu \circ \pi^{-1} = \nu} \left(h_X(\mu) - \int I_{\mu_0} d\mu \right) \quad (3.3.3)$$

So if we can show that

$$\sup_{\mu \circ \pi^{-1} = \nu} \left(h_X(\mu) - \int I_{\mu_0} d\mu \right) = h_Y(\nu) - \int I_{\nu_0} d\nu \quad (3.3.4)$$

then taking the supremum over $\nu \in \mathcal{M}(Y)$ in (3.3.3) gives us

$$\begin{aligned} & \sup_{\mu \in \mathcal{M}(X)} \left(h_X(\mu) + \int (I_{\nu_0} \circ \pi - I_{\mu_0} + \phi \circ \pi) d\mu \right) \\ &= \sup_{\nu \in \mathcal{M}(Y)} \left(\int (I_{\nu_0} + \phi) d\nu + h_Y(\nu) - \int I_{\nu_0} d\nu \right) \\ &= \sup_{\nu \in \mathcal{M}(Y)} \left(h_Y(\nu) + \int \phi d\nu \right) \end{aligned}$$

This will establish that f is a compensation function.

To show (3.3.4) we begin by seeking to maximize the term $h_X(\mu) + \int (-I_{\mu_0}) d\mu = \int (I_{\mu} - I_{\mu_0}) d\mu$ over μ such that $\mu \circ \pi^{-1} = \nu$. For simplicity in the following expressions, we drop the X subscript from the state partition on X and simply call it \mathcal{P} . We have

$$\begin{aligned}
& \int (I_\mu - I_{\mu_0}) d\mu \\
&= \int \sum_{P \in \mathcal{P}} -\chi_P \log \left(\frac{\mu(P | \mathcal{P}_{-\infty}^{-1})}{\mu_0(P | \mathcal{P}_{-\infty}^{-1})} \right) d\mu \\
&= \int \sum_{Q \in \mathcal{Q}} \sum_{P \subset Q} -\mu(P | \mathcal{P}_{-\infty}^{-1}) \log \left(\frac{\mu(P | \mathcal{P}_{-\infty}^{-1})}{\mu_0(P | \mathcal{P}_{-\infty}^{-1})} \right) d\mu \\
&= - \int \sum_{Q \in \mathcal{Q}} \mu(Q | \mathcal{P}_{-\infty}^{-1}) \sum_{P \subset Q} \left[\frac{\mu(P | \mathcal{P}_{-\infty}^{-1})}{\mu(Q | \mathcal{P}_{-\infty}^{-1})} \log \left(\frac{\mu(P | \mathcal{P}_{-\infty}^{-1}) / \mu(Q | \mathcal{P}_{-\infty}^{-1})}{\mu_0(P | \mathcal{P}_{-\infty}^{-1}) / \mu_0(Q | \mathcal{P}_{-\infty}^{-1})} \right) \right] d\mu \\
&\quad - \int \sum_{Q \in \mathcal{Q}} \mu(Q | \mathcal{P}_{-\infty}^{-1}) \sum_{P \subset Q} \left[\frac{\mu(P | \mathcal{P}_{-\infty}^{-1})}{\mu(Q | \mathcal{P}_{-\infty}^{-1})} \log \left(\frac{\mu(Q | \mathcal{P}_{-\infty}^{-1})}{\mu_0(Q | \mathcal{P}_{-\infty}^{-1})} \right) \right] d\mu \\
&= - \int \sum_{Q \in \mathcal{Q}} \mu(Q | \mathcal{P}_{-\infty}^{-1}) \sum_{P \subset Q} \left[\frac{\mu(P | \mathcal{P}_{-\infty}^{-1})}{\mu(Q | \mathcal{P}_{-\infty}^{-1})} \log \left(\frac{\mu(P | \mathcal{P}_{-\infty}^{-1}) / \mu(Q | \mathcal{P}_{-\infty}^{-1})}{\mu_0(P | \mathcal{P}_{-\infty}^{-1}) / \mu_0(Q | \mathcal{P}_{-\infty}^{-1})} \right) \right] d\mu \\
&\quad - \int \sum_{Q \in \mathcal{Q}} \mu(Q | \mathcal{P}_{-\infty}^{-1}) \log \left(\frac{\mu(Q | \mathcal{P}_{-\infty}^{-1})}{\mu_0(Q | \mathcal{P}_{-\infty}^{-1})} \right) d\mu
\end{aligned}$$

Now we apply corollary 2.5.4 to the sum over $P \subset Q$ in the first term.

$$\sum_{P \subset Q} \frac{\mu(P | \mathcal{P}_{-\infty}^{-1})}{\mu(Q | \mathcal{P}_{-\infty}^{-1})} \log \left(\frac{\mu(P | \mathcal{P}_{-\infty}^{-1}) / \mu(Q | \mathcal{P}_{-\infty}^{-1})}{\mu_0(P | \mathcal{P}_{-\infty}^{-1}) / \mu_0(Q | \mathcal{P}_{-\infty}^{-1})} \right) \leq \log \left(\sum_{P \subset Q} \frac{\mu_0(P | \mathcal{P}_{-\infty}^{-1})}{\mu_0(Q | \mathcal{P}_{-\infty}^{-1})} \right) = 0$$

With equality exactly when $\mu_0(P | \mathcal{P}_{-\infty}^{-1}) / \mu_0(Q | \mathcal{P}_{-\infty}^{-1}) = \mu(P | \mathcal{P}_{-\infty}^{-1}) / \mu(Q | \mathcal{P}_{-\infty}^{-1})$.

Thus the first term is nonpositive. This means we have shown that

$$\int (I_\mu - I_{\mu_0}) d\mu \leq - \int \sum_{Q \in \mathcal{Q}} \mu(Q | \mathcal{P}_{-\infty}^{-1}) \log \left(\frac{\mu(Q | \mathcal{P}_{-\infty}^{-1})}{\mu_0(Q | \mathcal{P}_{-\infty}^{-1})} \right) d\mu$$

We must now show that

$$- \int \sum_{Q \in \mathcal{Q}} \mu(Q | \mathcal{P}_{-\infty}^{-1}) \log \left(\frac{\mu(Q | \mathcal{P}_{-\infty}^{-1})}{\mu_0(Q | \mathcal{P}_{-\infty}^{-1})} \right) d\mu \leq h_Y(\nu) - \int I_{\nu_0} d\nu$$

We have

$$\begin{aligned}
& - \int \sum_{Q \in \mathcal{Q}} \mu(Q | \mathcal{P}_{-\infty}^{-1}) \log \left(\frac{\mu(Q | \mathcal{P}_{-\infty}^{-1})}{\mu_0(Q | \mathcal{P}_{-\infty}^{-1})} \right) d\mu \\
&= - \int \sum_{Q \in \mathcal{Q}} \left[\frac{\mu(Q | \mathcal{P}_{-\infty}^{-1})}{\mu(Q | \mathcal{Q}_{-\infty}^{-1})} \mu(Q | \mathcal{Q}_{-\infty}^{-1}) \log \left(\frac{\mu(Q | \mathcal{Q}_{-\infty}^{-1})}{\mu_0(Q | \mathcal{Q}_{-\infty}^{-1})} \right) \right] d\mu \\
& - \int \sum_{Q \in \mathcal{Q}} \left[\frac{\mu(Q | \mathcal{P}_{-\infty}^{-1})}{\mu(Q | \mathcal{Q}_{-\infty}^{-1})} \mu(Q | \mathcal{Q}_{-\infty}^{-1}) \log \left(\frac{\mu(Q | \mathcal{P}_{-\infty}^{-1})/\mu(Q | \mathcal{Q}_{-\infty}^{-1})}{\mu_0(Q | \mathcal{P}_{-\infty}^{-1})/\mu_0(Q | \mathcal{Q}_{-\infty}^{-1})} \right) \right] d\mu \\
&= - \int \sum_{Q \in \mathcal{Q}} \mu(Q | \mathcal{Q}_{-\infty}^{-1}) \log \left(\frac{\mu(Q | \mathcal{Q}_{-\infty}^{-1})}{\mu_0(Q | \mathcal{Q}_{-\infty}^{-1})} \right) d\mu \\
& - \int \sum_{Q \in \mathcal{Q}} \left[\mu(Q | \mathcal{Q}_{-\infty}^{-1}) \frac{\mu(Q | \mathcal{P}_{-\infty}^{-1})}{\mu(Q | \mathcal{Q}_{-\infty}^{-1})} \log \left(\frac{\mu(Q | \mathcal{P}_{-\infty}^{-1})/\mu(Q | \mathcal{Q}_{-\infty}^{-1})}{\mu_0(Q | \mathcal{P}_{-\infty}^{-1})/\mu_0(Q | \mathcal{Q}_{-\infty}^{-1})} \right) \right] d\mu
\end{aligned}$$

The first term given here is exactly $h_Y(\nu) - \int I_{\nu_0} d\nu$. To analyze the second term, we note that $\mathbb{E}_\mu(\mu(Q | \mathcal{P}_{-\infty}^{-1})/\mu(Q | \mathcal{Q}_{-\infty}^{-1}) | \mathcal{Q}_{-\infty}^{-1}) = 1$ and $\mu_0(Q | \mathcal{P}_{-\infty}^{-1})/\mu_0(Q | \mathcal{Q}_{-\infty}^{-1}) \geq 0$. So taking these as h and g in corollary 3.3.2 gives us

$$\begin{aligned}
& - \int \sum_{Q \in \mathcal{Q}} \left[\mu(Q | \mathcal{Q}_{-\infty}^{-1}) \frac{\mu(Q | \mathcal{P}_{-\infty}^{-1})}{\mu(Q | \mathcal{Q}_{-\infty}^{-1})} \log \left(\frac{\mu(Q | \mathcal{P}_{-\infty}^{-1})/\mu(Q | \mathcal{Q}_{-\infty}^{-1})}{\mu_0(Q | \mathcal{P}_{-\infty}^{-1})/\mu_0(Q | \mathcal{Q}_{-\infty}^{-1})} \right) \right] d\mu \\
&= - \int \sum_{Q \in \mathcal{Q}} \left[\mu(Q | \mathcal{Q}_{-\infty}^{-1}) \mathbb{E}_\mu \left(\frac{\mu(Q | \mathcal{P}_{-\infty}^{-1})}{\mu(Q | \mathcal{Q}_{-\infty}^{-1})} \log \left(\frac{\mu(Q | \mathcal{P}_{-\infty}^{-1})/\mu(Q | \mathcal{Q}_{-\infty}^{-1})}{\mu_0(Q | \mathcal{P}_{-\infty}^{-1})/\mu_0(Q | \mathcal{Q}_{-\infty}^{-1})} \right) \middle| \mathcal{Q}_{-\infty}^{-1} \right) \right] d\mu \\
&\leq - \int \sum_{Q \in \mathcal{Q}} \left[\mu(Q | \mathcal{Q}_{-\infty}^{-1}) \log \mathbb{E}_\mu \left(\frac{\mu_0(Q | \mathcal{P}_{-\infty}^{-1})}{\mu_0(Q | \mathcal{Q}_{-\infty}^{-1})} \middle| \mathcal{Q}_{-\infty}^{-1} \right) \right] d\mu
\end{aligned}$$

By assumption, $\mu_0(Q | \mathcal{P}_{-\infty}^{-1}) = \mu_0(Q | \mathcal{Q}_{-\infty}^{-1})$, so we have shown that this term is nonpositive. \square

In [22] Walters gives the following generalization of Israel's result on ergodic measures as equilibrium states.

Lemma 3.3.3. *Suppose $f \in C(X)$ is given. If $\nu \in \mathcal{M}(Y)$ is ergodic then there is some $\mu \in \mathcal{M}(X)$ with $\mu \circ \pi^{-1} = \nu$ which is an equilibrium state of $f + \phi \circ \pi$ for some $\phi \in C(Y)$.*

This fact will allow us to prove a fundamental relationship between the equilibrium

states of $f + \phi \circ \pi$ and ϕ . First we will need a corollary to Lemma 3.3.3.

Corollary 3.3.4. *If $f \in C(X)$ is a compensation function for a factor triple (X, Y, π) with X and Y subshifts, then for all $\nu \in \mathcal{M}(Y)$*

$$\int \mathcal{P}_\pi(f) d\nu = 0$$

Proof. By the relative variational principle, we know that

$$\begin{aligned} \int \mathcal{P}_\pi(f) d\nu_0 &= \sup_{\mu \circ \pi^{-1} = \nu_0} \left(h_\pi(\mu) + \int f d\mu \right) \\ &\leq \sup_{\nu} \sup_{\mu \circ \pi^{-1} = \nu} \left(h_\pi(\mu) + \int f d\mu \right) = W(f) \\ &= \sup_{\phi \in C(Y)} (\mathcal{P}_X(f + \phi \circ \pi) - \mathcal{P}_Y(\phi)) && \text{by thm 3.2.3} \\ &= 0 && \text{because } f \text{ is a compensation function} \end{aligned}$$

Now for any ergodic measure $\nu_0 \in \mathcal{M}(Y)$ Lemma 3.3.3 tells us that there exists $\mu_0 \in \mathcal{M}(X)$ and $\phi_0 \in C(Y)$ such that μ_0 is an equilibrium state of $f + \phi_0 \circ \pi$ and $\mu_0 \circ \pi^{-1} = \nu_0$. Thus

$$\begin{aligned} \mathcal{P}_X(f + \phi_0 \circ \pi) &= h(\mu_0) + \int f d\mu_0 + \int \phi_0 d\nu_0 \\ &= \sup_{\nu} \left(h(\nu) + \int \phi_0 d\nu \right) \text{ because } f \text{ is a compensation function} \\ &\geq h(\nu_0) + \int \phi_0 d\nu_0 \end{aligned}$$

From this we see that $h_\pi(\mu_0) + \int f d\mu_0 \geq 0$ so we have

$$\begin{aligned} \int \mathcal{P}_\pi(f) d\nu_0 &= \sup_{\mu \circ \pi^{-1} = \nu_0} \left(h_\pi(\mu) + \int f d\mu \right) \\ &\geq h_\pi(\mu_0) + \int f d\mu_0 \\ &\geq 0 \end{aligned}$$

If ν is not ergodic, let $\nu = \int m d\tau(m)$ be its ergodic decomposition. Then

$$\int \mathcal{P}_\pi(f) d\nu = \int \left(\int \mathcal{P}_\pi(f) dm \right) d\tau(m) \geq 0$$

□

Theorem 3.3.5. *Let f be a compensation function, $\mu_0 \in \mathcal{M}(X)$ and $\phi \in C(Y)$. Then μ_0 is an equilibrium state of $f + \phi \circ \pi$ iff $\mu_0 \circ \pi^{-1}$ is the equilibrium state of ϕ and μ_0 is a relative equilibrium state of f over $\mu_0 \circ \pi^{-1}$.*

Proof. Let $\nu_0 = \mu_0 \circ \pi^{-1}$. If μ_0 is an equilibrium state of $f + \phi \circ \pi$ then

$$\begin{aligned}
\mathcal{P}_Y(\phi) &= \mathcal{P}_X(f + \phi \circ \pi) \\
&= h(\mu_0) + \int f d\mu_0 + \int \phi d\nu_0 \\
&= \sup_{\mu \circ \pi^{-1} = \nu_0} \left(h(\mu) + \int f d\mu + \int \phi d\nu \right) \\
&= h(\nu_0) + \int \mathcal{P}_\pi(f) d\nu_0 + \int \phi d\nu_0 \\
&= h(\mu_0 \circ \pi^{-1}) + \int \phi \circ \pi d\mu_0 \text{ by corollary 3.3.4}
\end{aligned}$$

So $\mu_0 \circ \pi^{-1}$ is an equilibrium state of ϕ . In addition, the first and fourth lines of the equality above show that μ_0 is a relative equilibrium state of f over $\mu_0 \circ \pi^{-1}$.

Now if we assume that $\mu_0 \circ \pi^{-1}$ is an equilibrium state of ϕ and μ_0 is a relative equilibrium state of f over $\mu_0 \circ \pi^{-1}$ then

$$\begin{aligned}
&h(\mu_0) + \int f d\mu_0 + \int \phi \circ \pi d\mu_0 \\
&= h_\pi(\mu_0) + \int f d\mu_0 + h(\nu_0) + \int \phi d\nu_0 \\
&= \int \mathcal{P}_\pi(f) d\nu_0 + \mathcal{P}_Y(\phi) \\
&= \mathcal{P}_Y(\phi) = \mathcal{P}_X(f + \phi \circ \pi)
\end{aligned}$$

Thus μ_0 is an equilibrium state of $f + \phi \circ \pi$. □

3.4 Walters functions and Compensation Functions

Next we will attempt to give conditions which ensure that the hypotheses of Theorem 3.3.1 are satisfied. First we will require some definitions. Two functions $f, g \in C(X)$

are said to be *cohomologous* if there exists a function $h \in C(X)$ such that

$$f(x) = g(x) + h \circ T(x) - h(x)$$

For $x \in X$ let $x^+ = (x_0, x_1, \dots)$. If $a = (a_{-n}, \dots, a_{-1})$ and $a_{-n} \dots a_{-1}x_0$ is an allowable word then let $ax^+ = (a_{-n} \dots a_{-1}x_0x_1 \dots)$. For f_0 a one sided function on X (in other words a function which only depends on x^+) we define

$$C_{f_0}(x^+, z^+) = \sup_{n \geq 1} \left(\sup_{ax^+ \text{ allowable}} \left[\sum_{i=1}^n (f_0(a_{-i} \dots x^+) - f_0(a_{-i} \dots z^+)) \right] \right)$$

Of course, this is only well defined if $x_0 = z_0$.

Definition 3.4.1. If a function $f \in C(X)$ is cohomologous to a one sided function f_0 such that

- i. $C_{f_0}(x^+, z^+)$ exists when $x_0 = z_0$.
- ii. $C_{f_0}(x^+, z^+)$ is bounded above by a constant
- iii. $C_{f_0}(x^+, z^+) \rightarrow 0$ as $d(x^+, z^+) \rightarrow 0$

then we say that f is a Walters function. We denote the set of Walters functions by $\mathcal{W}(X)$.

It is fairly easy to see that $\mathcal{W}(X)$ contains all functions which depend on a finite set of coordinates of x . Thus, $\mathcal{W}(X)$ is dense in $C(X)$. It also contains many other familiar subsets of $C(X)$, such as the Hölder continuous functions.

The advantage of Walters functions is that much has been proved about their equilibrium states. We summarize those properties which are useful for our discussion in the following theorem.

Theorem 3.4.1. [20] *Let (X, T) be an irreducible shift of finite type and $f \in \mathcal{W}(X)$. Then f has a unique equilibrium state μ_f and I_{μ_f} has a continuous version which is a one-sided Walters function. Furthermore, I_{μ_f} is cohomologous to $\mathcal{P}(f) - f$.*

We can relate $\mathcal{W}(X)$ and $\mathcal{W}(Y)$ in the following manner

Lemma 3.4.2. *If $\phi \in \mathcal{W}(Y)$ then $\phi \circ \pi \in \mathcal{W}(X)$.*

We will carry out the proof assuming that π is a 1-block code. The general result is similar.

Proof. We know $\phi \in \mathcal{W}(Y)$ implies it is cohomologous to a one sided function ϕ_0 with $C_{\phi_0}(y^+, u^+) \rightarrow 0$ as $d(y^+, u^+) \rightarrow 0$. Say $\phi - \phi_0 = h \circ S - h$. Then

$$\phi \circ \pi - \phi_0 \circ \pi = h \circ S \circ \pi - h \circ \pi = h \circ \pi \circ T - h \circ \pi$$

So $\phi \circ \pi$ is cohomologous to $\phi_0 \circ \pi$. Say for each allowable ax^+ that $\pi(a^+) = by^+$. Then

$$\begin{aligned} C_{\phi_0 \circ \pi}(x^+, z^+) &= \sup_{n \geq 1} \left(\sup_{ax^+ \text{ allowable}} \left[\sum_{i=1}^n (\phi_0 \circ \pi(a_{-i} \dots x^+) - \phi_0 \circ \pi(a_{-i} \dots z^+)) \right] \right) \\ &= \sup_{n \geq 1} \left(\sup_{by^+ \text{ allowable}} \left[\sum_{i=1}^n (\phi_0(b_{-i} \dots y^+) - \phi_0(b_{-i} \dots u^+)) \right] \right) \\ &\leq C_{\phi_0}(y^+, u^+) \end{aligned}$$

Thus the boundedness and convergence of $C_{\phi_0}(y^+, u^+)$ give us that $\phi \circ \pi$ is in $\mathcal{W}(X)$. \square

Now suppose that there exists a compensation function which is in $\mathcal{W}(X)$. Let $\phi \in \mathcal{W}(Y)$. Lemma 3.4.2 tells us that $\phi \circ \pi \in \mathcal{W}(X)$ so $f + \phi \circ \pi$ is in $\mathcal{W}(X)$ as well. By Theorem 3.3.5 we know that if ν_ϕ is the equilibrium state of ϕ then there exists μ which is a relative equilibrium state of f over ν_ϕ , and $\mu = \mu_{f+\phi \circ \pi}$ is an equilibrium state of $f + \phi \circ \pi$. In addition, because of Theorem 3.4.1, ν_ϕ and μ_ϕ are unique and their information cocycles have continuous versions. This is one of the hypotheses we need in order to apply Theorem 3.3.1.

In fact, compensation functions of the type given in Theorem 3.3.1 will exist if we have a compensation function in $\mathcal{W}(X)$. Theorem 3.4.1 tells us that I_{ν_ϕ} and $I_{\mu_{f+\phi \circ \pi}}$ are cohomologous to $\mathcal{P}_Y(\phi) - \phi$ and $\mathcal{P}_X(f + \phi \circ \pi) - (f + \phi \circ \pi)$ respectively. In other words, $I_{\nu_\phi} \circ \pi - I_{\mu_{f+\phi \circ \pi}}$ is cohomologous to $\mathcal{P}_Y(\phi) - \phi - \mathcal{P}_X(f + \phi \circ \pi) + (f + \phi \circ \pi)$ which is cohomologous to f . Given that f is a compensation function, we see that $I_{\nu_\phi} \circ \pi - I_{\mu_{f+\phi \circ \pi}}$ is also a compensation function. Thus, if we have a compensation function in $\mathcal{W}(X)$ we can build many more compensation functions. In the next section, we will give an explicit case in which we can construct a compensation function in $\mathcal{W}(X)$ without prior access to such a compensation function.

3.4.1 Constructing compensation functions in $\mathcal{W}(X)$

The first appearance of compensation functions was in [3]. In that work, they were employed to show that for a factor triple (X, Y, π) with X and Y shifts of finite type, if some Markov measure $\nu \in \mathcal{M}(Y)$ lifts to a Markov measure $\mu \in \mathcal{M}(X)$ then in fact every Markov measure on Y lifts to a Markov measure. The central idea of the proof is to construct a compensation function based on the information cocycles of ν and μ . Let us attempt to describe how this is carried out.

We have seen that the unique equilibrium state of a function of two coordinates is a Markov measure in Theorem 2.5.6. As one might expect, it is also true that every Markov measure appears as the equilibrium state of some function of two coordinates. To see this, let m be a Markov measure and consider the function $f(x_0, x_1) = \log m([x_0x_1])$. Applying Theorem 2.5.6 to this function returns the measure m . In fact, this function is exactly $-I_m$.

Now suppose that we know that some Markov measure $\nu \in \mathcal{M}(Y)$ lifts to a Markov measure $\mu \in \mathcal{M}(X)$. Furthermore, suppose we know that $I_\nu \circ \pi - I_\mu$ is a compensation function. If p is another Markov measure on Y we know that it is the unique equilibrium state of $-I_p$. We also know that $I_\nu \circ \pi - I_\mu - I_p \circ \pi$ is a function of two coordinates, and thus has a unique equilibrium state m which is also a Markov measure. Now observe that $I_\nu \circ \pi - I_\mu - I_p \circ \pi$ is of the form $f + \phi \circ \pi$ where f is a compensation function, so by Theorem 3.3.5 its equilibrium state m is a lift of p .

The assumption that $I_\nu \circ \pi - I_\mu$ is a compensation function may seem to be a strong one, but we will see in the following theorem that something even stronger is true. It should be noted that this theorem is independent of Theorem 3.3.1. Although 3.3.1 provides sufficient conditions for the construction of compensation functions which look like the difference of two information cocycles, we will independently prove their existence in the following theorem by following a proof more similar to the one given in [3].

Theorem 3.4.3. *Let (X, T) and (Y, S) be shifts of finite type and $\pi : X \rightarrow Y$ be a 1-block code. Suppose there exist $f \in \mathcal{W}(X)$ and $g \in \mathcal{W}(Y)$ with equilibrium states μ_f and ν_g respectively, such that $\mu_f \circ \pi^{-1} = \nu_g$. Then $I_{\nu_g} \circ \pi - I_{\mu_f}$ is a compensation function for (X, Y, π) .*

This theorem was observed in both [3] and [22]. An analogous statement was proved for Markov measures in [3], and the proof given here was sketched in [22].

An even more general statement can be found in [2]. We give the full proof here for completeness, and to highlight the similarities between the relating of functions of two coordinates to Markov measures, and the relating the Walters functions to the Ruelle operator.

Our definition of the Ruelle operator will be given for one-sided shifts of finite type. If X is a two-sided shift, it is quite easy to convert it into a one-sided shift. For each $x \in X$ one can define a one-sided sequence $x^+ = .x_0x_1 \dots$ and let X^+ be the collection of all such points. If F is a one-sided function on X , by letting π^+ be the projection of X onto X^+ and $F^+ = F \circ \pi^+$ then $F^+(x^+) = F(x)$. In the proofs that follow, we will drop the $+$ notation, and simply treat X and F as one sided where it makes sense.

Definition 3.4.2. Let X be an irreducible, one-sided SFT. The Ruelle operator of a one-sided function ϕ is the operator defined by

$$(L_\phi h)(x) = \sum_{y \in T^{-1}x} \exp(\phi(y))h(y)$$

for $h \in C(X)$.

The significance of the Ruelle operator to this proof is given by a theorem which is analogous to, though much more general than, Theorem 2.5.6.

Theorem 3.4.4 ([20, Theorem 3.3]). *Let X be an irreducible one-sided subshift of finite type. Let $\phi \in C(X)$ satisfy $\sum_{n=0}^{\infty} \text{var}_n(\phi) < \infty$. Then there exist a number $\lambda > 0$, $h \in C(X)$ and $\mu \in \mathcal{M}(X)$ such that $h > 0$, $\int h d\mu = 1$, $L_\phi h = \lambda h$, $L_\phi^* \mu = \lambda \mu$ and $L_\phi^n f / \lambda^n \rightarrow h \int f d\mu$ for $f \in C(X)$.*

Our first step in proving Theorem 3.4.3 is the following lemma.

Lemma 3.4.5. *Let $\nu \in \mathcal{M}(Y)$ be the equilibrium state of some Walters function on Y . Let μ_f and ν_g be equilibrium states of $f \in \mathcal{W}(X)$ and $g \in \mathcal{W}(Y)$ such that $\mu_f \circ \pi^{-1} = \nu_g$. Then $\mathcal{P}(I_{\nu_g} \circ \pi - I_{\mu_f} - I_\nu) = 0$.*

Proof. From Theorem 3.4.1 we know that I_{μ_f} , I_{ν_g} , and I_ν have continuous versions which are one-sided Walters functions. Furthermore, from Lemma 3.4.2 we know that $I_{\nu_g} \circ \pi$ and $I_\nu \circ \pi$ are also one-sided Walters functions. So $F = I_{\nu_g} \circ \pi - I_{\mu_f} - I_\nu \circ \pi$ is a one-sided Walters function.

Treating F and X as their one-sided versions, Definition 3.4.2 gives us $(L_F 1)(x) = \sum_{y \in T^{-1}x} \exp(F(y))$. From this we see that

$$(L_F^n 1)(x) = \sum_{y \in T^{-n}x} \exp(F(y) + F(Ty) + \dots + F(T^{n-1}y))$$

For each $i \in \mathcal{A}(X)$ pick $x(i) \in X$ such that $x(i)_0 = i$. Let A be the transition matrix of X . For any word w such that $A(w_{n-1}, i) = 1$ let $wx(i) \in [w]$ be a point such that $(wx(i))_{j+n} = x(i)_j$ for $j \geq 0$. It follows that

$$(L_{F+1})(x(i)) = \sum_{\substack{w \in W_n \\ A(w_{n-1}, i) = 1}} \exp((S_n F)(wx(i)))$$

This allows us to show that

$$\begin{aligned} & \sum_{w \in W_n} \exp((S_n F)(wx(i))) \\ & \leq \left(\sum_{i \in \mathcal{A}(X)} (L_F 1)(x(i)) \right) \\ & \leq |\mathcal{A}(X)| \sum_{w \in W_n} \exp((S_n F)(wx(i))) \end{aligned}$$

And thus

$$\begin{aligned} & \frac{1}{|\mathcal{A}(X)|} \sum_{i \in \mathcal{A}(X)} (L_F 1)(x(i)) \\ & \leq \sum_{w \in W_n} \exp((S_n F)(wx(i))) \\ & \leq \sum_{i \in \mathcal{A}(X)} (L_F 1)(x(i)) \end{aligned}$$

This allows us to calculate the pressure of F as

$$\mathcal{P}(F) = \lim_{n \rightarrow \infty} \frac{1}{n} \log \sum_{i \in \mathcal{A}(X)} (L_F 1)(x(i)) \quad (3.4.1)$$

Next we observe that F satisfies the hypotheses of the Ruelle operator Theorem 3.4.4. So there exists a number $\lambda > 0$, $h \in C(X)$, and $m \in \mathcal{M}(X)$ such that $h > 0$, $\int h dm = 1$, $L_F h = \lambda h$, $L_F^* m = \lambda m$, and $L_F^n 1 / \lambda^n \rightarrow h$.

So for all $\epsilon > 0$ there exists an $N > 0$ such that for all $n > N$

$$\sup_x |(L_F^n 1)/\lambda^n - h| < \epsilon$$

From this we have

$$\lambda^n(h - \epsilon) < (L_F^n 1) < \lambda^n(h + \epsilon)$$

Thus

$$\begin{aligned} & \sum_{i \in \mathcal{A}(X)} \lambda^n(h(x(i)) - \epsilon) \\ & \leq \sum_{i \in \mathcal{A}(X)} (L_F^n 1)(x(i)) \\ & \leq \sum_{i \in \mathcal{A}(X)} \lambda^n(h(x(i)) + \epsilon) \end{aligned}$$

Combining this with our previous expression for pressure allows us to say that

$$\mathcal{P}(F) = \log(\lambda)$$

For our particular F we see that

$$\begin{aligned} F(x) &= I_{\nu_g} \circ \pi(x) - I_{\mu_f}(x) - I_{\nu} \circ \pi(x) \\ &= -\log \nu_g([\pi(x_0)] | (\mathcal{P}_Y)_{-\infty}^{-1}) + \log \mu_f([x_0] | (\mathcal{P}_X)_{-\infty}^{-1}) + \log \nu([\pi(x_0)] | (\mathcal{P}_Y)_{-\infty}^{-1}) \\ &= \log \left(\frac{\mu_f([x_0] | (\mathcal{P}_X)_{-\infty}^{-1}) \nu([\pi(x_0)] | (\mathcal{P}_Y)_{-\infty}^{-1})}{\nu_g([\pi(x_0)] | (\mathcal{P}_Y)_{-\infty}^{-1})} \right) \end{aligned}$$

So

$$\begin{aligned} & (L_F^n 1)(x) \\ &= \sum_{\substack{w \in W_n \\ A(w_{n-1}, x_0)=1}} \prod_{i=0}^{n-1} \frac{\mu_f([w_i] | (\mathcal{P}_X)_{-\infty}^{-1}) \nu([\pi(w_i)] | (\mathcal{P}_Y)_{-\infty}^{-1})}{\nu_g([\pi(w_i)] | (\mathcal{P}_Y)_{-\infty}^{-1})} \\ &= \sum_{\substack{w \in W_n \\ A(w_{n-1}, x_0)=1}} \frac{\mu_f([w] | (\mathcal{P}_X)_{-\infty}^{-n}) \nu([\pi(w)] | (\mathcal{P}_Y)_{-\infty}^{-n})}{\nu_g([\pi(w)] | (\mathcal{P}_Y)_{-\infty}^{-n})} \end{aligned} \tag{3.4.2}$$

Note that μ_f is also the equilibrium state of $-I_{\mu_f}$. From this, [22, Prop 4.1] tells

us that there exist $\alpha_1, \beta_1 > 0$ satisfying

$$\alpha_1 \leq \mu_f([x_0 \dots x_{n-1}]) \exp(n\mathcal{P}(-I_{\mu_f}) - S_n(-I_{\mu_f})(x)) \leq \beta_1$$

In addition,

$$\exp(S_n(-I_{\mu_f})(x)) = \mu_f([x_0 \dots x_{n-1}] \mid (\mathcal{P}_X)_{-\infty}^{-n})$$

and $\mathcal{P}(-I_{\mu_f}) = 0$. So we have

$$\alpha_1 \leq \frac{\mu_f([x_0 \dots x_{n-1}])}{\mu_f([x_0 \dots x_{n-1}] \mid (\mathcal{P}_X)_{-\infty}^{-n})} \leq \beta_1$$

Similar conclusions hold for ν_g and ν . Combining these we have that there exist $c, C > 0$ such that

$$\begin{aligned} & c \left(\frac{\mu_f([x_0 \dots x_{n-1}]) \nu([\pi(x_0 \dots x_{n-1})])}{\nu_g([\pi(x_0 \dots x_{n-1})])} \right) \\ & \leq \frac{\mu_f([x_0 \dots x_{n-1}] \mid (\mathcal{P}_X)_{-\infty}^{-n}) \nu([\pi(x_0 \dots x_{n-1})] \mid (\mathcal{P}_Y)_{-\infty}^{-n})}{\nu_g([\pi(x_0 \dots x_{n-1})] \mid (\mathcal{P}_Y)_{-\infty}^{-n})} \\ & \leq C \left(\frac{\mu_f([x_0 \dots x_{n-1}]) \nu([\pi(x_0 \dots x_{n-1})])}{\nu_g([\pi(x_0 \dots x_{n-1})])} \right) \end{aligned} \quad (3.4.3)$$

Next, letting B be the transition matrix of Y and V_n the set of words of length n in Y , we can show that

$$\begin{aligned} \sum_{\substack{w \in W_n \\ A(w_{n-1}, i)=1}} \frac{\mu_f([w]) \nu([\pi(w)])}{\nu_g([\pi(w)])} &= \sum_{\substack{v \in V_n \\ B(v_{n-1}, \pi(i))=1}} \sum_{w \in \pi^{-1}(v)} \frac{\mu_f([w]) \nu([v])}{\nu_g([v])} \\ &= \sum_{\substack{v \in V_n \\ B(v_{n-1}, \pi(i))=1}} \nu([v]) \end{aligned} \quad (3.4.4)$$

Where

$$\sum_{w \in \pi^{-1}(v)} \mu_f([w]) / \nu_g([v]) = 1$$

because $\mu_f \circ \pi^{-1} = \nu_g$.

Combining equations (3.4.1) - (3.4.4) shows us that

$$\begin{aligned}
\mathcal{P}(F) &= \lim_{n \rightarrow \infty} \frac{1}{n} \log \sum_{i \in \mathcal{A}(X)} (L_{F+1})(x(i)) \\
&= \lim_{n \rightarrow \infty} \frac{1}{n} \log \sum_{i \in \mathcal{A}(X)} \sum_{\substack{w \in W_n \\ A(w_{n-1}, x_0)=1}} \frac{\mu_f([w] | (\mathcal{P}_X)_{-\infty}^{-n}) \nu([\pi(w)] | (\mathcal{P}_Y)_{-\infty}^{-n})}{\nu_g([\pi(w)] | (\mathcal{P}_Y)_{-\infty}^{-n})} \\
&= \lim_{n \rightarrow \infty} \frac{1}{n} \log \sum_{i \in \mathcal{A}(X)} \sum_{\substack{w \in W_n \\ A(w_{n-1}, x_0)=1}} \frac{\mu_f([w]) \nu([\pi(w)])}{\nu_g([\pi(w)])} \\
&= \lim_{n \rightarrow \infty} \frac{1}{n} \log \sum_{i \in \mathcal{A}(X)} \sum_{\substack{v \in V_n \\ B(v_{n-1}, \pi(i))=1}} \nu([v]) \\
&= \lim_{n \rightarrow \infty} \frac{1}{n} \log |\mathcal{A}(X)| \\
&= 0
\end{aligned}$$

□

Lemma 3.4.6. *Let (X, Y, π) be a factor triple with X and Y shifts of finite type and π a 1-block code. Let $f \in \mathcal{W}(X)$ and $g, \phi \in \mathcal{W}(Y)$ with equilibrium states μ_f, ν_g and ν_ϕ such that $\mu_f \circ \pi^{-1} = \nu_g$. If $F = I_{\nu_g} \circ \pi - I_{\mu_f} - I_{\nu_\phi} \circ \pi$ then F is cohomologous to $-I_{\mu_F}$ where μ_F is the unique equilibrium state of F .*

Proof. This is a direct consequence of Lemma 3.4.5 and the fact (from [22, Prop 4.1]) that for any Walters function,

$$F \sim -I_{\mu_F} + \mathcal{P}(F)$$

□

We are now prepared to prove the main theorem of this section.

Proof of Theorem 3.4.3. First, let $\phi \in \mathcal{W}(Y)$. We know then that $\phi \sim -I_{\nu_\phi} + \mathcal{P}(\phi)$ where ν_ϕ is the equilibrium state of ϕ . Pre-composing with π , this expression becomes

$$\phi \circ \pi \sim -I_{\nu_\phi} \circ \pi + \mathcal{P}(\phi)$$

By Lemma 3.4.6 we know that for $F = I_{\nu_g} \circ \pi - I_{\mu_f} - I_{\nu_\phi} \circ \pi$

$$-I_{\mu_F} \sim I_{\nu_g} \circ \pi - I_{\mu_f} - I_{\nu_\phi} \circ \pi$$

Combining these expressions we obtain

$$-I_{\mu_F} \sim I_{\nu_g} \circ \pi - I_{\mu_f} + \phi \circ \pi - \mathcal{P}(\phi)$$

Taking the pressure of both sides of this yields

$$\begin{aligned} \mathcal{P}(I_{\nu_g} \circ \pi - I_{\mu_f} + \phi \circ \pi - \mathcal{P}(\phi)) &= \mathcal{P}(I_{\nu_g} \circ \pi - I_{\mu_f} + \phi \circ \pi) - \mathcal{P}(\phi) \\ &= \mathcal{P}(-I_{\mu_F}) \\ &= 0 \end{aligned}$$

This proves the theorem for all Walters class functions on Y . The Walters functions contain the all functions which depend on finitely many coordinates, and are thus dense in $C(Y)$. By the continuity of the pressure function, this proves the theorem for all $\phi \in C(Y)$. \square

Now that we have the machinery to do so, let us explicitly construct a compensation function. For simplicity, we will work in the setting of [3]. In other words, we will build our compensation function out of Markov measures.

Example 3.4.1 (Compensation function for a split golden mean shift). Let (X, Y, π) be the factor triple defined by the following directed graphs: The map π is the

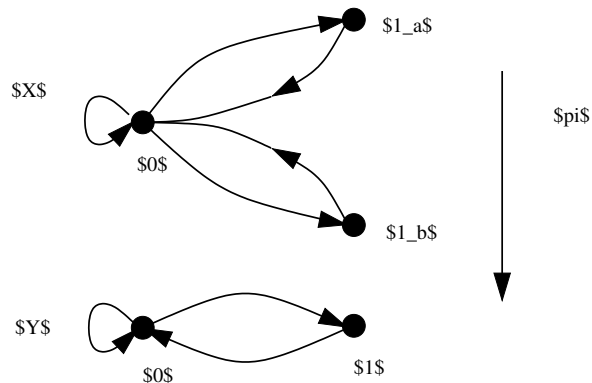


Figure 3.1: The split golden mean shift

1-block code defined by $\pi(0) = 0$ and $\pi(1_a) = \pi(1_b) = 1$. The space Y is commonly called the golden mean shift, so we will refer to X as the split golden mean shift. To construct our compensation function for this factor, we need to find Markov measures ν_0 on Y and μ_0 on X such that $\mu_0 \circ \pi^{-1} = \nu_0$. Let $\phi = \frac{1+\sqrt{5}}{2}$, the golden mean. Let ν_0 be the Markov measure defined by the matrix

$$P_0 = \begin{bmatrix} \phi^{-2} & \phi^{-1} \\ 1 & 0 \end{bmatrix}$$

This is in fact the measure of maximal entropy for Y , from which it gets its name.

It is easy to check that for any choice of α , a lift of ν_0 is given by the matrix

$$M_0 = \begin{bmatrix} \phi^{-2} & \alpha\phi^{-1} & (1-\alpha)\phi^{-1} \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

We will express the information cocycles of these measures as matrices with the symbol ∞ used as a placeholder for transitions which are not allowed in the shift.

$$I_{\mu_0} = \begin{bmatrix} -\log \phi^{-2} & -\log \alpha\phi^{-1} & -\log(1-\alpha)\phi^{-1} \\ 0 & \infty & \infty \\ 0 & \infty & \infty \end{bmatrix}$$

$$I_{\nu_0} = \begin{bmatrix} -\log \phi^{-2} & -\log \phi^{-1} \\ 0 & \infty \end{bmatrix}, \quad I_{\nu_0} \circ \pi = \begin{bmatrix} -\log \phi^{-2} & -\log \phi^{-1} & -\log \phi^{-1} \\ 0 & \infty & \infty \\ 0 & \infty & \infty \end{bmatrix}$$

Now by Theorem 3.4.3 the function

$$I_{\nu_0} \circ \pi - I_{\mu_0} = \begin{bmatrix} 0 & \log \alpha & \log(1-\alpha) \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

is a compensation function for π . In the proof of Theorem 3.4.3, we saw that if we take another measure ν on Y which is also an equilibrium state of a Walters function, then the cohomology equation $I_{\nu_0} \circ \pi - I_{\mu_0} - I_{\nu} \circ \pi \sim -I_{\mu}$ should uniquely determine a measure μ such that $\mu \circ \pi^{-1} = \nu$. Let ν be the Markov measure (which we know is

the equilibrium state of some function of two coordinates) given by

$$P = \begin{bmatrix} p & 1-p \\ 1 & 0 \end{bmatrix}$$

$$I_\nu \circ \pi = \begin{bmatrix} -\log p & -\log(1-p) & -\log(1-p) \\ 0 & \infty & \infty \\ 0 & \infty & \infty \end{bmatrix}$$

Then the cohomology equation above gives us

$$I_\mu = \begin{bmatrix} -\log p & -\log \alpha(1-p) & -\log(1-\alpha)(1-p) \\ 0 & \infty & \infty \\ 0 & \infty & \infty \end{bmatrix}$$

We conclude that the unique measure μ determined by ν_0 , μ_0 and ν is given by the matrix

$$M = \begin{bmatrix} p & \alpha(1-p) & (1-\alpha)(1-p) \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

Now let us verify that $I_{\nu_0} \circ \pi - I_{\mu_0}$, as computed above, is in fact a compensation function. We will verify this for a generic function $f(x) \in C(Y)$ which only depends of the first two coordinates of x .

$$f = \begin{bmatrix} f(00) & f(01) \\ f(10) & 0 \end{bmatrix}, f \circ \pi + I_{\nu_0} \circ \pi - I_{\mu_0} = \begin{bmatrix} f(00) & f(01) + \log \alpha & f(01) + \log(1-\alpha) \\ f(10) & 0 & 0 \\ f(10) & 0 & 0 \end{bmatrix}$$

If $I_{\nu_0} \circ \pi - I_{\mu_0}$ is a compensation function, then the pressure of these two functions above should be equal. First we compute $\mathcal{P}_Y(f)$. We can do this using Theorem 2.5.6. According to that theorem, we first form the matrix

$$Q_0 = \begin{bmatrix} e^{f(00)} & e^{f(01)} \\ e^{f(10)} & 0 \end{bmatrix}$$

The eigenvalues of this matrix are given by

$$\lambda(\lambda - e^{f(00)}) - e^{f(10)+f(01)} = \lambda^2 - \lambda e^{f(00)} - e^{f(10)+f(01)} = 0$$

So the largest eigenvalue is $\lambda_+ = \frac{e^{f(00)} + \sqrt{e^{2f(00)} + 4e^{f(01)+f(10)}}}{2}$. Thus the pressure of f is $\mathcal{P}_Y(f) = \log \lambda_+$

To compute $\mathcal{P}_X(f \circ \pi + I_{\nu_0} \circ \pi - I_{\mu_0})$ we form the matrix

$$Q_1 = \begin{bmatrix} e^{f(00)} & \alpha e^{f(01)} & (1-\alpha)e^{f(01)} \\ e^{f(10)} & 0 & 0 \\ e^{f(10)} & 0 & 0 \end{bmatrix}$$

Which has eigenvalues given by

$$\begin{aligned} & \lambda^2(\lambda - e^{f(00)}) - \alpha e^{f(01)}(\lambda e^{f(10)}) - (1-\alpha)e^{f(10)}(\lambda e^{f(01)}) \\ & = \lambda(\lambda^2 - \lambda e^{f(00)} - e^{f(10)+f(01)}) = 0 \end{aligned}$$

Clearly this will have the same largest eigenvalue as Q_0 , so $\mathcal{P}_Y(f) = \mathcal{P}_X(f \circ \pi + I_{\nu_0} \circ \pi - I_{\mu_0}) = \log \lambda_+$.

Chapter 4

Structure of Compensation Functions and Relative Equilibrium States

In the previous chapter we were able to construct, for certain factor triples, a compensation function in the Walters class of continuous functions. We also saw that Walters functions have a unique equilibrium state which is fully supported. What we have not addressed is another obvious question. What can we say about the relative equilibrium states of a compensation function? The primary focus of this chapter will be the two central theorems of this work, both of which address this question.

4.1 Two Types of Compensation Functions

If we recall our discussion of how a compensation function acts as an “oracle” for relative entropy, then we can see that we are already making an implicit assumption about its relative equilibrium states. First, if the relative equilibrium state were not fully supported we would only see how relative entropy is being generated on its support. Second, if the relative equilibrium state did not have positive relative entropy itself, then it certainly could not be identifying the way in which relative topological entropy appears. In this chapter, we will see that compensation functions like the ones constructed in [3] do in fact have the oracle-like properties advertised above. More generally, we will see that all functions which meet a particular continuity condition

have fully supported, positive relative entropy relative equilibrium states. However, we will also see that if we relax this continuity at all then there exist compensation functions who achieve their compensation in a fundamentally different way, and whose relative equilibrium states live on a very small subset of the cover. In this way, we observe a kind of phase transition in the structure of relative equilibrium states.

Let us give a precise definition of what we mean when we call a compensation function an oracle, which should match the intuition we gained by comparing them to the relative topological entropy.

Definition 4.1.1. Let (X, Y, π) be a subshift factor triple. Let $f \in C(X)$ be a compensation function for this factor triple. Then f is an *oracle-type* compensation function if for every fully supported $\nu \in \mathcal{M}(Y)$, the relative equilibrium states of f over ν are fully supported and have positive relative entropy.

In his original survey of compensation functions and relative equilibrium states [22], Walters proves that for a broad range of subshift factor triples (although not all of them, as is claimed in the work) there exists a continuous compensation function. His method of constructing them relies on the existence of a subshift $X_0 \subset X$ for which $\pi|_{X_0}$ is surjective and finite-to-one. His error lies in believing that every subshift factor triple possesses such an X_0 . For the moment, let us suppose that we are working in a setting where this is a safe assumption. In fact, let us restrict ourselves to the setting of the main theorems of this chapter, namely factor triples (X, Y, π) where X is an irreducible shift of finite type and π is a 1-block code. As we will see in the coming sections, these factor triples always possess a finite-to-one subfactor.

Walters method for constructing a continuous compensation function is to build a function f such that $f \leq 0$, $f|_{X_0} = 0$ and f becomes negative very quickly as we move away from X_0 . By showing that the maximal relative pressure of his function is zero, he uses 3.2.3 to show that it is a compensation function. Why should a function f constructed like this have $W(f) = 0$? To understand this, we need to understand a little bit more about the difference between infinite-to-one and finite-to-one block codes.

By finite-to-one, we mean that there exists a K such that $|\pi^{-1}(y)| \leq K$ for all $y \in Y$. Imagine that there are two words w and v of length n in X such that $w_0 = v_0$, $w_{n-1} = v_{n-1}$ and $\pi(w) = \pi(v)$. Then above any point y which contains $\pi(w)$ infinitely often we could choose either w or v at each each occurrence of $\pi(w)$. Such

a y would certainly have infinitely many preimages. Additionally, every word in Y with k occurrences of $\pi(w)$ would contribute at least 2^k corresponding words to X . It would be reasonable to expect that this makes the topological entropy of X greater than that of Y . If such a pair of words exist we say that π *collapses a diamond*. The following lemma will make these notions precise. For proofs, we refer the reader to [10, Chapter 8].

Lemma 4.1.1. *Let (X, Y, π) be a factor triple with X an irreducible shift of finite type and π a 1-block code. Then the following are equivalent.*

- $h_{\text{top}}(X) > h_{\text{top}}(Y)$
- π *collapses a diamond*
- π *is infinite-to-one*

The converse to this is that π is finite-to-one if and only if $h_{\text{top}}(X) = h_{\text{top}}(Y)$. In this case, there is no relative topological entropy. We can now see how Walters compensation functions achieves their task. Recall that a relative equilibrium state maximizes $h_{\pi}(\mu) + \int f d\mu$. Unlike oracle-type compensation functions, these functions relative equilibrium states live on a set which is finite-to-one over Y , because the penalty incurred in the integral for straying from X_0 is so negative that it outweighs any potential to generate relative entropy. This motivates the our next definition.

Definition 4.1.2. Let (X, Y, π) be a subshift factor triple and $f \in C(X)$ be a compensation function. Then f is a *Walters-type* compensation function if there exists a subshift $X_0 \subset X$ with $\pi|_{X_0}$ surjective and finite-to-one such that the relative equilibrium states of f are supported on X_0 . The relative equilibrium states of such an f will have zero relative entropy.

We should be careful to point out that Walters-type compensation functions are not related to the Walters class of continuous functions.

One issue with Walters proof is that it is not constructive. We are guaranteed the existence of a compensation function with relative equilibrium states on X_0 , but we don't get to see what one looks like. In particular, we may want to know how sharply f has to decline off X_0 to achieve compensation. One type of well understood smoothness condition for continuous functions is summable variation. The n th variation of of a function $f \in C(X)$ is defined by $\text{var}_n(f) = \max_{d(x,y) \leq 2^{-n}} |f(x) - f(y)|$.

In other words, this is the maximum distance two points which agree on a $2n$ -window about the origin can be mapped apart by f . If this quantity is summable over n then, because of the ergodic theorem, it actually helps to bound the integral of f . Functions with summable variation are also Walters class functions, so they have fully supported equilibrium states. This leads us to conjecture that summable variation is sufficient to ensure that the relative equilibrium states are fully supported as well. The proof of this fact will be given in the subsequent section.

If summable variation implies that our compensation function is oracle-type, how much less smoothness is needed to build a Walters-type compensation function? To explore this concept, we introduce a class of continuity conditions which depend on a single parameter, and which encompass the idea of a function becoming “less smooth” as we vary this parameter.

Definition 4.1.3. A function $f \in C(X)$ will be called p -Dini if

$$\sum_{n=0}^{\infty} (\text{var}_n(f))^p < \infty$$

If f is 1-Dini, then f has summable variation.

As p grows, the functions which are p -Dini are allowed to have much larger variation. In the final section of this chapter, we will see that relaxing our continuity from 1-Dini to p -Dini for any $p > 1$ is enough to build a Walters-type compensation function. Continuing to borrow terminology from physical thermodynamics, this type of a fundamental shift in the possible behaviour of equilibrium states at a sharp boundary can be appropriately labelled a phase transition.

4.2 1-Dini functions and oracle-type compensation functions

In this section we will focus on functions which are 1-Dini (in other words, those which have summable variation). Our inspiration for studying these functions was the hope that, like their equilibrium states, their relative equilibrium states would prove to be fully supported. Additionally, we would like for 1-Dini compensation functions to be oracle-type. Thus, our goal will be to prove

Theorem 4.2.1. *If (X, Y, π) is a factor triple with X an irreducible shift of finite type and π an infinite-to-one 1-block code, and f is a 1-Dini function on X then for any fully supported measure $\nu \in \mathcal{M}(Y)$*

- i. The relative equilibrium states of f over ν are fully supported.*
- ii. The relative equilibrium states of f over ν have positive relative entropy.*

In particular, if f is a compensation function then f is oracle-type.

4.2.1 Preparation for the proof of Theorem 4.2.1

In [24] a result similar to 4.2.1 was proved for measures of maximal relative entropy. Measures of maximal relative entropy are those measures $\mu \in \mathcal{M}(X)$ such that $h(\mu) = \sup_{m \circ \pi^{-1} = \mu \circ \pi^{-1}} h(m)$. In other words, they are the relative equilibrium states of 0 over their own push forward to Y . These measures were shown to have full support if their push forward had full support. Our proof will make use of techniques for interleaving measures similar to those seen in [24], [15] and [1]. To best utilize these techniques, we give here several key lemmas from [24].

The first lemma has to do with the ability to construct diamonds which live half-in and half-out of a proper subshift of X .

Lemma 4.2.2. *Let (X, Y, π) be a factor triple with X an irreducible shift of finite type and π a 1-block code. If $Z \subset X$ is a proper subshift such that $\pi(Z) = Y$, then there exist two words u and v with the following properties:*

- i. $\pi(u) = \pi(v)$*
- ii. $|u| = |v| = n$, $u_0 = v_0$ and $u_{n-1} = v_{n-1}$*
- iii. u is a word appearing in Z , v does not occur in Z*
- iv. For any words s and t with sut a word in Z , there is exactly one occurrence of v in svt .*

The second two lemmas actually apply to any invariant measure μ on a σ -algebra \mathcal{F} . For a set $A \in \mathcal{F}$ (or a sub- σ -algebra of \mathcal{F}) with $\mu(A) > 0$ we let $\mathcal{F}_A = \{B \cap A \mid B \in \mathcal{F}\}$, and use similar notation for the restriction of sub- σ -algebras and partitions

to A . Similarly, we let $\mu_A(B) = \frac{\mu(B)}{\mu(A)}$ and for a partition \mathcal{P} and sub- σ -algebra \mathcal{B} we let $H_A(\mathcal{P} \mid \mathcal{C}) = H_{\mu_A}(\mathcal{P}_A \mid \mathcal{B}_A)$.

Lemma 4.2.3. *Let \mathcal{P} be a measurable partition of X and let \mathcal{B} and \mathcal{C} be sub- σ -algebras of \mathcal{F} such that \mathcal{C} and $\mathcal{P} \vee \mathcal{B}$ are independent. If $C \in \mathcal{C}$ and $\mu(C) > 0$, then*

$$H_C(\mathcal{P} \mid \mathcal{B} \vee \mathcal{C}) = H(\mathcal{P} \mid \mathcal{B})$$

Lemma 4.2.4. *Let $A \in \mathcal{F}$ and $\mu(A) > 0$. Then for a measurable partition \mathcal{P} ,*

$$H(\mathcal{P} \mid \mathcal{B}) \geq \mu(A)H_A(\mathcal{P} \mid \mathcal{B})$$

Another key concept required for the proof will be that of a joining of two measures.

Definition 4.2.1. Let (X, T, \mathcal{B}) and (Y, S, \mathcal{C}) be irreducible shifts of finite type, $\mu \in \mathcal{M}(X)$ and $\nu \in \mathcal{M}(Y)$. A *joining* of μ and ν is a $T \times S$ invariant measure $\hat{\mu} \in \mathcal{M}(X \times Y)$ such that for all $A \in \mathcal{B}$ and $B \in \mathcal{C}$ we have

$$\begin{aligned} \hat{\mu}(A \times Y) &= \mu(A) \\ \hat{\mu}(X \times B) &= \nu(B) \end{aligned}$$

We let $J(\mu, \nu)$ be the set of all joinings of μ and ν .

The crucial feature of joinings for our purposes is that there exists a metric \bar{d} on measures which allows us to compare the entropy of μ and ν in terms of their \bar{d} distance. The definition we give is not the most general one, but is sufficient for our purposes.

Definition 4.2.2. Let (X, T, \mathcal{B}) and (Y, S, \mathcal{C}) be irreducible shifts of finite type on a common alphabet \mathcal{A} , $\mu \in \mathcal{M}(X)$ and $\nu \in \mathcal{M}(Y)$. The \bar{d} -distance between ν and μ is

$$\bar{d}(\mu, \nu) = \inf_{\hat{\mu} \in J(\mu, \nu)} \int \delta_0(x, y) d\hat{\mu}$$

where $\delta_0(x, y)$ is 1 if $x_0 \neq y_0$ and 0 otherwise.

We can think of the \bar{d} -distance between μ and ν in the following way. There is a natural pairing of typical points from μ and ν so that in each pair we can choose

the point which is typical for one measure, say ν , alter its first n symbols on a set of roughly size $n\bar{d}(\mu, \nu)$ and obtain the first n symbols of the typical point for μ from the pair. We give a few basic properties of \bar{d} in the next lemma. Proofs of Lemma 4.2.5 and Theorem 4.2.6 can be found in [17, Chapter 7.3]

Lemma 4.2.5. *The \bar{d} -distance is symmetric, reflexive and satisfies the triangle inequality. Furthermore, the infimum in Definition 4.2.2 is in fact a minimum. In other words, there exists a joining $\hat{\mu} \in J(\mu, \nu)$ such that*

$$\bar{d}(\mu, \nu) = \int \delta_0(x, y) d\hat{\mu}$$

The next theorem establishes a connection between $\bar{d}(\mu, \nu)$ and the entropies of μ and ν . One interpretation of the theorem is that the entropy is continuous with respect to the \bar{d} -distance.

Theorem 4.2.6. *If $\bar{d}(\mu, \nu) < \epsilon$ then*

$$|h(\mu) - h(\nu)| = O(\epsilon \log \epsilon)$$

A final theorem which will be of use to use is Pinsker's formula, which can be found in [12, Theorem 6.3].

Lemma 4.2.7. *Let (X, T) be a subshift. Let \mathcal{P} and \mathcal{Q} be partitions of X . Then*

$$H(\mathcal{P} \vee \mathcal{Q} \mid \mathcal{P}_{-\infty}^{-1} \vee \mathcal{Q}_{-\infty}^{-1}) = H(\mathcal{P} \mid \mathcal{P}_{-\infty}^{-1} \vee \mathcal{Q}_{-\infty}^{\infty}) + H(\mathcal{Q} \mid \mathcal{Q}_{-\infty}^{-1})$$

In particular, if \mathcal{Q} is coarser than \mathcal{P} then

$$H(\mathcal{P} \mid \mathcal{P}_{-\infty}^{-1}) = H(\mathcal{P} \mid \mathcal{P}_{-\infty}^{-1} \vee \mathcal{Q}_{-\infty}^{\infty}) + H(\mathcal{Q} \mid \mathcal{Q}_{-\infty}^{-1})$$

4.2.2 Proof of Theorem 4.2.1 part 1

. The first part of the theorem is proved by assuming that a proposed relative equilibrium state μ is not fully supported. We are then able to spread measure around in a way which increases relative entropy. We will exploit f being 1-Dini to bound its integral and show that it is smaller than our entropy gain. Lemma 4.2.2 allows us

to spread our old measure around in a way which is invariant over Y , and which is detectable by our new measure.

Let ν be a fully supported measure on Y and μ a relative equilibrium state of f over ν . Thus μ is one of the measures for which the relative pressure of f is equal to $h(\mu) - h(\nu) + \int f d\mu$. Assume that μ is not fully supported. Then by applying Lemma 4.2.2 to $Z = \text{spt}(\mu)$, we obtain two words u and v which have the same image and meet at their endpoints. In addition, u occurs in Z and v does not, which implies $\mu(u) > 0$ and $\mu(v) = 0$. So blocks of u occur with positive frequency in typical points for μ .

The method of constructing the new higher entropy measure $\bar{\mu}$ is to occasionally swap u for v , spreading measure outside of Z and generating entropy.

In order to avoid ambiguity about the order u 's get swapped, because they may overlap, we make sure swapping does not occur closer than $|u| = n$ steps. The fourth property of the lemma makes sure that when we see a v in the image of a point from Z , we know that if we swap it back to a u we recover the original point.

To be more precise, let $0 < p < 1$ and consider the full 2-shift $\Omega = \{0, 1\}^{\mathbb{Z}}$. Let $\psi : \Omega \rightarrow S$ be a factor map defined by letting $\psi(\omega)_0 = 1$ when $\omega_{-n+1}^0 \in [0^n 1]_{-n+1}^0$. Otherwise, $\psi(\omega)_0 = 0$. Now we construct a measure η on S by pushing forward a Bernoulli measure that has 1 occurring with frequency p .

In order to make a map for switching u 's into v 's, we construct $\phi : X \times S \rightarrow X$ as follows. If $s_i = 1$ and $T^i(x) \in [u]$ then $\phi(x, s)_i^{i+n-1} = v$. Otherwise $\phi(x, s)_i = x_i$. The construction of S ensures that all 1's are at least n apart. As discussed above, this implies that we can recover the original point and thus ϕ is well defined and commutes with the natural shift on $X \times S$. Furthermore, we have $\pi(x) = \pi(\phi(x, s))$ for all $(x, s) \in X \times S$. In addition, if $x \in Z$ then $\phi(x, s)_{-n+1}^{n-1}$ determines x_0 . This is due to the last two properties of Lemma 4.2.2.

The measure $\bar{\mu}$ will be a push forward of $\mu \times \eta$ on $X \times S$. We need to compare the relative entropies of μ and $\bar{\mu}$, which is the same as comparing their entropies because both measures live over ν . Let \mathcal{Q} be the partition of $X \times S$ generated by the first symbol of $\phi(x, s)$. Let \mathcal{P} and \mathcal{R} be the state partitions of X and S . Where appropriate, these will also stand for the pullback of the partitions to $X \times S$. Using

the notation $\mathcal{P}_i^j = \bigvee_{k=i}^j T^k(\mathcal{P})$, we can express the three principal entropies as:

$$h(\bar{\mu}) = H(\mathcal{Q} \mid \mathcal{Q}_{-\infty}^{-1})$$

$$h(\mu) = H(\mathcal{P} \mid \mathcal{P}_{-\infty}^{-1})$$

$$h(\eta) = H(\mathcal{R} \mid \mathcal{R}_{-\infty}^{-1})$$

Note that for $\mu \times \eta$ -a.e. point in $X \times S$, x_0 is determined by $\phi(x, s)_{-n+1}^{n-1}$. So the measurable partition \mathcal{P} generated by x_0 is coarser than $\tilde{\mathcal{Q}} = \mathcal{Q}_{-n+1}^{n-1}$. This allows us to apply 4.2.7 to say that

$$\begin{aligned} h(\bar{\mu}) &= H(\tilde{\mathcal{Q}} \mid \tilde{\mathcal{Q}}_{-\infty}^{-1}) \\ &= H(\mathcal{P} \mid \mathcal{P}_{-\infty}^{-1}) + H(\tilde{\mathcal{Q}} \mid \mathcal{P}_{-\infty}^{\infty} \vee \tilde{\mathcal{Q}}_{-\infty}^{-1}) \\ &= h(\mu) + H(\mathcal{Q} \mid \mathcal{P}_{-\infty}^{\infty} \vee \mathcal{Q}_{-\infty}^{-1}) \end{aligned}$$

We note that

$$\begin{aligned} H(\mathcal{Q}_0^{n-1} \mid \mathcal{P}_{-\infty}^{\infty} \vee \mathcal{Q}_{-\infty}^{-1}) &= H(\mathcal{Q}_0 \mid \mathcal{P}_{-\infty}^{\infty} \vee \mathcal{Q}_{-\infty}^{-1}) + H(\mathcal{Q}_1 \mid \mathcal{P}_{-\infty}^{\infty} \vee \mathcal{Q}_{-\infty}^{-1}) \\ &\quad + \dots + H(\mathcal{Q}_{n-1} \mid \mathcal{P}_{-\infty}^{\infty} \vee \mathcal{Q}_{-\infty}^{-1}) = nH(\mathcal{Q} \mid \mathcal{P}_{-\infty}^{\infty} \vee \mathcal{Q}_{-\infty}^{-1}) \end{aligned}$$

Now applying Lemma 4.2.4 allows us to say that

$$\begin{aligned} H(\mathcal{Q} \mid \mathcal{P}_{-\infty}^{\infty} \vee \mathcal{Q}_{-\infty}^{-1}) &= \frac{1}{n} H(\mathcal{Q}_0^{n-1} \mid \mathcal{P}_{-\infty}^{\infty} \vee \mathcal{Q}_{-\infty}^{-1}) \\ &\geq \frac{\mu([u])}{n} H_{[u] \times S}(\mathcal{Q}_0^{n-1} \mid \mathcal{P}_{-\infty}^{\infty} \vee \mathcal{Q}_{-\infty}^{-1}) \end{aligned}$$

When we restrict our view to $[u]$, \mathcal{Q}_0^{n-1} is either u or v , and this entirely depends

on s_0 . In addition, for $x \in [u]$, $\mathcal{P}_{-\infty}^{\infty} \vee \mathcal{R}_{-\infty}^{-1}$ determines $\mathcal{Q}_{-\infty}^{-1}$. Thus

$$\begin{aligned} H_{[u] \times S}(\mathcal{Q}_0^{n-1} \mid \mathcal{P}_{-\infty}^{\infty} \vee \mathcal{Q}_{-\infty}^{-1}) & \\ &= H_{[u] \times S}(\mathcal{R} \mid \mathcal{P}_{-\infty}^{\infty} \vee \mathcal{Q}_{-\infty}^{-1}) \\ &\geq H_{[u] \times S}(\mathcal{R} \mid \mathcal{P}_{-\infty}^{\infty} \vee \mathcal{R}_{-\infty}^{-1}) \\ &= H(\mathcal{R} \mid \mathcal{R}_{-\infty}^{-1}) \\ &= h(\eta) \end{aligned}$$

The equality in the fourth line is due to Lemma 4.2.3

In the end, we have shown that

$$H(\mathcal{Q} \mid \mathcal{P}_{-\infty}^{\infty} \vee \mathcal{Q}_{-\infty}^{-1}) \geq \frac{\mu([u])}{n} h(\eta) \quad (4.2.1)$$

Now we bound the entropy of η by comparing it to the Bernoulli measure b on Ω . We will compare the measures using the \bar{d} metric. Let us construct a joining of η and b . For cylinder sets $C \subset S$ and $D \subset \Omega$ let $\hat{\mu}(C \times D) = b(\psi^{-1}(C) \cap D)$. This measure is invariant under $\sigma \times \sigma$. We also have that $\hat{\mu}(C \times \Omega) = b(\psi^{-1}(C) \cap \Omega) = \eta(C)$ and $\hat{\mu}(S \times D) = b(D)$. So $\hat{\mu}$ is a joining, and thus

$$\bar{d}(\eta, b) \geq \int \delta_0(s, \omega) d\hat{\mu}$$

If (s, ω) is a typical point for $\hat{\mu}$ then $s_0 = \omega_0$ except when $\omega_0 = 1$ and ω_{-n+1}^{-1} contains at least one additional 1, which is an $O(p^2)$ event. We know 1's occur with probability p in b , thus

$$\bar{d}(\eta, b) = O(p^2)$$

As we will be constructing a series of upper and lower bounds, it will be convenient to introduce the terminology $f(n)$ is $\Omega^+(g(n))$ if $f(n) \geq |cg(n)|$ for some constant c . From Theorem 4.2.6 we have

$$h(\eta) - h(b) = O(p^2 \log p)$$

Combining this with our estimate (4.2.1), we have

$$h(\bar{\mu}) - h(\mu) = \frac{\mu([u])}{n} (\Omega^+(p \log p) + O(p^2 \log p)) = \Omega^+(p \log p)$$

To bound the difference of the integrals of f , note that for $\mu \times \eta$ -a.e. $(x, s) \in X \times S$, x is a typical point for μ and $\phi(x, s) = \bar{x}$ is a typical point for $\bar{\mu}$. So the ergodic theorem tells us that

$$\left| \int f d\mu - \int f d\bar{\mu} \right| \leq \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=0}^{N-1} |f(T^i x) - f(T^i \bar{x})|$$

The only places where x and \bar{x} differ are where a u from x has been changed to a v in \bar{x} . This happens on average $\mu([u])\eta([1])N$ times in a run of N . Since f is 1-Dini we can say $\sum_n (\text{var}_n(f)) = L$. So the most $f(T^i x)$ and $f(T^i \bar{x})$ can differ by is L and each v can contribute at most nL to the sum above. Accounting for possible tail contributions, we have that for all $\epsilon > 0$ there is an N_0 such that for $N > N_0$

$$\frac{1}{N} \sum_{i=0}^{N-1} |f(T^i x) - f(T^i \bar{x})| \leq \mu([u])\eta([1])Ln + \frac{2L}{N} + \epsilon$$

Taking limits and using the fact that $\eta([1]) = O(p)$, we estimate the integral as

$$\left| \int f d\mu - \int f d\bar{\mu} \right| = O(p)$$

Finally, since the entropy term is $\Omega^+(p \log p)$ and the integral is only $O(p)$ we have that for p small enough, $h(\bar{\mu}) + \int f d\bar{\mu} - h(\mu) - \int f d\mu > 0$. However, μ was assumed to have maximal relative entropy over ν . This proves that μ is fully supported. \square

4.2.3 Proof of Theorem 4.2.1 part 2

. Now that we know the relative equilibrium states of a 1-Dini function are fully supported, we seek to prove that they also have positive relative entropy. Let $\nu \in \mathcal{M}(Y)$ be fully supported and $\mu \in \mathcal{M}(X)$ be a relative equilibrium state of f over ν .

Assume that $h(\mu | \nu) = 0$. Proceeding similarly to the first part of the proof we will construct a new measure $\bar{\mu}$ which has greater relative pressure. Our estimate will be showing that the difference in relative entropies is greater than in the integrals.

Once again, $\bar{\mu}$ is constructed by taking a pair of words u and v which can be swapped in a point without affecting the eventual image in Y . Here, we do not need the strong detectability conditions provided by Lemma 4.2.2. We can find such a pair

of paths from the fact that π is infinite-to-one, and thus there exist a pair of paths u and v which form a diamond over Y .

Let $|u| = n$. Construct the approximately Bernoulli process (S, σ, η) as before. Thus no two 1's occur in S closer than n . We construct $\bar{\mu}$ by pushing $\mu \times \eta$ forward by ϕ which changes u into v and v into u whenever $s_0 = 1$.

The entropy term we need to estimate is $h(\bar{\mu} | \nu) - h(\mu | \nu) = h(\bar{\mu} | \nu)$. Let \mathcal{P} be the partition on $X \times S$ from the (X, μ) process, $\bar{\mathcal{P}}$ from the $(X, \bar{\mu})$ process, \mathcal{Q} from the (S, η) process and \mathcal{R} from the (Y, ν) process. Then

$$\begin{aligned} h(\bar{\mu} | \nu) &= H(\bar{\mathcal{P}}_0 | \bar{\mathcal{P}}_{-\infty}^{-1} \vee \mathcal{R}_{-\infty}^{\infty}) \\ &\geq H(\bar{\mathcal{P}}_0 | \bar{\mathcal{P}}_{-\infty}^{-1} \vee \mathcal{P}_{-\infty}^{\infty}) \\ &\geq \mu([u] \cup [v]) H_{[u] \cup [v] \times S}(\bar{\mathcal{P}}_0 | \bar{\mathcal{P}}_{-\infty}^{-1} \vee \mathcal{P}_{-\infty}^{\infty}) \\ &= \frac{\mu([u] \cup [v])}{n} H_{[u] \cup [v] \times S}(\bar{\mathcal{P}}_0^{n-1} | \bar{\mathcal{P}}_{-\infty}^{-1} \vee \mathcal{P}_{-\infty}^{\infty}) \end{aligned}$$

The third inequality is from Lemma 4.2.4. In this last term, we have restricted our view to being at either u or v . Knowing $\mathcal{P}_{-\infty}^{\infty}$ means that finding out which one $\bar{\mathcal{P}}_0^{n-1}$ is exactly tells us if $s_0 = 1$ or $s_0 = 0$. This allows us to say

$$\begin{aligned} H_{[u] \cup [v] \times S}(\bar{\mathcal{P}}_0^{n-1} | \bar{\mathcal{P}}_{-\infty}^{-1} \vee \mathcal{P}_{-\infty}^{\infty}) &= H_{[u] \cup [v] \times S}(\mathcal{Q}_0 | \bar{\mathcal{P}}_{-\infty}^{-1} \vee \mathcal{P}_{-\infty}^{\infty}) \\ &\geq H_{[u] \cup [v] \times S}(\mathcal{Q}_0 | \mathcal{Q}_{-\infty}^{-1} \vee \mathcal{P}_{-\infty}^{\infty}) \\ &= H(\mathcal{Q}_0 | \mathcal{Q}_{-\infty}^{-1}) \\ &= h(\eta) \end{aligned}$$

Here, the last equality is due to the independence of X and S , and a final use of Lemma 4.2.3. We have already calculated $h(\eta)$, and so our estimate on the relative entropies is

$$h(\bar{\mu} | \nu) - h(\mu | \nu) = \Omega^+(p \log p)$$

From here, we must again estimate the integral term. One way to do this is to construct a joining of μ and $\bar{\mu}$. Our joining will be $\hat{\mu}$ defined on products of cylinder sets from X as $\hat{\mu}(A \times B) = \mu(\pi_X^{-1}(A) \cap \phi^{-1}(B))$. Note that one way to construct a typical pair of points for $\hat{\mu}$ is to pick (x, \bar{x}) from $X \times X$ where x is typical for μ and $\bar{x} = \phi(x, s)$ for some typical s . The number of times x and \bar{x} will then differ is $O(p)$.

The difference in the integrals is bounded by $\lim_{N \rightarrow \infty} A_N(|f(x) - f(\bar{x})|)$, where A_N is the ergodic average. In each place where x and \bar{x} differ we receive a constant bounded contribution to the sum because f is 1-Dini. Each tail end also gives a bounded contribution, which is constant in N . The contribution from differences is $O(Np)$ and the tail contribution is $O(1)$ so in the limit we have that the integral is $O(p)$. Together with our earlier estimate that the relative entropy gain is $\Omega^+(p \log p)$, this tells us that $\bar{\mu}$ has strictly greater relative pressure than μ , a contradiction. \square

4.3 A Walters-type compensation function

We saw in the previous section that if f is a 1-Dini function and μ a measure which is not fully supported and does not have positive relative entropy, then we can create a new measure $\bar{\mu}$ which generates extra entropy by occasionally swapping a word in the support of μ with one outside of its support. This entropy can be generated in such a way that it is always greater than any new contributions to the integral made by $\bar{\mu}$. This is in direct contrast to Walters method of building compensation functions, which force their relative equilibrium states to live on a set where there is no relative entropy by making the penalty for straying from that set very high. In this section, we will explicitly construct a function which behaves in such a manner. It is the explicit construction of our compensation function which sets it apart from [22][Theorem 3.1]. In that construction, a compensation function which lives on a finite-to-one subfactor is shown to exist, but the argument given is not constructive. By explicitly constructing such a function, we can explore its continuity properties more thoroughly. In particular, we can show that the function we construct will be as close to a 1-Dini function as is possible without actually being 1-Dini. In addition, we will construct an interesting extension of our cover X based on a well known construction (first used in [11]) of an equal-entropy subfactor for shifts of finite type.

Theorem 4.3.1. *Let (X, Y, π) be a factor triple with X a shift of finite type and π an infinite-to-one 1-block code. Then there exists a function f which is simultaneously p -Dini for all $p > 1$ and whose relative equilibrium states live on a subshift $X_0 \subset X$ such that $\pi|_{X_0}$ is surjective and finite to one. From this, it is clear that their relative equilibrium states have zero relative entropy, that they are not fully supported, and that f is a Walters-type compensation function.*

4.3.1 The Marcus-Petersen-Williams subshift and induced subsystems

We will now review two concepts which will play a key role in the proof of Theorem 4.3.1. The first is that of an induced transformation. Suppose μ is an ergodic invariant measure on X and A a set with $\mu(A) > 0$. If we consider the set $A_R = \{x \in A \mid T^k x \in A \text{ for infinitely many } k > 0\}$ then by Theorem 2.4.2 we can say that the $\mu(A \setminus A_R)$ is equal to the average frequency with which a point in $A \setminus A_R$ returns to $A \setminus A_R$, which is clearly 0. In light of this, the function $\tau_A(x) = \min\{k \geq 1 \mid T^k x \in A\}$ is defined for μ -a.e. $x \in A$. We call this function the *first return time* of x .

Definition 4.3.1. If (X, T, \mathcal{B}) is a subshift and $\mu \in \mathcal{M}(X)$ an ergodic measure then the system (A, T_A, \mathcal{B}_A) defined by $T_A(x) = T^{\tau(x)}(x)$ and $\mathcal{B}_A = \{B \cap A \mid B \in \mathcal{B}\}$ is called the *induced subsystem* of A . The measure $\mu_A(B) = \frac{\mu(A \cap B)}{\mu(A)}$ is an invariant probability measure on this system.

The function $\tau^{(n)}(x) = \sum_{i=0}^{n-1} \tau_A \circ T_A^i(x)$ is the n th return time.

Lemma 4.3.2 (Kac's theorem). *For an induced subsystem (A, T_A, \mathcal{B}_A) we have that*

$$\int_A \tau_A d\mu = 1$$

and

$$\lim_{n \rightarrow \infty} \frac{\tau^{(n)}(x)}{n} = \frac{1}{\mu(A)}$$

Proof. These proofs follow [8][Theorem 2.4.1]. For the first part of the lemma, note that $\chi_A(T^k x) \tau_A(T^k x) = \tau_A(T_A^j(x))$ if $k = \tau^{(j)}(x)$ and 0 otherwise. By Theorem 2.4.2, for μ -a.e. $x \in A$

$$\int \chi_A \tau_A d\mu = \lim_{n \rightarrow \infty} \frac{1}{\tau^{(n)}(x)} \sum_{i=0}^{n-1} \tau_A(T_A^i(x)) = 1$$

The second part is proved in an analogous way by applying the ergodic theorem to χ_A . □

The relationship between the entropy of μ_A in induced subsystem and the entropy of μ is given in the next lemma.

Lemma 4.3.3 (Abramov's formula). *For an induced subsystem (A, T_A, \mathcal{B}_A) and the measure μ_A induced by an ergodic measure $\mu \in \mathcal{M}(X)$*

$$h_{T_A}(\mu_A) = \frac{1}{\mu(A)} h_T(\mu)$$

Proof. Let $Q_n = \{x \in A \mid \tau_A(x) = n\}$. In other words, $\{Q_n\}_{n=1}^\infty$ is the partition of A by first-return times. Let us call this partition \mathcal{Q} . Further refine each element Q_n by W_n , the partition of words of length n . We will call this partition $\bar{\mathcal{Q}}$. Then knowing which element of $\bar{\mathcal{Q}}$ a point $x \in A$ lies in tells us $[x_0 \dots x_{\tau_A(x)-1}]$. Thus, if we are given $\bar{\mathcal{Q}}_{-n+1}^0$ we know $[x_0 \dots x_{\tau^{(n)}(x)-1}]$.

We would like to use Theorem 2.4.3 to estimate $h_{T_A}(\mu_A)$ based on the partition $\bar{\mathcal{Q}}$. As $\bar{\mathcal{Q}}$ is not a finite partition, we need to check that $H(\bar{\mathcal{Q}}) < \infty$. We will need to utilize yet another corollary to Jensen's inequality which states that

$$\sum_i a_i \log \frac{a_i}{b_i} \geq \left(\sum_i a_i \right) \log \frac{\sum_i a_i}{\sum_i b_i}$$

Let $|\mathcal{A}(X)| = k$. If we fix an element Q_n of the \mathcal{Q} partition and index the elements of $\bar{\mathcal{Q}}$ for which $\bar{Q}_i \subset Q_n$, we can apply this inequality with $a_i = \mu_A(\bar{Q}_i)$ and $b_i = 1$ to say

$$\begin{aligned} \sum_{\bar{Q}_i \subset Q_n} -\mu_A(\bar{Q}_i) \log \mu_A(\bar{Q}_i) &\leq \mu_A(Q_n) \left(\log \left(\sum_{\bar{Q}_i \subset Q_n} 1 \right) - \log \mu_A(Q_n) \right) \\ &\leq \mu_A(Q_n) (n \log k - \log \mu_A(Q_n)) \end{aligned}$$

Now we estimate the entropy of \mathcal{Q} by

$$\begin{aligned} H(\bar{\mathcal{Q}}) &= \sum_{n=1}^{\infty} \sum_{\bar{Q}_i \subset Q_n} -\mu_A(\bar{Q}_i) \log \mu_A(\bar{Q}_i) \\ &\leq \sum_{n=1}^{\infty} \mu_A(Q_n) (n \log k - \log \mu_A(Q_n)) \end{aligned}$$

Notice that $\sum_{n=1}^{\infty} n \mu_A(Q_n)$ is the expected first return time, which by Lemma 4.3.2 is finite. So clearly the first term in this sum is finite. To deal with the second term,

we split $\{Q_n\}_{n=1}^\infty$ into two sets. Let $B = \{n \mid \mu_A(Q_n) \leq 2^{-n}\}$. First we see that

$$\sum_{n \in B} -\mu_A(Q_n) \log \mu_A(Q_n) \leq \sum_{n \in B} n2^{-n} < \infty$$

For $n \notin B$, note that

$$\begin{aligned} \log \mu_A(Q_n) &> -n \\ -\mu_A(Q_n) \log \mu_A(Q_n) &< n\mu_A(Q_n) \\ \sum_{n \notin B} -\mu_A(Q_n) \log \mu_A(Q_n) &< \sum_{n \notin B} n\mu_A(Q_n) \end{aligned}$$

This last quantity is again bounded by the expected first return time, and thus is finite. So we have that $H(Q) < \infty$.

Let $\epsilon > 0$. Pick some $x \in A$ for which Theorem 2.4.3 applies to both $h_T(\mu)$ and $h_{T_A}(\mu_A)$. Then for all n large enough,

$$\begin{aligned} \left| \frac{-\log \mu_A[x_0 \dots x_{\tau^{(n)}(x)-1}]}{n} - h_{T_A}(\mu_A) \right| &< \epsilon \\ \left| \frac{-\log \mu[x_0 \dots x_{\tau^{(n)}(x)-1}]}{\tau^{(n)}(x)} - h_T(\mu) \right| &< \epsilon \end{aligned}$$

By dividing these two expressions, we are able to show that

$$\frac{(-\log \mu[x_0 \dots x_{\tau^{(n)}(x)-1}]/\tau^{(n)}(x)) - \epsilon}{(-\log \mu_A[x_0 \dots x_{\tau^{(n)}(x)-1}]/n) + \epsilon} < \frac{h_T(\mu)}{h_{T_A}(\mu_A)} < \frac{(-\log \mu[x_0 \dots x_{\tau^{(n)}(x)-1}]/\tau^{(n)}(x)) + \epsilon}{(-\log \mu_A[x_0 \dots x_{\tau^{(n)}(x)-1}]/n) - \epsilon}$$

Since $x \in A$ we have that $\mu_A[x_0 \dots x_{\tau^{(n)}(x)-1}] = \frac{\mu[x_0 \dots x_{\tau^{(n)}(x)-1}]}{\mu(A)}$, from which

$$\frac{-\log \mu_A[x_0 \dots x_{\tau^{(n)}(x)-1}]}{n} = \frac{-\log \mu[x_0 \dots x_{\tau^{(n)}(x)-1}] + \log \mu(A)}{n}$$

Also, by the second statement in Lemma 4.3.2

$$\lim_{n \rightarrow \infty} \frac{\tau^{(n)}(x)}{n} = \frac{1}{\mu(A)}$$

So by taking limits as $n \rightarrow \infty$ and letting $\epsilon \rightarrow 0$ we see that

$$\frac{h_T(\mu)}{h_{T_A}(\mu_A)} = \mu(A)$$

and so the theorem is proved. \square

We now move away from induced subsystems to discuss the method by which we will construct a subshift $X_0 \subset X$ which maps to Y in a finite-to-one way. This construction comes from [11], where it was used to show that every factor triple (X, Y, π) with X sofic and $h(X) > h(Y)$ has an equal-entropy subfactor. We will give the theorem in a less general setting that matches Theorem 4.3.1. Although none of our proofs will rely on the fact that X_0 is sofic, we include this interesting fact and its proof.

Theorem 4.3.4. *Let (X, Y, π) be a factor triple with X a shift of finite type and π a 1-block code. Then there exists a sofic subshift $X_0 \subset X$ which is closed, shift invariant and for which $\pi_0 = \pi|_{X_0}$ is a surjective, finite-to-one block code.*

Proof. Assign a linear ordering to the symbols in $\mathcal{A}(X)$. This establishes a natural lexicographic order on sets of words in X which share the same image under π and which have the same endpoints (in other words, those words which form diamonds in X). If a word w of length n has the least lexicographic order among $\{w' \mid w'_0 = w_0, w'_{n-1} = w_{n-1}, \pi(w') = \pi(w)\}$, then we will say w is Marcus-Petersen-Williams (MPW)-minimal. Let $\mathcal{U} = \{w \mid w \text{ is MPW-minimal}\}$. We will define the MPW-subshift of X to be $X_0 = \{x \in X \mid \forall i, j \in \mathbb{Z}, x_i^j \in \mathcal{U}\}$. Note that if $w \in \mathcal{U}$ then every subword of w is also in \mathcal{U} .

To see that $\pi_0(X_0) = Y$, first let $y \in Y$. For each $n \geq 1$ choose any point $x^{(n)} \in X$ such that $\pi(x_{-n}^n) = y_{-n}^n$ and $x_{-n}^n \in \mathcal{U}$. By the compactness of X we can take a convergent subsequence $x^{(n_i)}$ which converges to some $x \in X$. Clearly $\pi(x) = y$. In addition, x_i^j is a subword of $x_{-n_i}^{n_i}$ for some large enough i , which implies $x_i^j \in \mathcal{U}$. This implies $x \in X_0$. Furthermore, if two words form a diamond over Y then at most one of them can be MPW-minimal, so X_0 contains no diamonds. By Lemma 4.1.1, this ensures π_0 is finite-to-one.

Now we check that X_0 is in fact sofic. One characterization of sofic shifts is that the follower sets of words form a finite collection (see for example [10, Prop. 3.2.9]).

Although not every word in \mathcal{U} appears as a word in X_0 , it will still be sufficient to bound the number of follower sets of words in \mathcal{U} .

For every $u \in \mathcal{U}$, let $F(u) = \{v \in \mathcal{U} \mid uv \in \mathcal{U}\}$. We wish to show that $|\{F(u) \mid u \in \mathcal{U}\}|$ is finite. Let $u, v \in \mathcal{U}$, with $|u| = n$ and $|v| = m$. There are two possible obstructions to uv being in \mathcal{U} . First, if $u_{n-1}v_0$ is not an admissible transition in X , then clearly $uv \notin \mathcal{U}$.

If uv appears as a word in X , then there must exist a word $r \in U$ with $r_0 = u_0$, $r_{n+m-1} = v_{m-1}$ and $\pi(r) = \pi(uv)$. We know that $r_{n-1} \neq u_{n-1}$ because r_0^{n-1} and u are both MPW-minimal. If $r_0^{n-1} > u$ in the purely lexicographic order on words, then the fact that $v \in \mathcal{U}$ would imply $r > uv$, which is not the case. Together, these imply that $r_0^{n-1} < u$ in lexicographic order.

Now let $v' \in \mathcal{U}$ with $|v'| = k$ and suppose there exists s such that $s_0 = r_n$, $s_{k-1} = v'_{k-1}$ and $\pi(s) = \pi(v')$. Then $uv' \notin \mathcal{U}$ because $r_0^{n-1}s_0^{k-1} < uv'$. Next let $u' \in \mathcal{U}$ with $|u'| = l$ and $u'_{l-1} = u_{n-1}$. Suppose there exists t such that $t_0 = u'_0$, $t_{l-1} = r_{l-1}$ and $t_0^{l-1} < u'$. Then by considering the word $t_0^{l-1}r_n^{n+m-1}$ we see that $u'v \notin \mathcal{U}$. Let $G(u)$ be the set of all $a \in \mathcal{A}$ such that $u_{n-1}a$ is not admissible, and $E(u)$ be the set of all $a \in \mathcal{A}$ such that there exists an r with $r_0 = u_0$, $r_{n-1} = a$, $\pi(r) = \pi(u)$ and $r < u$ in lexicographic order. By the above discussion, $G(u)$ and $E(u)$ completely classify the words which are not in the follower set of u , and if $G(u) = G(u')$ and $E(u) = E(u')$ then $F(u) = F(u')$. There are only a finite number of choices for $G(u)$ and $E(u)$, each limited by the number of subsets of \mathcal{A} , and thus only a finite number of distinct follower sets. This proves X_0 is sofic. \square

In order to prove Theorem 4.3.1 we will construct a function f which is simultaneously p -Dini for all $p > 1$, and whose relative equilibrium states live on a MPW-subshift of X . From this it is clear that their relative entropy over measures on Y is zero. The function will also satisfy $f|_{X_0} = 0$. This will be sufficient to ensure f is a

Walters-type compensation function. To see this we first check that for any $\phi \in C(Y)$

$$\begin{aligned} & \sup_{\mu \in \mathcal{M}(X_0)} \left\{ h(\mu) + \int \phi \circ \pi_0 d\mu \right\} \\ &= \sup_{\nu \in \mathcal{M}(Y)} \sup_{\substack{\mu \in \mathcal{M}(X_0) \\ \mu \circ \pi_0^{-1} = \nu}} \left\{ h(\mu) - h(\nu) + h(\nu) + \int \phi \circ \pi_0 d\mu \right\} \\ &= \sup_{\nu \in \mathcal{M}(Y)} \left\{ h(\nu) + \int \phi d\nu \right\} \end{aligned}$$

So $\mathcal{P}_{X_0}(\phi \circ \pi_0) = \mathcal{P}_Y(\phi)$. Now

$$\begin{aligned} \sup_{\mu \in \mathcal{M}(X)} \left\{ h(\mu) + \int f + \phi \circ \pi d\mu \right\} &\geq \sup_{\mu \in \mathcal{M}(X_0)} \left\{ h(\mu) + \int f + \phi \circ \pi_0 d\mu \right\} \\ &= \sup_{\mu \in \mathcal{M}(X_0)} \left\{ h(\mu) + \int \phi \circ \pi_0 d\mu \right\} \end{aligned}$$

So $\mathcal{P}_X(f + \phi \circ \pi) \geq \mathcal{P}_Y(\phi)$. However, from Theorem 3.2.3 we know that

$$\begin{aligned} \sup_{\phi \in \mathcal{C}(Y)} \{ \mathcal{P}_X(f + \phi \circ \pi) - \mathcal{P}_Y(\phi) \} &= \sup_{\mu \in \mathcal{M}(X)} \left\{ h(\mu) - h(\mu \circ \pi^{-1}) + \int f d\mu \right\} \\ &= \sup_{\mu \in \mathcal{M}(X_0)} \left\{ h(\mu) - h(\nu) + \int f d\mu \right\} \\ &= 0 \end{aligned}$$

From which $\mathcal{P}_X(f + \phi \circ \pi) \leq \mathcal{P}_Y(\phi)$. So f is a compensation function.

4.3.2 Clothspinning sequences

One of the main ideas of our proof will be to construct a finite extension of X based on subdividing points into minimal subwords and considering the space of points paired with dividers which delineate these subwords. This extension encodes a sufficient amount of information to reproduce the relative entropy and is easily partitioned into sets where the integral can be bounded. By inducing on returns to the dividers we will be able to show that the potential to increase relative entropy by straying from the MPW-subshift is outweighed by the penalty incurred in the integral.

For points x which are not in the MPW-subshift we will define a set of “clothespins” of x such that the words between adjacent clothespins are minimal, but the words between nonadjacent clothespins are not. By showing that there are only finitely many distinct ways to accomplish this clothespinning we will be able to define a space of clothespinned points which lives over X in a finite-to-one way.

Throughout this section we will be working with a factor triple (X, Y, π) where X is a shift of finite type and Y is a sofic image of X under a 1-block code π . We will assume that some MPW-order has been put on the symbols in X and that X_0 is the associated MPW-subshift. Recall that the set of words which are MPW-minimal (given their endpoints and image) is \mathcal{U} . We also let $\Omega = \{0, 1\}^{\mathbb{Z}}$.

Definition 4.3.2. For $x \in X$ consider the following process. Let $n_0^N(x) = -N$. Given $n_{k-1}^N(x)$ define $n_k^N(x) = \min \{i > n_{k-1}^N(x) \mid x_{n_{k-1}^N} \dots x_{i+1} \notin \mathcal{U}\}$. If some $n_k^N(x) = \infty$, terminate the process. Define $s^{(N)}(x) \in \Omega$ by $s_i^{(N)}(x) = 1$ if $i \in \{n_k^N(x)\}_{k=0}^{\infty}$. If s is a limit of $s^{(N)}(x)$ then we will call s a *clothespinning sequence* of x . A position i for which $s_i = 1$ will be called a *clothespin* of x .

Note that if $x \in X_0$ then $s^{(N)}(x)$ has exactly one clothespin at $-N$ and all 0's otherwise. So the only clothespinning sequence of a point in X_0 is the fixed 0 sequence.

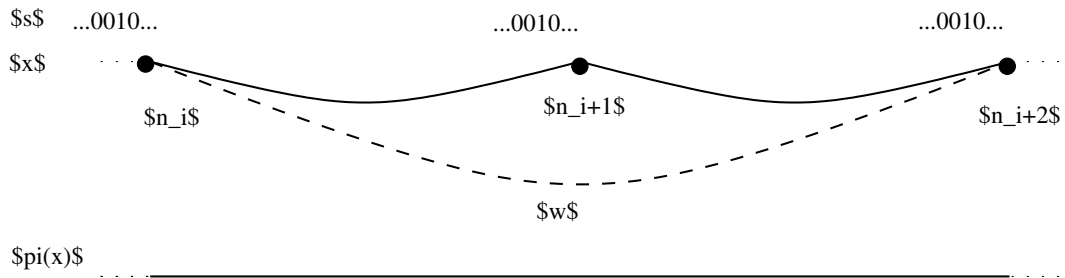


Figure 4.1: Between adjacent clothespins x is minimal, but if we pull out n_{i+1} we see the word w hangs lower than $x_{n_i}^{n_{i+2}}$

The most important property of clothespinning sequences is that if $s_i^j = 10^{j-i-1}1$ then $x_i^j \in \mathcal{U}$, but $x_i^{j+1} \notin \mathcal{U}$. A few additional facts about clothespinning are summarized in the following lemma.

Lemma 4.3.5. *Let $\mu \in \mathcal{M}(X)$ be an ergodic measure such that $\mu(X \setminus X_0) > 0$. Then there exists a k and a set $CP(X)$ with $\mu(CP(X)) = 1$ such that every point in $CP(X)$ has exactly k distinct clothespinning sequences.*

Proof. Let $CP_0(X) = \{x \in X \mid \forall i \in \mathbb{Z} \exists j > i \text{ such that } x_i^j \notin \mathcal{U}\}$. The ergodic theorem tells us that since $\mu(X \setminus X_0) > 0$, μ -a.e. point must encounter words which are not MPW-minimal infinitely often. Thus μ -a.e. point is in $CP_0(X)$.

Let $x \in CP_0(X)$. Each new clothespin's construction only depends on the previous clothespin, so if s and s' are two distinct clothespinning sequences for x and $s_i = s'_i = 1$ then $s_j = s'_j$ for all $j > i$. Additionally, for any two clothespinning sequences s and s' , if i_1 and i_2 are adjacent clothespins in s , and j_1 and j_2 are adjacent in s' with $j_1 \leq i_1$, then $j_2 \leq i_2$. For if $j_2 > i_2$ then $x_{j_1}^{j_2}$ contains $x_{i_1}^{i_2+1}$ which is nonminimal. Thus s' has a pin between i_1 and i_2 , and the position of this pin determines all pins to the right. From this we can conclude that for any $n \in \mathbb{Z}$ the number of distinct clothespinning sequences over x_n^∞ is finite, and decreasing as $n \rightarrow \infty$.

Define $k_n(x)$ to be $|\{s_n^\infty \mid s \text{ is a clothespinning sequence of } x\}|$. The preceding explanation shows us that $k_n(x)$ is decreasing and $k_n(x) \geq 1$, so we can let $k(x) = \lim_{n \rightarrow \infty} k_n(x)$. Consider the set $A_m = \{x \in CP_0(X) \mid k_m(x) = k(x) \text{ and } k_{m-1}(x) > k(x)\}$. As we observe the clothespinning sequences of x going toward ∞ , we lose a sequence whenever two clothespins from previously distinct pinnings coincide. The set A_m represents the set of x whose ‘‘last coincident pin’’ occurs at time m . Note that if x is a point which is not in A_m for any $m \in \mathbb{Z}$ then x has a finite number of clothespinning sequences in total. However, A_m is not a set we can return to once we shift a point by T . So by the Poincaré recurrence theorem, $\mu(A_m) = 0$ for all $m \in \mathbb{Z}$. So μ -a.e. point in $CP_0(X)$ has finitely many distinct clothespinning sequences. Now, because μ is ergodic and $k(x)$ is shift invariant, $k(x)$ is equal to some constant k almost everywhere. In particular, μ -a.e. x has $k_n(x) = k$, so in a sense the number of clothespinning sequences we see locally is constant. Thus we can define $CP(X) = \{x \in CP_0(X) \mid k_n(x) = k \forall n\}$. \square

The next lemma is needed to ensure that we can build a measure on the clothespinned space which projects to a given measure on X .

Lemma 4.3.6. *Let $\mu \in \mathcal{M}(X)$ be an ergodic measure such that $\mu(X \setminus X_0) > 0$. Let \bar{X} be the set of pairs (x, s) where $x \in X$ and s is a clothespinning sequence for x and $\bar{T}(x, s)$ be the usual shift on x and s . Let $\pi_X(x, s) = x$. Then there exists a measure $\bar{\mu} \in \mathcal{M}(\bar{X})$ such that $\bar{\mu} \circ \pi_X^{-1} = \mu$.*

Proof. First let us note that by Lemma 4.3.5 we know the factor map π_X is finite-to-one. Let μ be a measure such that $\mu(X \setminus X_0) > 0$. By Lemma 4.3.5 we know that

there is a k such that μ -a.e. point in $CP(X)$ has exactly k clothespinning sequences.

Let $B_n(\Omega)$ be the words of length n in Ω . For every word $v \in B_n(\Omega)$ and $x \in CP(X)$ let us define $f_v(x) = |\{(x, s) \in \bar{X} \mid s_0^{n-1} = v\}|$. If $x \notin CP(X)$, just set $f_v(x) = 0$. In other words, $f_v(x)$ is the number of clothespinning sequences over x that “look like” v at the origin. If we let w be a word of length n in X then for any $x \in [w]$ we have $\sum_{v \in B_n(\Omega)} f_v(x) = k$. This is because each of the k clothespinning sequences of x appear in at least one of the v 's, and they cannot belong to more than one. Furthermore, we can see that for any $v \in B_n(\Omega)$, $f_{0v}(x) + f_{1v}(x) = f_{v0}(x) + f_{v1}(x) = f_v(x)$.

Let $C = ([w] \times [v])_i^{i+n-1}$ be a cylinder set of length n in \bar{X} . We will define $\bar{\mu}$ on such a cylinder by $\bar{\mu}(C) = \frac{1}{k} \int_{[w]_0^{n-1}} f_v(x) d\mu$. We should check the appropriate consistency conditions to ensure $\bar{\mu}$ extends to an invariant measure on \bar{X} . In particular, if $B = [w] \times [v]$ is a cylinder in \bar{X} , we can see that

$$\begin{aligned} \sum_{i \in \mathcal{A}(X)} \sum_{j \in \mathcal{A}(\Omega)} \bar{\mu}([wi] \times [vj]) &= \sum_{i \in \mathcal{A}(X)} \int_{[wi]} \sum_{j \in \mathcal{A}(\Omega)} f_{vj}(x) d\mu \\ &= \int_{[w]} f_v(x) d\mu \\ &= \bar{\mu}([w] \times [v]) \end{aligned}$$

Similarly, it is not hard to see that $\sum_{i \in \mathcal{A}(X)} \sum_{j \in \mathcal{A}(\Omega)} \bar{\mu}([iw] \times [jv]) = \bar{\mu}([w] \times [v])$. Together with the shift invariance of the clothespinning sequences, this tells us that $\bar{\mu}$ defines an invariant measure on \bar{X} . Noting that $\pi_X^{-1}([w]) = \bigcup_{v \in B_n(\Omega) \cap \bar{X}} [w] \times [v]$ we can then show that

$$\begin{aligned} \bar{\mu} \circ \pi_X^{-1}([w]) &= \frac{1}{k} \int_{[w]} \left(\sum_{v \in B_n(\Omega)} f_v(x) \right) d\mu \\ &= \frac{1}{k} \int_{[w]} k d\mu \\ &= \mu([w]) \end{aligned}$$

So $\bar{\mu}$ projects to μ . □

4.3.3 Proof of Theorem 4.3.1

Note that with a different choice of f , an essentially identical argument to the one which follows could be used to prove an analogous statement which reads: ... for all $p > 1$ there exists a function f which is q -Dini for all $q > p$, but which is not q -Dini for $q \leq p$ such that the relative equilibrium states of f ...

Let X_0 be a MPW subshift of X . Let \mathcal{U} be the set of words in X which are MPW-minimal.

In order to build a function whose relative equilibrium states live on X_0 we need a function which attains its maximum on X_0 and which penalizes us heavily for straying from it. However, the function cannot be too sharp, or it will fail to be p -Dini for some p near 1. Additionally, we would like for the function to rely on some condition which can be checked on a nice partition around the origin.

For $x \in X$, define $n(x)$ to be the least integer such that $x_{-n(x)} \dots x_{n(x)} \in \mathcal{U}$ but $x_{-n(x)-1} \dots x_{n(x)+1} \notin \mathcal{U}$. If $x \in X_0$ then define $n(x) = \infty$. This number gives us a substitute for having the distance from X_0 in our function.

If we want to phrase our function as $f(x) = g(n(x))$ and satisfy that f increases to 0 as $n \rightarrow \infty$ then the p -Dini condition becomes $\sum_n g(n)^p < \infty$. Our choice of function which satisfies that condition will be $f(x) = \frac{-t \log n(x)}{n(x)}$, with t to be determined later. This function also penalizes us heavily enough to beat the relative entropy which a relative equilibrium state could gain by straying from X_0 .

Let μ be an ergodic measure on X (a candidate relative equilibrium state) which is not supported on X_0 . We will show $h(\mu | \mu \circ \pi^{-1}) + \int f d\mu < 0$ to prove the relative equilibrium states live on X_0 . The reason this is sufficient is that for any measure μ' which is supported on X_0 , the fact that X_0 is finite-to-one over Y gives $h(\mu') = h(\mu' \circ \pi^{-1})$. Since $f|_{X_0} = 0$, we have $h(\mu') - h(\mu' \circ \pi^{-1}) + \int f d\mu' = 0$. This line of justification is similar to the one used by Walters to build continuous compensation functions in [22][Lemma 3.2].

First, we move to the clothespinned space \bar{X} and let $\bar{\mu}$ be the measure defined by Lemma 4.3.6. Our next step is to induce on the set $C = \{(x, s) \in \bar{X} \mid s_0 = 1\}$, in other words the clothespinnings with a pin at 0. It is clear that $\bar{\mu}(C) > 0$. We will call the induced system $(X_C, T_C, \mathcal{B}_C, \mu_C)$. Note that the action of T_C is to shift $(x, s) \in C$ so that the next clothespin in s is at the origin.

If we define $\bar{f} = f \circ \pi_X$ then $\int f d\mu = \int \bar{f} d\bar{\mu}$. Let $G(n_1, n_2, a, b, c)$ be the set of $(x, s) \in C$ such that n_1 and n_2 are the first two pins to the right of 0 with $x_0 = a$, $x_{n_1} = b$ and $x_{n_2} = c$. Let $G(n_1, n_2) = \bigcup_{a,b,c} G(n_1, n_2, a, b, c)$ and $p(n_1, n_2) = \mu_C(G(n_1, n_2))$. As these partitions indicate, we will be bounding the sum of \bar{f} up to the second return time and to this end we define $f_C^2(x, s) = \sum_{i=0}^{n_2-1} \bar{f}(\bar{T}^i(x, s))$.

If $(x, s) \in G(n_1, n_2)$ then we know $x_{-n_2} \dots x_{n_2} \notin \mathcal{U}$ because the word between 0 and n_2 is nonminimal. This allows us to say that $n(x) \leq n_2$, and thus $\bar{f}(x, s) \leq \frac{-t \log(n_2)}{n_2}$. In fact, for all $0 \leq i \leq n_2$ we have $n(T^i(x)) \leq n_2$ so we bound f_C^2 as

$$\begin{aligned} f_C^2(x, s) &\leq -t \sum_{i=0}^{n_2} \frac{\log n_2}{n_2} \\ &= -t \log n_2 \end{aligned}$$

So the bound obtained for the integral is

$$\int f_C^2 d\mu_C \leq \sum_{n_1, n_2} -p(n_1, n_2) B(n_1, n_2)$$

where $B(n_1, n_2) = -t \log n_2$.

To estimate the relative entropy term, let us first recall that because π_X is finite-to-one, $h(\mu \mid \mu \circ \pi^{-1}) = h(\bar{\mu} \mid \bar{\mu} \circ (\pi \circ \pi_X)^{-1})$. So it makes sense to work on bounding the relative entropy in the clothespinned space. Let \mathcal{Q} be σ -algebra on \bar{X} given by knowing $\pi \circ \pi_X(x, s)_{-\infty}^{\infty}$. Let \mathcal{Q}_C be the elements of \mathcal{Q} intersected with C . Then Theorem 4.3.3 tells us that

$$\begin{aligned} h(\bar{\mu} \mid \bar{\mu} \circ (\pi \circ \pi_X)^{-1}) &= h(\bar{\mu} \mid \mathcal{Q}) \\ &= \frac{1}{2} \bar{\mu}(C) h_{T_C^2}(\mu_C \mid \mathcal{Q}_C) \end{aligned}$$

If we know $\pi \circ \pi_X(x, s)$ then learning $(x, s)_0^{n_2}$ is the same as learning n_1, n_2, x_0, x_{n_1} and x_{n_2} . So if we let \mathcal{P} be the $G(n_1, n_2, a, b, c)$ partition we can say

$$h_{T_C^2}(\mu_C \mid \mathcal{Q}_C) = h(T_C^2, \mathcal{P}) \leq \sum_{n_1, n_2} \sum_{a, b, c} -p(n_1, n_2, a, b, c) \log p(n_1, n_2, a, b, c)$$

We want this estimate to be finite to reasonably combine it with the integral bound. Let $p(n) = \sum_{n_2=n} p(n_1, n_2)$ and $d = |\mathcal{A}(X)|^3$. We will again use the version

of Jensen's inequality which appeared in the proof of Theorem 4.3.3. In fact, this estimate mirrors the one given in that proof quite closely. The inequality will be applied twice, first with a_i as choices of (a, b, c) , then as choices of (n_1, n_2) . Both times $b_i = 1$.

$$\begin{aligned} \sum_{n_1, n_2} \sum_{a, b, c} -p(n_1, n_2, a, b, c) \log p(n_1, n_2, a, b, c) &\leq \sum_n \sum_{n_2=n} -p(n_1, n_2) \log \frac{p(n_1, n_2)}{d} \\ &\leq \sum_n \left(-p(n) \log \frac{p(n)}{n-1} + p(n) \log d \right) \\ &= \sum_n (-p(n) \log p(n) + p(n) \log(n-1) + p(n) \log d) \end{aligned}$$

We will deal with the finiteness of the three terms in this sum separately. The sum of the third terms is clearly finite. From Lemma 4.3.2 we know the first return times have finite expectation, and thus so do the second return times. This allows us to say $\sum_n np(n) < \infty$, which implies the second set of terms $\sum_n p(n) \log(n-1) < \infty$. The first term can be dealt with by first defining $A = \{n \mid p(n) \leq \frac{1}{n^2}\}$. Then we have

$$\sum_{n \in A} -p(n) \log p(n) < \sum_{n \in A} \frac{2 \log n}{n^2} < \infty$$

When $n \in A^c$,

$$\begin{aligned} \log p(n) &> -2 \log n \\ -p(n) \log p(n) &< 2p(n) \log n \\ \sum_{n \in A^c} -p(n) \log p(n) &< \sum_{n \in A^c} 2p(n) \log n \end{aligned}$$

Again from Lemma 4.3.2 this last sum is finite, and thus the estimate on the relative entropy is finite.

Combining the two bounds we arrive at an inequality of the form

$$\begin{aligned} \frac{2}{\bar{\mu}(C)} \left\{ h(\bar{\mu} \mid \bar{\mu} \circ (\pi \circ \pi_X)^{-1}) + \int \bar{f} d\bar{\mu} \right\} &\leq \\ \sum_{n_1, n_2} \sum_A B(n_1, n_2) p(n_1, n_2, A) - p(n_1, n_2, A) \log p(n_1, n_2, A) & \end{aligned}$$

If we define $a(n_1, n_2) = e^{B(n_1, n_2)} = \left(\frac{1}{n_2}\right)^t$ then for a suitable value of t these terms are summable, and in fact their sum is as small as we like. Assume that t is chosen to make these summable. For simplicity in the following argument, we will temporarily reindex (n_1, n_2, A) to n . Define $C = \sum_n a_n$. We are seeking to show $\sum_n p_n \log \frac{a_n}{p_n}$ is maximized when $\frac{a_n}{p_n} = C$ for all n . The tangent line to \log at C is given by $L_C(x) = \log(C) - \frac{1}{C}(x - C)$. Note that because $\log(x)$ is concave, the tangent at C lies entirely above $\log(x)$. So, taking $x = \frac{a_n}{p_n}$ gives us the inequality $\frac{a_n}{p_n C} - 1 + \log(C) \geq \log \frac{a_n}{p_n}$. Using this Jensenesque reasoning allows us to show

$$\begin{aligned} \sum_n p_n \log \frac{a_n}{p_n} &\leq \sum_n p_n \left(\frac{a_n}{p_n C} - 1 + \log(C) \right) \\ &= \log(C) \end{aligned}$$

When $\frac{a_n}{p_n} = C$ this bound is obtained. Indexing back to (n_1, n_2, A) we obtain the following inequality.

$$\sum_{n_1, n_2} \sum_A B(n_1, n_2) p(n_1, n_2, A) - p(n_1, n_2, A) \log p(n_1, n_2, A) \leq \log \left(\sum_{n_1, n_2} \left(\frac{1}{n_2} \right)^t \right)$$

Through adjusting t we can ensure that this bound is negative, and thus the theorem is proved. \square

Chapter 5

Outlook

We have discussed the fact that functions with summable variation (the 1-Dini functions) have unique equilibrium states (see for example [20]). In [1], the following result was shown for measures of relative maximal entropy:

Theorem 5.0.7. *Let (X, Y, π) be a factor triple with X a shift of finite type. Let $\nu \in \mathcal{M}(Y)$ be a fully supported ergodic measure. There is a computable, conjugacy invariant number c_π called the **class degree** of the factor triple, such that the number of ergodic measures of relative maximal entropy over ν is at most c_π .*

The class degree c_π is related to the degree of a finite-to-one code, and many properties which are true for the degree carry over to class degree. We now know that the relative equilibrium states of a 1-Dini function are fully supported, and we conjecture that this bound applies to these measures as well.

Conjecture 5.0.8. *Let (X, Y, π) be a factor triple with X a shift of finite type. Let $\nu \in \mathcal{M}(Y)$ be a fully supported ergodic measure and $f \in C(X)$ a 1-Dini function. Then there are at most c_π ergodic relative equilibrium states of f over ν , where c_π is the class degree.*

The general structural theory of infinite-to-one factor maps for shifts of finite type has been improved greatly over the last few years. The importance of the Marcus-Petersen-Williams construction in this work leads one to wonder how it fits into this structural picture. Class degree is derived from the notion of the number of communicating classes over Y , sometimes called the transition classes. It would

be desirable to understand if for a clever choice of our MPW-ordering, the resulting MPW-subshift bears any significance for the transition classes.

Chapter 6

Bibliography

- [1] Mahsa Allahbakhshi and Anthony Quas. Class degree and relative maximal entropy. *Trans. Amer. Math. Soc.*, 365(3):1347–1368, 2013.
- [2] Mike Boyle and Karl Petersen. Hidden Markov processes in the context of symbolic dynamics. In *Entropy of hidden Markov processes and connections to dynamical systems*, volume 385 of *London Math. Soc. Lecture Note Ser.*, pages 5–71. Cambridge Univ. Press, Cambridge, 2011.
- [3] Mike Boyle and Selim Tuncel. Infinite-to-one codes and Markov measures. *Trans. Amer. Math. Soc.*, 285(2):657–684, 1984.
- [4] F. R. Gantmacher. *The theory of matrices. Vols. 1, 2*. Translated by K. A. Hirsch. Chelsea Publishing Co., New York, 1959.
- [5] P. Góra and B. Schmitt. Un exemple de transformation dilatante et C^1 par morceaux de l'intervalle, sans probabilité absolument continue invariante. *Ergodic Theory Dynam. Systems*, 9(1):101–113, 1989.
- [6] Franz Hofbauer. Examples for the nonuniqueness of the equilibrium state. *Trans. Amer. Math. Soc.*, 228(223–241.), 1977.
- [7] Robert B. Israel. *Convexity in the theory of lattice gases*. Princeton University Press, Princeton, N.J., 1979. Princeton Series in Physics, With an introduction by Arthur S. Wightman.
- [8] Gerhard Keller. *Equilibrium states in ergodic theory*, volume 42 of *London Mathematical Society Student Texts*. Cambridge University Press, Cambridge, 1998.

- [9] François Ledrappier and Peter Walters. A relativised variational principle for continuous transformations. *J. London Math. Soc. (2)*, 16(3):568–576, 1977.
- [10] Douglas Lind and Brian Marcus. *An introduction to symbolic dynamics and coding*. Cambridge University Press, Cambridge, 1995.
- [11] Brian Marcus, Karl Petersen, and Susan Williams. Transmission rates and factors of Markov chains. In *Conference in modern analysis and probability (New Haven, Conn., 1982)*, volume 26 of *Contemp. Math.*, pages 279–293. Amer. Math. Soc., Providence, RI, 1984.
- [12] William Parry. *Entropy and generators in ergodic theory*. W. A. Benjamin, Inc., New York-Amsterdam, 1969.
- [13] Yuval Peres and Anthony Quas. Entropy rate for hidden Markov chains with rare transitions. In *Entropy of hidden Markov processes and connections to dynamical systems*, volume 385 of *London Math. Soc. Lecture Note Ser.*, pages 172–178. Cambridge Univ. Press, Cambridge, 2011.
- [14] Karl Petersen. *Ergodic theory*, volume 2 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1989.
- [15] Karl Petersen, Anthony Quas, and Sujin Shin. Measures of maximal relative entropy. *Ergodic Theory Dynam. Systems*, 23(1):207–223, 2003.
- [16] Karl Petersen and Sujin Shin. On the definition of relative pressure for factor maps on shifts of finite type. *Bull. London Math. Soc.*, 37(4):601–612, 2005.
- [17] Daniel J. Rudolph. *Fundamentals of measurable dynamics*. Oxford Science Publications. The Clarendon Press Oxford University Press, New York, 1990. Ergodic theory on Lebesgue spaces.
- [18] David Ruelle. *Thermodynamic formalism*. Cambridge Mathematical Library. Cambridge University Press, Cambridge, second edition, 2004. The mathematical structures of equilibrium statistical mechanics.
- [19] Sujin Shin. Measures that maximize weighted entropy for factor maps between subshifts of finite type. *Ergodic Theory Dynam. Systems*, 21(4):1249–1272, 2001.

- [20] Peter Walters. Ruelle's operator theorem and g -measures. *Trans. Amer. Math. Soc.*, 214:375–387, 1975.
- [21] Peter Walters. *An introduction to ergodic theory*, volume 79 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1982.
- [22] Peter Walters. Relative pressure, relative equilibrium states, compensation functions and many-to-one codes between subshifts. *Trans. Amer. Math. Soc.*, 296(1):1–31, 1986.
- [23] Yuki Yayama. Existence of a measurable saturated compensation function between subshifts and its applications. *Ergodic Theory Dynam. Systems*, 31(5):1563–1589, 2011.
- [24] Jisang Yoo. Measures of maximal relative entropy with full support. *Ergodic Theory Dynam. Systems*, 31(6):1889–1899, 2011.