THE UNIVERSITY OF WARWICK

AUTHOR: **Thomas Kempton** DEGREE: **Ph.D.**

TITLE: **Thermodynamic Formalism for Symbolic Dynamical Systems**

DATE OF DEPOSIT: ....

I agree that this thesis shall be available in accordance with the regulations governing the University of Warwick theses.

I agree that the summary of this thesis may be submitted for publication.

I **agree** that the thesis may be photocopied (single copies for study purposes only).

Theses with no restriction on photocopying will also be made available to the British Library for microfilming. The British Library may supply copies to individuals or libraries. subject to a statement from them that the copy is supplied for non-publishing purposes. All copies supplied by the British Library will carry the following statement:

> "Attention is drawn to the fact that the copyright of this thesis rests with its author. This copy of the thesis has been supplied on the condition that anyone who consults it is understood to recognise that its copyright rests with its author and that no quotation from the thesis and no information derived from it may be published without the author's written consent."

AUTHOR'S SIGNATURE: ....

USER'S DECLARATION

1. I undertake not to quote or make use of any information from this thesis without making acknowledgement to the author.
2. I further undertake to allow no-one else to use this thesis while it is in my care.

DATE SIGNATURE ADDRESS

....

....

....

....

....

www.warwick.ac.uk

Thermodynamic Formalism for Symbolic Dynamical Systems

by

Thomas Kempton

Thesis

Submitted to the University of Warwick

for the degree of

Doctor of Philosophy

Mathematics Department

February 2011

THE UNIVERSITY OF WARWICK

Contents

Acknowledgments **iv**

Declarations **v**

Abstract **vi**

Chapter 1 Introduction **1**

Chapter 2 Preliminaries **7**

2.1 Topological Markov Shifts 9

2.1.1 Suspension Flows 11

2.2 Thermodynamic Formalism 12

2.2.1 Metric Entropy 12

2.2.2 Topological Entropy and Pressure 14

2.2.3 Gibbs Measures and Equilibrium States 17

2.2.4 Coboundaries 19

Chapter 3 Counting β-Expansions **21**

3.1 Introduction 21

3.2 Generating β-expansions 23

3.2.1 A More General Method Including $\beta > 2$. 24

3.3 Proof of Theorem 3.1.1 26

3.4 Does There Exist a Growth Rate? 30

Chapter 4 Topological Pressure for Suspension Flows over Countable Markov Shifts **32**

4.1 Introduction 32

4.2 Preliminaries 35

4.2.1 Suspension Flows 35

4.3 An Analogue of Gurevich Pressure for Suspension Flows 42

4.4 Compact Invariant Subsets 48

4.5 The Definition of Barreira and Iommi 51

4.6 A Variational Principle for P_ϕ 53

4.7 Equilibrium States 66

4.8 Applications to the Positive Geodesic Flow 68

Chapter 5 Zero Temperature Limit Laws **70**

5.1 Introduction 70

5.2 Set Up 73

5.3 Recasting the Question 76

5.4 A Related Finite Markov Shift 83

5.4.1 Convergence 85

5.5 Proof of Technical Lemma 90

Chapter 6 Factors of Gibbs Measures for Subshifts of Finite Type **96**

6.0.1 Preliminaries and Technical Hypotheses 97

6.0.2 Results 100

6.1 Defining the Potential ψ_2 101

6.2 Proof that ψ_2 is Well Defined 107

6.3 Regularity of the Potential ψ_2 116

6.4 Examples and Comments 119
6.5 Comments on the Technical Hypothesis 122
6.6 Renormalization of Gibbs Measures in Statistical Mechanics 124

Acknowledgments

I would like to thank my supervisor Mark Pollicott for his excellent supervision during the last few years. Mark has been incredibly generous with his time and has provided me with a wide range of interesting and accessible open problems to work on.

I would also like to thank Omri Sarig, Oliver Jenkinson, Charlene Kalle, Dan Thompson, Ian Morris, Nikita Sidorov, Andy Ferguson and Evgeny Verbitskiy for some useful discussions and suggestions.

Finally, I would like to thank my family and Aspasia Bisopoulou for their love and support.

This work was supported by the EPSRC.

Declarations

I declare that the work in this thesis is, to the best of my knowledge, original and my own work, except where otherwise indicated, cited, or commonly known. This work has not been submitted for any other degree.

An article based on the work of chapter 5 is to appear in the Journal of Statistical Physics.

The work of chapter 6 has been written up as two articles. The first of these, [KP11], deals with the special case of factors of full shifts and is joint work with Mark Pollicott. It has been accepted for publication in *Entropy of Hidden Markov Processes and Connections to Dynamical Systems, papers from the Banff International Research Station Workshop, October 2007.* The more general case of factors of subshifts of finite type is independent work and has been published in the Bulletin of the London Mathematical Society, [Kem11]. The exposition of chapter 6 mainly follows that of [Kem11], although many of the technical lemmas of [KP11] are included, and thus this chapter contains both individual work and work joint with Mark Pollicott.

I have submitted a further article for publication based on the ideas of chapter 4.

Abstract

We derive results in the ergodic theory of symbolic dynamical systems.

Our first result concerns β-expansions of real numbers. We show that for a fixed non-integer $\beta > 1$ and a fixed real number $x \in \left[0, \frac{\lfloor \beta \rfloor}{\beta - 1}\right]$, the number of words $(x_1, \cdots, x_n)$ that can be extended to β-expansions of x grows at least exponentially in n.

Our second result concerns definitions of topological pressure for suspension flows over countable Markov shifts. Previously, topological pressure had been considered for a restricted class of suspension flows upon which the thermodynamic formalism can be well understood using the base transformation. We consider a more general class of suspension flows and show the equivalence of several natural definitions of topological pressure, including a definition analogous to that of Gurevich pressure for a Markov shift.

Our third result concerns zero temperature limit laws for countable Markov shifts. We show that for a uniformly locally constant potential f on a topologically mixing countable Markov shift satisfying the big images and preimages property, the equilibrium states μ_{tf} associated to the potential tf converge as t tends to infinity.

Finally we consider the image under a one-block factor map Π of a Gibbs measure μ supported on a finite alphabet Markov shift. We give sufficient conditions on Π for the image measure $\Pi^*(\mu)$ to be a Gibbs measure and discuss regularity properties of the potential associated to $\Pi^*(\mu)$ in terms of the regularity of the potential associated to μ.

Chapter 1

Introduction

The work in this thesis concerns ergodic theory for symbolic dynamical systems. Basic definitions and theorems relevant to the work are given in chapter 2 and results are presented in chapters 3 to 6, each of which discusses a different problem and can be read independently of the others. In this chapter we give a brief outline of the questions considered in the work. Further introduction and motivation can be found at the beginning of each chapter.

Counting β-expansions:

In chapter 3 we discuss β-expansions of real numbers. Chapter 3 is rather different in nature to the other chapters in this work as it does not concern thermodynamic formalism. Furthermore, while the question that we answer is one about a symbolic space, the space is not Markov and our method of study is not typical of the methods of symbolic dynamical systems. It does however serve to highlight the link between symbolic and non-symbolic dynamical systems, providing further motivation for the study of Markov shifts later in the thesis.

Given a real number $\beta > 1$, a β-expansion of $x \in \left[0, \frac{\lfloor\beta\rfloor}{\beta-1}\right]$ is a sequence $(x_n)_{n=1}^{\infty} \in$

$\{0, 1, \cdots, \lfloor\beta\rfloor\}^{\mathbb{N}}$ such that

$$x = \sum_{n=1}^{\infty} x_n \beta^{-n}.$$

For non-integer $\beta > 1$ and typical $x \in [0, \frac{\lfloor\beta\rfloor}{\beta-1}]$ there are uncountably many β-expansions of x. In [SF], Feng and Sidorov defined $\mathcal{N}_n(x;\beta)$ to be the number of words of length n that can be extended to β-expansions of x, and proved that, for $\beta < \dfrac{1+\sqrt{5}}{2}$, there exists a positive constant $c > 1$ such that

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathcal{N}_n(x;\beta) \geq c.$$

We extend the above result to all non-integer $\beta > 1$, giving a positive answer to a question posed in that paper.

Thermodynamic Formalism and Symbolic Dynamical Systems:

The work of chapters 4, 5 and 6 concerns thermodynamic formalism for symbolic dynamical systems. In chapters 5 and 6 we consider Markov shifts, which provide symbolic models for a wide variety of dynamical systems including hyperbolic automorphisms of the torus, certain billiard maps and the Gauss map. In chapter 4 we consider suspension flows over Markov shifts, which provide models for continuous dynamical systems such as the geodesic flow on the modular surface and the Teichmüller flow. Through studying these symbolic models one is able to gain a great deal of understanding about the original dynamical systems being modelled.

Ergodic theory has its origins in statistical mechanics and the study of the long term behaviour of systems of large numbers of particles. In such systems precise computation of the behaviour of each particle may be unfeasible, but through the ergodic theorems one is able to gain an understanding of the long term behaviour of a typical point and link the macroscopic behaviour of the system with the microscopic laws governing individual particles. When we refer here to a typical point, we mean almost every point with respect to some suitable measure invariant under the transformation, but this leads to the question, with respect to which measure should one use the ergodic theorem? The empirical data available to physicists led them to the conclusion that the Gibbs measure is the most suitable such measure.

In the 1950s, Ruelle and Sinai translated the idea of the Gibbs measure to the setting of dynamical systems. The body of research based around this idea became known as thermodynamic formalism. Thermodynamic formalism has been used to great effect in the study of dynamical systems, for example in understanding the behaviour of hyperbolic flows, where through symbolic dynamics techniques one is able to construct Gibbs measures and prove various results such as exponential decay of correlations.

Topological Pressure for Suspension Flows over Countable Markov Shifts:

The notion of topological pressure, which is of crucial importance in the development of thermodynamic formalism, is well understood for topologically mixing flows and transformations on compact spaces. It has several equivalent definitions which, in general, are no longer equivalent if the underlying space is non-compact, and so there exist various different notions of pressure for flows and transformations on non-compact spaces. In the case of topologically mixing countable Markov shifts (Σ, σ), Sarig proved in [Sar99] that the following notions are equivalent.

$$\begin{aligned} P_\sigma(g) &:= \lim_{n\to\infty} \frac{1}{n} \log \left(\sum_{\sigma^n(\underline{x})=x} \exp(g^n(\underline{x}))\chi_{[a]}(\underline{x}) \right) \\ &= \sup\{h_\mu(\sigma) + \int g d\mu | \mu \in \mathcal{M}_\sigma, \int g d\mu > -\infty\} \\ &= \sup\{P_\sigma(g|K) |\ K \text{ is a compact invariant subset of } \Sigma\}, \end{aligned}$$

where $\mathcal{M}_\sigma$ is the set of σ invariant measures on the Markov shift Σ, $g^n(\underline{x}) := \sum_{k=0}^{n-1} g(\sigma^k(\underline{x}))$ and a is allowed to be any letter of $\mathcal{A}$. The choice of a does not affect $P_\sigma(g)$. P_σ is known as Gurevich pressure.

The results in [Sar99] have led to a large volume of work studying the thermodynamic properties of transformations on non-compact spaces that can be modelled by countable Markov shifts.

In chapter 4 we introduce a natural analogue of Gurevich pressure for suspension flows over countable Markov shifts. Previously a definition of topological entropy for suspension flows over countable Markov shifts with locally constant roof function was given by Savchenko in [Sav98]. Subsequently, a definition of topological pressure was given by Barreira and Iommi in [BI06] for suspension flows with Hölder continuous roof functions that are bounded away from zero. We extend the definitions of Savchenko and of Barreria and Iommi to a wider class of suspension flows, and show their equivalence to our generalised Gurevich pressure.

We prove that, for a suspension flow (Σ_f, ϕ) over base Σ with roof function $f : \Sigma \to \mathbb{R}^+$ and for suitable conditions on f and g, the following notions of topological pressure are equivalent.

$$\begin{aligned}
P_\phi(g) &:= \lim_{t\to\infty} \frac{1}{t} \log \left(\sum_{\phi_s(\underline{x},0)=(\underline{x},0), s\le t} \exp\left(\int_0^s g(\phi_k(\underline{x},0))dk \right) \chi_{[a]}(\underline{x}) \right) \\
&= \sup_{K\in\mathcal{K}_{\Sigma_f}} P_\phi(g|K) \\
&= \inf\{t\in\mathbb{R} : P_\sigma(\Delta_g - tf) \le 0\} = \sup\{t\in\mathbb{R} : P_\sigma(\Delta_g - tf) \ge 0\} \\
&= \sup\{h_\nu(\phi) + \int g d\nu : \nu \in \mathcal{E}_\phi, \int g d\nu > -\infty\}
\end{aligned}$$

where a is any element of $\mathcal{A}$, $\mathcal{E}_\phi$ is the set of ergodic flow invariant probability vectors on Σ_f and $\mathcal{K}_{\Sigma_f}$ is the set of compact flow invariant subsets of Σ_f.

As an application we consider the entropy of the positive geodesic flow on the modular surface.

Zero Temperature Limit Laws:

In chapter 5 we consider Gibbs measures on a countable Markov shift Σ. Under suitable conditions on f and Σ, there exists a unique Gibbs measure μ_{tf} associated to the potential tf for each $t > 1$. Then given a sequence $t_n \to \infty$, one can ask what happens to the sequence $\mu_{t_n f}$. In statistical mechanics this corresponds to studying a system of particles at temperature $\frac{1}{t_n}$ as $t_n \to \infty$, and so limit points of the sequence $\mu_{t_n f}$ are referred to as zero temperature limits. Zero temperature limit laws are also of relevance to ergodic optimisation, since any limit point μ of the sequence $\mu_{t_n f}$ will be a maximising measure for f.

We prove that, given a uniformly locally constant potential $f : \Sigma \to \mathbb{R}$ on a countable Markov shift and suitable conditions of f and Σ to ensure the existence of the Gibbs measures μ_{tf}, the sequence $\mu_{t_n f}$ converges in the weak* topology for any sequence $t_n \to \infty$.

Factors of Gibbs measures:

There are many natural situations in which one is required to study factors of Markov shifts. For example, if a Markov system is subject to imperfect observation under which two or more states are indistinguishable, then one observes only some factor transformation on the space of equivalence classes of indistinguishable states. This observed transformation may no longer be Markov. If the original transformation preserved an invariant Gibbs measure then it may be natural to study the properties of the observed transformation with respect to the original Gibbs measure projected on to our factor space.

In chapter 6 we consider the images under factor maps Π of Gibbs measures supported on finite alphabet Markov shifts. We give sufficient conditions on Π for the image measure to be a Gibbs measure, and discuss the regularity of the potential associated to the image measure in terms of the regularity of the potential associated to the original measure. We also give an example of a mapping which does not satisfy our conditions and for which the image measure is not a Gibbs measure. This generalises work by Chazottes and Ugalde in [CU03] and [CU11], and by Verbitskiy in [Ver11].

Chapter 2

Preliminaries

In this chapter we introduce some basic definitions and theorems for dynamical systems and ergodic theory. Given a space X, we will be interested in the behaviour of transformations T and flows ϕ on X. A flow $\phi : X \times \mathbb{R} \to X$ is a function such that for each $t \in \mathbb{R}$, $\phi_t(x) := \phi(x, t)$ is a transformation on X. Flows must also be continuous in t and satisfy $\phi_0(x) = x$ and $\phi_{s+t}(x) = \phi_s(\phi_t(x))$ for all $s, t \in \mathbb{R}$ and $x \in X$. Pairs (X, T) and (X, ϕ) will be called dynamical systems.

We let the triple $(X, \mathcal{B}, \mu)$ denote a space X, the σ-algebra $\mathcal{B}$ of measurable subsets of X, and a measure μ on X.

Let $T : X \to X$ be a transformation. Given a set $A \in \mathcal{B}$ we define the set $T^{-1}(A) := \{x \in X | T(x) \in A\}$. T is said to be measurable if $T^{-1}(A) \in \mathcal{B}$ for all $A \in \mathcal{B}$. T is said to preserve measure μ if $\mu(T^{-1}(A)) = \mu(A)$ for all $A \in \mathcal{B}$. The set of all measures μ invariant under T is denoted $\mathcal{M}_T$. For transformations T preserving a finite measure μ we have the following theorem.

Theorem 2.0.1 (Poincaré recurrence theorem). *Let $T : (X, \mathcal{B}, \mu) \to (X, \mathcal{B}, \mu)$ be a measure preserving transformation with $\mu(X) < \infty$ and suppose that $A \in \mathcal{B}$ has $\mu(A) > 0$. Then for almost every $x \in A$, $T^n(x) \in A$ for infinitely many $n \in \mathbb{N}$.*

A measure preserving transformation $T : (X, \mathcal{B}, \mu) \to (X, \mathcal{B}, \mu)$ is called conservative if for any set A with $\mu(A) > 0$ and for almost every $x \in A$ there exists an $n \in \mathbb{N}$ such that $T^n(x) \in A$. Any transformation preserving a finite measure is automatically conservative by the Poincaré recurrence theorem.

A measure preserving transformation $T : (X, \mathcal{B}, \mu) \to (X, \mathcal{B}, \mu)$ is said to be ergodic if for all $A \in \mathcal{B}$ with $T^{-1}(A) = A$ we have $\mu(A) = 0$ or $\mu(A^c) = 0$. Perhaps the most famous result of ergodic theory is the Birkhoff ergodic theorem:

Theorem 2.0.2 (Birkhoff ergodic theorem). *Let $T : (X, \mathcal{B}, \mu) \to (X, \mathcal{B}, \mu)$ be an ergodic measure preserving transformation with $\mu(X) = 1$. Then for all $f \in \mathcal{L}^1(\mu)$,*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(T^k(x)) = \int_X f d\mu$$

for almost every x (with respect to μ).

There are alternative statements of the theorem that do not require T to be ergodic or $\mu(X)$ to be finite, but we shall use only this standard form. Sometimes we will require the following two topological notions.

A measurable transformation $T : (X, \mathcal{B}) \to (X, \mathcal{B})$ is said to be topologically transitive if for all open sets $A, B \in \mathcal{B}$ there exists an $n \in \mathbb{N}$ such that

$$T^{-n}(A) \cap B \neq \phi.$$

T is said to be topologically mixing if for all open sets $A, B \in \mathcal{B}$ there exists an $N \in \mathbb{N}$ such that for all $n > N$ we have

$$T^{-n}(A) \cap B \neq \phi.$$

Clearly if a transformation is topologically mixing then it is also topologically transitive.

2.1 Topological Markov Shifts

Topological Markov shifts are symbolic dynamical systems which are useful models for various other dynamical systems. In this section we define them formally and introduce some structure on the space Σ.

Definition 2.1.1. *Given a finite or countable alphabet $\mathcal{A} = \{1, \cdots, k\}$ or $\mathbb{N}$ and a matrix M of zeros and ones with rows and columns indexed by $\mathcal{A}$, we define the one sided topological Markov shift (Σ, σ) to be the shift space*

$$\Sigma := \{\underline{x} = (x_i)_{i=0}^{\infty} \in \mathcal{A}^{\mathbb{Z}_+} : M_{x_i x_{i+1}} = 1 \forall i \in \mathbb{Z}_+\}$$

together with the transformation $\sigma : \Sigma \to \Sigma$, $\sigma(x_0 x_1 \cdots) = (x_1 x_2 \cdots)$.

We call a finite word $x_m \cdots x_n$ admissible if $M_{x_i x_i+1} = 1$ for $i \in \{m, \cdots, n-1\}$. Given an admissible word $x_m \cdots x_n$ we define the cylinder set $[x_m \cdots x_n]$ to be the set of sequences $\underline{y} = (y_i)_{i=1}^{\infty} \in \Sigma$ satisfying $y_m \cdots y_n = x_m \cdots x_n$. We sometimes write $x_m \cdots x_n \in \Sigma$ to mean $x_m \cdots x_n$ is an admissible word, but it will always be clear whether we are discussing infinite sequences or finite admissible words.

We define a metric on Σ by $d(\underline{x}, \underline{y}) = 2^{-\inf\{n \in \mathbb{Z}_+ : x_n \neq y_n\}}$.

The metric d defines a topology on Σ. The σ-algebra of open sets is generated by the set of cylinder sets. Σ is compact if $\mathcal{A}$ is finite and non-compact if $\mathcal{A}$ is infinite.

We further define the n-th variation of a function $\psi : \Sigma \to \mathbb{R}$ by

$$var_n(\psi) = \sup\{|\psi(\underline{x}) - \psi(\underline{y})| : \underline{x}, \underline{y} \in \Sigma, x_0 \cdots x_{n-1} = y_0 \cdots y_{n-1}\} \text{ for } n \geq 1$$

and $var_0(\psi) = \sup\{|\psi(\underline{x}) - \psi(\underline{y})| : \underline{x}, \underline{y} \in \Sigma\}$.

For a continuous function ψ we have $\lim_{n\to\infty} var_n(\psi) = 0$. The speed of this convergence gives us the regularity of the function. In particular a function $\psi : \Sigma \to \mathbb{R}$ is Hölder continuous if there exist constants $c > 0$ and $\theta \in (0, 1)$ such that $var_n(\psi) < c\theta^n$ for all $n \geq 0$, and is called weakly Hölder continuous if $var_n(\psi) < c\theta^n$ for all $n \geq 1$.

There are two commonly used definitions of summable variation for a function $\psi : \Sigma \to \mathbb{R}$ which are not equivalent if $\mathcal{A}$ is not finite. The first convention defines ψ to have have summable variation if $\sum_{n=1}^{\infty} var_n(\psi) < \infty$, while the second requires the extra condition that $var_0(\psi) < \infty$, i.e. that ψ is bounded. We follow the first convention, and should we require a particular function of summable variation to also be bounded we shall state so explicitly.

One sided Markov shifts are not invertible and so we define two sided Markov shifts, which are their natural extension.

Definition 2.1.2. *Given $\mathcal{A}$ and M as in the definition of one sided Markov shifts, we define the two sided Markov shift associated to M as the set of sequences $\{\underline{x} \in \mathcal{A}^{\mathbb{Z}}$ such that $M_{x_i x_{i+1}} = 1$ for all $i \in \mathbb{Z}\}$ together with the shift operator $\sigma((x_i)_{i\in\mathbb{Z}}) = (x_{i+1})_{i\in\mathbb{Z}}$.*

In the case of two sided Markov shifts (Σ, σ) and $\psi : \Sigma \to \mathbb{R}$, we define $d(\underline{x}, \underline{y}) = 2^{-\inf\{n\in\mathbb{Z}_+ : x_{-n}\cdots x_n \neq y_{-n}\cdots y_n\}}$ and

$$var_n(\psi) = \sup\{|\psi(\underline{x}) - \psi(\underline{y})| : \underline{x}, \underline{y} \in \Sigma, x_{-(n-1)} \cdots x_{n-1} = y_{-(n-1)} \cdots y_{n-1}\}.$$

Summable variation and Hölder continuity are defined using the metric d as with the one sided shift.

2.1.1 Suspension Flows

In the study of continuous time dynamical systems ψ on a space X, it is often useful to take a Poincaré section $A \subset X$ and study the properties of the induced transformation on A. We can define $A_\infty := \{x \in A : \psi_t(x) \in A$ for infinitely many $t > 0\}$ and $T : A_\infty \to A_\infty$ by $T(x) = \psi_t(x)$ where $t = f(x) = \inf\{s > 0 : \phi_s(x) \in X_0\}$, which is always finite. Intelligent choices of A may yield a comparatively simple induced transformation from a complicated flow, and many properties of the flow can be inferred from properties of T. However since T does not tell us how long it took to flow from x to $T(x)$, certain properties of the flow, such as return time statistics, cannot be studied purely through the study of T. To this end, we define the space

$$A_f := \{(x, y) : x \in A_\infty, 0 \leq y \leq f(x)\}.$$

We define the flow ϕ on A_f by

$$\phi_t(x, y) = (x, y + t)$$

for $y + t \in [0, f(x))$ and extend this to a flow for all time t using the identification $(x, f(x)) = (T(x), 0)$. The flow ϕ is called the suspension flow over A with roof function f. The study of the suspension flow ϕ may allow us to prove results about the original flow ψ.

2.2 Thermodynamic Formalism

The ergodic theorem gives a good description of the behaviour of a transformation T with respect to an ergodic invariant measure μ. However there are many further questions that we can ask. Is there a rate of convergence in the ergodic theorem? With respect to which invariant measure is it most natural to apply the ergodic theorem in order to measure the long term behaviour of T? What can be said topologically about the set of points for which the ergodic averages $\lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1} f(T^k(x))$ do not converge to $\int f d\mu$? The notions of entropy, pressure and the Gibbs measure, which are generalisations of concepts from statistical mechanics, have been of crucial importance in developing answers to these questions. The body of research studying the properties of dynamical systems using these notions is termed thermodynamic formalism. We introduce some key ideas from thermodynamic formalism for use later, more comprehensive introductions can be found in [Wal82, Sar].

2.2.1 Metric Entropy

Kolmogorov-Sinai entropy, or metric entropy, was introduced by Kolmogorov in 1958 as a measure of the complexity of a transformation $T : X \to X$ with respect to some invariant measure μ. The definition that we give is a refinement by Sinai of Kolmogorov's original definition.

Definition 2.2.1. *Let T be a measure preserving transformation of the finite measure space $(X, \mathcal{B}, \mu)$ and $\mathcal{A} = \{A_1, \cdots, A_k\}$ be a finite measurable partition of X. We define the entropy of the partition $\mathcal{A}$ by*

$$H(\mathcal{A}) = -\sum_{i=1}^{k} \mu(A_i)\log(\mu(A_i)).$$

We further define the entropy of the transformation T with respect to partition $\mathcal{A}$ as

$$h(T, \mathcal{A}) = \lim_{n\to\infty} \frac{1}{n} H(\bigvee_{i=0}^{n-1} T^{-i}(\mathcal{A}))$$

where $T^{-i}(\mathcal{A})$ is the partition $\{T^{-i}(A_j), j \in \{1, \cdots, k\}\}$ and the elements of $\bigvee_{i=0}^{n-1} T^{-i}(\mathcal{A})$ are sets of the form $\bigcap_{i=0}^{n-1} T^{-i} A_{j_i}$ for $j_i \in \{1, \cdots, k\}$.

Finally, we define the metric entropy of T, $h_\mu(T)$, to be the supremum over all finite measurable partitions $\mathcal{A}$ of X of the quantity $h(T, \mathcal{A})$.

This definition was extended by Krengel in [Kre67] to spaces (X, T, μ) for which $\mu(X)$ is conservative but need not be finite by defining

$$h_\mu(T) = \sup\{h_{\mu|_{E_\infty}}(T|_{E_\infty}) : E \subset X, 0 < \mu(E) < \infty\}.$$

Here $E_\infty = \{x \in E : T^n(x) \in E \text{ for infinitely many } n \in \mathbb{N}\}$, $\mu|_{E_\infty}(A) := \mu(E_\infty \cap A) = \mu(E \cap A)$ since T, μ is conservative, and $T|_{E_\infty} : E_\infty \to E_\infty$ is the induced transformation

$$T|_{E_\infty}(x) := T^n(x),$$

where $n = n(x) := \min\{m \geq 1 : T^m(x) \in E_\infty\}$.

We call a set $E \subset X$ a sweep out set if almost every point of X enters E infinitely often under the action of T. If T is conservative and ergodic then every set of positive measure is a sweep out set. It was proved by Krengel in [Kre67] that $h_\mu(T) = h_{\mu|_E}(T|_E)$ for any sweep out set E.

2.2.2 Topological Entropy and Pressure

The following sequence of definitions defines topological pressure for a transformation T on a compact metric space (X, d). While the definition uses the metric d, any two metrics d and d' inducing the same topology will give the same value for the pressure of a function, and thus pressure is a topological invariant, see [Wal82] for details.

Definition 2.2.1. *Let T be a topologically mixing transformation of a compact metric space X. We let dynamical balls be defined by*

$$B_n(x, \epsilon) := \{y \in X : d(T^i(x), T^i(y)) < \epsilon \ \forall i \in \{0, \cdots, n-1\}\}.$$

A set S is said to be an (n, ϵ)-spanning set if $\bigcup_{x \in S} B_n(x, \epsilon)$ covers X. For $g \in C(X, \mathbb{R})$, $n \in \mathbb{N}$ and $\epsilon > 0$ we let

$$Q_n(T, g, \epsilon) = \inf \left\{ \sum_{x \in S} \exp(g^n(x)) | \ S \text{ is an } (n, \epsilon) \text{ spanning set for } X \right\}$$

where $g^n(x) := \sum_{k=0}^{n-1} g(T^k(x))$. We let

$$Q(T, g, \epsilon) = \limsup_{n \to \infty} \frac{1}{n} Q_n(T, g, \epsilon).$$

Finally, we define the topological pressure of a function $g \in C(X, \mathbb{R})$ by

$$P_T(g) = \lim_{\epsilon \to 0} Q(T, g, \epsilon).$$

Topological pressure is a natural generalisation of the earlier notion of topological entropy. We define the topological entropy $h(T)$ of a topologically mixing transformation T of a compact metric space X by

$$h(T) := P_T(0).$$

Where there is no confusion about the transformation T we write $P(g)$ instead of $P_T(g)$. $P(g)$ takes values in $(-\infty, \infty]$. The following theorem gives an equivalent formulation of topological pressure, see ([Wal82]).

Theorem 2.2.1 (The Variational Principle). *Let T be a topologically mixing transformation of a compact metric space X and $g \in C(X, \mathbb{R})$. Then*

$$P_T(g) = \sup\{h_\mu(T) + \int g d\mu | \mu \in \mathcal{M}_T\}.$$

In particular, by putting $g = 0$ this gives us that topological entropy is the supremum over all invariant probability measures of the Kolmogorov-Sinai entropy.

In [Bow70], Bowen showed that the topological entropy of an Axiom A diffeomorphism is equal to the growth rate of the number of periodic orbits. This has been extended to deal with topological pressure and shown to be true for various classes of dynamical system, we state it in terms of finite alphabet Markov shifts.

Theorem 2.2.2. *Let (Σ, σ) be a topologically mixing finite alphabet Markov shift and $g \in C(\Sigma, \mathbb{R})$. Then*

$$P_\sigma(g) = \lim_{n\to\infty} \frac{1}{n} \log \left(\sum_{\underline{x}\in\Sigma:\sigma^n(\underline{x})=\underline{x}} \exp(g^n(\underline{x})) \right).$$

The above definitions and theorems relating to pressure have been for topologically mixing transformations of a compact metric space. In general, pressure does not behave so well on non-compact spaces. Generalisations of definition 2.2.1 to non-compact spaces need not satisfy the variational principle. Indeed, an example was given by Salama in [Sal88] to show that topological pressure as defined by definition 2.2.1 is no longer a topological invariant for non-compact spaces, because using definition 2.2.1 it is possible for two different metrics inducing the same topology to give different values of $P_T(g)$. For this reason, generalisations of the notion of topological pressure to non-compact sets or non-invariant subsets of a compact set tend to use the ideas of theorems 2.2.1 and 2.2.2 or ideas from dimension theory to define pressure. Various such definitions have been given by Bowen [Bow73], Pesin and Pitskel [PP84], Sarig [Sar99] and Thompson[Tho11]. In particular, the definition by Sarig of Gurevich pressure for countable Markov shifts will be used throughout the thesis.

Definition 2.2.2. *Given a mixing subshift of finite type Σ with finite or countably infinite alphabet $\mathcal{A}$ and a weakly Hölder continuous function $g : \Sigma \to \mathbb{R}$, the Gurevich pressure of g is defined as follows.*

$$P_\sigma(g) \quad := \quad \lim_{n\to\infty} \frac{1}{n} \log \left(\sum_{\sigma^n(\underline{x})=x} \exp(g^n(\underline{x}))\chi_{[a]}(\underline{x}) \right)$$

where a is allowed to be any element of $\mathcal{A}$. The choice of a does not affect $P_\sigma(g)$.

Two further properties of P_σ were proved in [Sar99].

Theorem 2.2.3. *For (Σ, σ) and g as in Definition 2.2.2,*

$$\begin{aligned} P_\sigma(g) &= \sup\{h_\mu(T) + \int g d\mu | \mu \in \mathcal{M}_\sigma, \int g d\mu > -\infty\} \\ &= \sup\{P_\sigma(g|K) \big| \ K \textit{ is a compact invariant subset of } \Sigma\}. \end{aligned}$$

Here $P_\sigma(g|K)$ means the topological pressure of $g|K$ on the space $K \subset \Sigma$.

Thus, in the case that $\mathcal{A}$ is finite (and hence Σ is compact), P_σ coincides with the classical definition of pressure. The restriction of the variational principle to measures for which $\int g d\mu > -\infty$ is to avoid the situation that $h_\mu = \infty$ and $\int g d\mu = -\infty$, in which case the sum $h_\mu + \int g d\mu$ is not defined.

2.2.3 Gibbs Measures and Equilibrium States

Gibbs measures, which are defined for topological Markov shifts as follows, are an important class of invariant measure. There are also non-invariant notions of Gibbs measure, but for our purposes Gibbs measures are defined to be invariant.

Definition 2.2.3. *We call an invariant measure μ supported on shift space Σ a Gibbs measure if there exists a function $f : \Sigma \to \mathbb{R}$ and constants $C_1, C_2 > 0$ such that*

$$C_1 \leq \frac{\mu[x_0 \cdots x_{n-1}]}{\exp(f^n(\underline{x}) - nP_\sigma(f))} \leq C_2 \tag{2.1}$$

for all $\underline{x} \in \Sigma$, where $f^n(\underline{x}) := \sum_{k=0}^{n-1} f(\sigma^k(\underline{x}))$ and $[x_0 \cdots x_{n-1}] = \{\underline{y} \in \Sigma : y_0 \cdots y_{n-1} = x_0 \cdots x_{n-1}\}$.

We call μ the Gibbs measure associated to potential f. It is a consequence of the above definition that any potential f associated to a Gibbs measure μ must be continuous. If $\mathcal{A}$ is finite and (Σ, σ) is topologically mixing then there exists a unique Gibbs measure associated to each Hölder continuous function $f : \Sigma \to \mathbb{R}$. Furthermore, each such Gibbs measure is also an equilibrium state for f, defined as follows:

Definition 2.2.4. *Given a Markov shift (Σ, σ) and a function $f : \Sigma \to \mathbb{R}$, we call a measure $\mu \in \mathcal{M}_\sigma$ an equilibrium state if*

$$h_\mu + \int_\Sigma f d\mu = \sup\{h_\nu(T) + \int_\Sigma f d\nu : \nu \in \mathcal{M}_\sigma, \int_\Sigma f d\nu > -\infty\}.$$

If $\mathcal{A}$ is infinite then Gibbs measures and equilibrium states may no longer exist, and it is possible that Gibbs measures exist but equilibrium states do not or vice versa. The following condition on a Markov shift (Σ, σ) is important for the existence of Gibbs measures.

Definition 2.2.5. *We say that a Markov shift (Σ, σ) over countable alphabet $\mathcal{A}$ satisfies the big images and preimages property (BIP) if there exists a finite set $\mathcal{I} \subset \mathcal{A}$ such that for any pair $a, b \in \mathcal{A}$ there exists some $i \in \mathcal{I}$ such that aib is an admissible word.*

In [MU01], Mauldin and Urbański proved that if (Σ, σ) is a topologically mixing BIP shift, $f : \Sigma \to \mathbb{R}$ has summable variation and

$$\sum_{i \in \mathcal{A}} \exp(\sup f|_{[i]}) < \infty,$$

then $P_\sigma(f) < \infty$ and there exists a Gibbs measure μ_f associated to f. There is no requirement that f should be bounded. It was further shown in [MU01] that if we also have that

$$\sum_{i \in \mathcal{A}} \sup(f|_{[i]}) \exp(\sup f|_{[i]}) < \infty,$$

then μ_f is also an equilibrium state. A good reference for Gibbs measures on countable Markov shifts, including many different conditions sufficient for their existence, is given by [MU03].

Sarig showed in [Sar03] that BIP is necessary and sufficient for the existence of a Gibbs measure μ_f associated to f in the case that (Σ, σ) is topologically mixing and f is bounded with summable variation and finite topological pressure.

Finally, Buzzi and Sarig proved in [BS03] that if (Σ, σ) is topologically transitive and f has summable variation then any equilibrium state associated to f is unique.

2.2.4 Coboundaries

Coboundaries are a useful class of function which allow us to manipulate potentials without affecting the thermodynamic quantities associated to them.

Definition 2.2.6. *We say two functions $f, g : \Sigma \to \mathbb{R}$ are cohomologous if there exists a function $\psi : \Sigma \to \mathbb{R}$ such that $f = g + \psi - \psi \circ \sigma$. A function which is cohomologous to zero is called a coboundary.*

If f and g are cohomologous then they have the same topological pressure, and Gibbs measures or equilibrium states associated to f coincide with those associated to g. Furthermore, for any invariant measure μ we have $\int f d\mu = \int g d\mu$. The following theorem was proved by Sinai in [Sin72]. A more modern exposition can be found in the lecture notes of Sarig, [Sar].

Theorem 2.2.4. *Let (Σ, σ) be a two sided countable Markov shift and $f : \Sigma \to \mathbb{R}$ be weakly Hölder continuous. Then there exists a weakly Hölder continuous function $h : \Sigma \to \mathbb{R}$ and a weakly Hölder continuous $g : \Sigma \to \mathbb{R}$ depending only on positive coordinates such that $g = f + h - h \circ \sigma$.*

Since g depends only on positive coordinates we can consider the one sided shift (Σ', σ') corresponding to (Σ, σ), and the natural relations between thermodynamic quantities related to g on Σ' and those related to g on Σ allow us to transfer many results between the one sided and two sided settings.

The following theorem and corollary will be required in chapter 5. Theorems of this type under various different conditions were proved in [JMU06], the statement we use here is a combination of lemma 4.2, lemma 4.4 and corollary 6.5 of that paper.

Theorem 2.2.5. *Let (Σ, σ) be a topologically mixing Markov shift and let $f : \Sigma \to \mathbb{R}$ have summable variation and an invariant Gibbs measure. Then there exists a* *function* ϕ_1^* *with* $var_0(\phi_1^*) < \infty$ *and* $var_i(\phi_1^*) \leq \sum_{j=i+1}^{\infty} var_j(f)$ *for* $i \geq 1$ *such that*

$$g := f + \phi_1^* - \phi_1^* \circ \sigma$$

has $g(\underline{x}) \leq \sup_{\mu \in \mathcal{M}_\sigma} \int_\Sigma g d\mu$.

This was used in [JMU06] to show that if f has summable variation then g must also have summable variation. However, it was pointed out to us by Oliver Jenkinson that if $f(\underline{x}) = f(x_0 x_1)$ then $var_i(\phi_1^*) = 0$ for $i \geq 1$ and $var_0(\phi_1^*) < \infty$, which gives the following corollary.

Corollary 2.2.1. *Let* (Σ, σ) *be a topologically mixing Markov shift and let* $f : \Sigma \to \mathbb{R}$ *have* $f(\underline{x}) = f(x_0 x_1)$ *and* $var_1(f) < \infty$. *Suppose that there exists an invariant Gibbs measure associated to* f. *Then there exists a function* $g : \Sigma \to \mathbb{R}$ *cohomologous to* f *with* $g(\underline{x}) = g(x_0 x_1)$ *and* $g(\underline{x}) \leq \sup_{\mu \in \mathcal{M}_\sigma} \int_\Sigma g d\mu$.

Chapter 3

Counting β-Expansions

3.1 Introduction

There are many ways of representing real numbers. For example, one can consider the decimal expansion $x = \sum_{i=1}^{\infty} x_i.10^{-i}$ of an element of $(0,1)$. A point can have at most two decimal expansions and almost every point with respect to Lebesgue measure has a unique decimal expansion. Alternatively, one can consider expansions in other bases. For $\beta > 1$, we consider expansions

$$x = \sum_{n=1}^{\infty} x_i \beta^{-i}$$

of real numbers $x \in [0, \frac{\lfloor \beta \rfloor}{\beta - 1}]$, where each $x_i \in \{0, 1, \cdots, \lfloor \beta \rfloor\}$. Any such code $(x_i)_{i=1}^{\infty}$ for x is called a β-expansion of x.

In [Sid03a] and [Sid03b], Sidorov proved that, for non-integer $\beta > 1$, almost every $x \in \left[0, \frac{\lfloor \beta \rfloor}{\beta - 1}\right]$ has uncountably many β-expansions, and that the set of exceptions to this rule has Hausdorff dimension strictly less than one. This result was extended to give quantitative information in [SF], where Feng and Sidorov defined

$$\mathcal{N}_n(x;\beta) := \left| \{(a_1, \cdots, a_n) \in \{0, 1, \cdots, \lfloor\beta\rfloor\}^n : \exists (a_{n+1}, a_{n+2}, \cdots) \text{ with } x = \sum_{k=1}^{\infty} a_k \beta^{-k}\} \right|,$$

and proved that for each $\beta \in \left(1, \frac{1+\sqrt{5}}{2}\right)$ there exists a constant $c(\beta) > 0$ such that for all $x \in \left(0, \frac{\lfloor\beta\rfloor}{\beta-1}\right)$,

$$\liminf_{n\to\infty} \frac{\log(\mathcal{N}_n(x;\beta))}{n} \geq c(\beta).$$

We extend the result in [SF] to a wider class of β, giving a positive answer to a question posed in that paper.

Theorem 3.1.1. *For every non-integer real number $\beta > 1$ there exists a constant $c(\beta) > 0$ such that for almost all $x \in \left(0, \frac{\lfloor\beta\rfloor}{\beta-1}\right)$ with respect to Lebesgue measure,*

$$\liminf_{n\to\infty} \frac{\log(\mathcal{N}_n(x;\beta))}{n} \geq c(\beta).$$

We give such a constant $c(\beta)$ explicitly in terms of the absolutely continuous invariant measure of a transformation that generates β-expansions.

For certain $\beta > \frac{1+\sqrt{5}}{2}$, including all $\beta \in \left(\frac{1+\sqrt{5}}{2}, 2\right)$, there exist points $x \in \left(0, \frac{\lfloor\beta\rfloor}{\beta-1}\right)$ which have a unique β-expansion. Therefore there exist real numbers $\beta > 1$ for which the above almost everywhere result does not extend to every $x \in \left(0, \frac{\lfloor\beta\rfloor}{\beta-1}\right)$.

The structure of this chapter goes as follows. In Section 3.2 we recall the construction by Dajani and Kraaikamp of the random β-transformation and explain how, given β, it can be used to generate all β-expansions of a point x. In Section 3.3 we write $\mathcal{N}_n(x;\beta)$ as an expression involving the random β-transformation and apply the ergodic theorem and some simple analysis to complete the proof of theorem 3.1.1. Finally in section 3.4 we discuss possible extensions of the theorem and the limitations of our method.

3.2 Generating β-expansions

We indicate a method of finding β-expansions. For simplicity we let $\beta \in (1, 2)$. It is well known that there exists a β-expansion of $x \in \mathbb{R}$ if and only if $x \in [0, \frac{1}{\beta-1}]$, see for example [DK02].

Question: When does there exist a β-expansion of $x \in \mathbb{R}$ starting with the digit 0?

A number x has a β-expansion starting with the digit 0 if and only if there exists a sequence $(x_n)_{n=1}^{\infty} \in \{0, 1\}^{\mathbb{N}}$ such that

$$x = 0 + \sum_{n=2}^{\infty} \frac{x_n}{\beta^n},$$

giving equivalently that

$$\beta x = \sum_{n=2}^{\infty} \frac{x_n}{\beta^{n-1}} = \sum_{n=1}^{\infty} \frac{x_{n+1}}{\beta^n}.$$

So $(0, x_2, x_3, \cdots)$ is a β-expansion for x if and only if $(x_2, x_3, \cdots)$ is a β-expansion for βx. There will be such a choice $(x_2, x_3, \cdots)$ if and only if $\beta x \in [0, \frac{1}{\beta-1}]$. Therefore x has a β-expansion starting with the digit 0 if and only if $x \in [0, \frac{1}{\beta(\beta-1)}]$.

Question: When does there exist a β-expansion of $x \in \mathbb{R}$ starting with the digit 1?

A number x has a β-expansion starting with the digit 1 if and only if there exists a sequence $(x_n) \in \{0,1\}^{\mathbb{N}}$ such that

$$x = \frac{1}{\beta} + \sum_{n=2}^{\infty} \frac{x_n}{\beta^n},$$

giving equivalently that

$$\beta x - 1 = \sum_{n=2}^{\infty} \frac{x_n}{\beta^{n-1}} = \sum_{n=1}^{\infty} \frac{x_{n+1}}{\beta^n}.$$

So $(1, x_2, x_3, \cdots)$ is a β-expansion for x if and only if $(x_2, x_3, \cdots)$ is a β-expansion for $\beta x - 1$. There will be such a choice if $\beta x - 1 \in [0, \frac{1}{\beta-1}]$. Therefore x has a beta expansion starting with the digit 1 if and only if $x \in [\frac{1}{\beta}, \frac{1}{\beta-1}]$.

Iterating to generate β-expansions:

Using the answers to the two questions above, we are able to generate the first digit of a β-expansion of $x \in [0, \frac{1}{\beta-1}]$, although we note that if $x \in S := [0, \frac{1}{\beta(\beta-1)}] \cap [\frac{1}{\beta}, \frac{1}{\beta-1}]$ then we are allowed a choice for the first digit. Furthermore, we see that if $x_1 = 0$ then x_2 must correspond to the first digit of a β expansion of βx, and so we repeat the above procedure for βx and this gives us a choice of x_2. Similarly if $x_1 = 1$ then x_2 will be the first digit of an expansion of $\beta x - 1$. Iterating this process n times we generate words $(x_1, \cdots, x_n)$ that can be extended to β-expansions of x.

3.2.1 A More General Method Including $\beta > 2$.

In [DK03], Dajani and Kraaikamp defined the random β-transformation K_β which generalises the above idea to include $\beta \geq 2$ and allows one to study all β expansions of x.

Let $\beta > 1$. For $n \in \{0, 1, \cdots, \lfloor \beta \rfloor\}$ we let $T_n(x) := \beta x - n$. We define the regions

S_n by

$$S_n := \left[\frac{n}{\beta}, \frac{\lfloor\beta\rfloor}{\beta(\beta-1)} + \frac{n-1}{\beta}\right], n \in \{1, 2, \cdots, \lfloor\beta\rfloor\},$$

and the switch region S by $S := \bigcup_{n=1}^{\lfloor\beta\rfloor} S_n \times \{0,1\}^{\mathbb{N}}$.

We further define the equality regions E_n by

$$\begin{aligned} E_n &:= \left(\frac{\lfloor\beta\rfloor}{\beta(\beta-1)} + \frac{n-1}{\beta}, \frac{n+1}{\beta}\right), n \in \{1, 2, \cdots, \lfloor\beta\rfloor - 1\}, \\ E_0 &:= \left[0, \frac{1}{\beta}\right) \text{ and } E_{\lfloor\beta\rfloor} := \left(\frac{\lfloor\beta\rfloor}{\beta(\beta-1)} + \frac{\lfloor\beta\rfloor - 1}{\beta}, \frac{\lfloor\beta\rfloor}{\beta-1}\right]. \end{aligned}$$

Then the collection of sets

$$\{E_n : n \in \{0, 1, \cdots, \lfloor\beta\rfloor\}\} \cup \{S_n : n \in \{1, 2, \cdots, \lfloor\beta\rfloor\}\}$$

partitions the interval $\left[0, \frac{\lfloor\beta\rfloor}{\beta-1}\right]$.

We define the random β-transformation $K_\beta : \{0,1\}^{\mathbb{N}} \times \left[0, \frac{\lfloor\beta\rfloor}{\beta-1}\right] \to \{0,1\}^{\mathbb{N}} \times \left[0, \frac{\lfloor\beta\rfloor}{\beta-1}\right]$ by

$$K_\beta(\omega, x) := \begin{cases} (\omega, T_n(x)) & x \in E_n \\ (\sigma(\omega), T_{n-1}(x)) & x \in S_n, w_0 = 0 \\ (\sigma(\omega), T_n(x)) & x \in S_n, w_0 = 1 \end{cases}.$$

Then given a pair $(\omega, x) \in \{0,1\}^{\mathbb{N}} \times \left[0, \frac{\lfloor\beta\rfloor}{\beta-1}\right]$, the sequence $(i_n) \in \{0, 1, \cdots, \lfloor\beta\rfloor\}^{\mathbb{N}}$ corresponding to the sequence of transformations $T_{i_1}, T_{i_2}, \cdots$ applied to the second coordinate in the iteration of K_β gives a β-expansion of x. Furthermore, any β-expansion of x can be given by such a sequence corresponding to (ω, x) for some ω. A proof of this is given in [DK03].

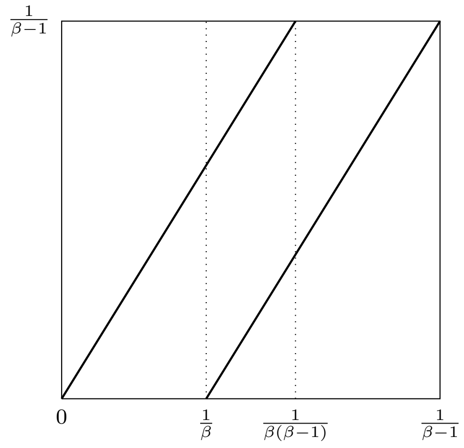

Figure 3.1: The projection onto the second coordinate of K_β for $\beta = \dfrac{1+\sqrt{5}}{2}$

3.3 Proof of Theorem 3.1.1

As seen in the last section, given $x \in \left[0, \frac{\lfloor\beta\rfloor}{\beta-1}\right]$ and $\omega \in \{0,1\}^{\mathbb{N}}$ the random β-transformation generates a unique β-expansion $(x_i)_{i=1}^{\infty}$ of x. We let x be fixed and call $(x_i)_{i=1}^{\infty}$ the β-expansion generated by ω. Similarly, we describe the finite word $(x_1, \cdots, x_n)$ as being the word of length n generated by ω.

In order to count β-expansions using the random β-transformation we want to understand the circumstances under which two sequences $\omega, \omega' \in \{0,1\}^{\mathbb{N}}$ generate different words $(x_1, \cdots, x_n)$.

We let $q = q(\omega, \omega') := \min\{k : \omega_1 \cdots \omega_k \neq \omega'_1 \cdots \omega'_k\}$. Then the first $q-1$ times that the orbits under K_β of (ω, x) and (ω', x) enter the switch region, the same decision is taken about how to continue the β expansion. However on the qth entry to the switch region a different decision is taken. Thus ω and ω' will produce different words of length n if and only if the orbit of (ω, x) enters S at least $q(\omega, \omega')$ times in the first n iterations of K_β. We define

$$h(\omega, x, n) := \#\{k \in \{0, \cdots, n-1\} : K_\beta^k(\omega, x) \in S\}.$$

Then ω, ω' generate the same word of length n if and only if

$$\omega_1, \cdots, \omega_{h(\omega,x,n)} = \omega'_1, \cdots, \omega'_{h(\omega',x,n)}$$

and so defining

$$\Omega(x,n) := \{\omega_1, \cdots, \omega_{h(\omega,x,n)} : \omega \in \{0,1\}^{\mathbb{N}}\},$$

we have $\mathcal{N}_n(x;\beta) = |\Omega(x,n)|$.

Defining m_p to be the $(p, 1-p)$ Bernoulli measure on $\{0,1\}^{\mathbb{N}}$, we have the following characterisation of $|\Omega(x,N)|$.

Lemma 3.3.1.

$$|\Omega(x,n)| = \int_{\{0,1\}^{\mathbb{N}}} 2^{h(\omega,x,n)} dm_{\frac{1}{2}}(\omega).$$

Proof. For $\omega \in \{0,1\}^{\mathbb{N}}$ we have $h(\omega,x,n) \in \{0,1,\cdots,n-1\}$. Then

$$\begin{aligned}\int_{\{0,1\}^{\mathbb{N}}} 2^{h(\omega,x,n)} dm_{\frac{1}{2}}(\omega) &= \sum_{k=0}^{n-1} \int_{\omega\in\{0,1\}^{\mathbb{N}}:h(\omega,x,n)=k} 2^k dm_{\frac{1}{2}}(\omega) \\ &= \sum_{k=0}^{n-1} 2^k \times m_{\frac{1}{2}} \left\{\omega \in \{0,1\}^{\mathbb{N}} : h(\omega,x,n) = k\right\}.\end{aligned}$$

But the set of ω for which $h(\omega,x,n) = k$ is a union of cylinders of the form $[\omega_1, \cdots, \omega_k]$, each of which have $m_{\frac{1}{2}}$ measure 2^{-k}. So

$$m_{\frac{1}{2}} \left\{\omega \in \{0,1\}^{\mathbb{N}} : h(\omega,x,n) = k\right\} = 2^{-k} \left|\{(\omega_1, \cdots, \omega_{h(\omega,x,n)}) : h(\omega,x,n) = k\}\right|$$

Then we can rewrite

$$\begin{aligned}\int_{\{0,1\}^{\mathbb{N}}} 2^{h(\omega,x,n)} dm_{\frac{1}{2}}(\omega) &= \sum_{k=0}^{n-1} 2^k \times 2^{-k} \times \left|\{(\omega_1, \cdots, \omega_{h(\omega,x,n)}) : h(\omega, x, n) = k\}\right| \\ &= |\Omega(x, n)|\end{aligned}$$

□

We want to study the growth of $|\Omega(x, n)|$ using the ergodic theorem, and therefore we need an invariant measure for K_β. In [DdV07], Dajani and de Vries studied invariant measures for the random β-transformation and proved the following theorem.

Theorem 3.3.1. *For each* $p \in [0, 1]$ *there exists a* K_β*-invariant probability measure* $\hat{\mu}_\beta$ *on* $\{0, 1\}^{\mathbb{N}} \times \left[0, \frac{\lfloor\beta\rfloor}{\beta-1}\right]$ *of the form* $\hat{\mu}_\beta = m_p \times \mu_\beta$*, where* μ_β *is absolutely continuous with respect to Lebesgue measure* λ*. Furthermore* $\hat{\mu}_\beta$ *is ergodic.*

We fix $p = \frac{1}{2}$. We will use the measure $\hat{\mu}_\beta$ to show that typical pairs (ω, x) enter the switch region S under the action of K_β with a certain limiting frequency. To that end, we note that it was proved in [DdV07] that $\hat{\mu}_\beta(S) > 0$. We have the following lemma.

Lemma 3.3.2. *For* λ*-a.e.* $x \in \left[0, \frac{\lfloor\beta\rfloor}{\beta-1}\right]$ *we have that for* $m_{\frac{1}{2}}$*-a.e.* $\omega \in \{0, 1\}^{\mathbb{N}}$,

$$\lim_{n\to\infty} \frac{h(\omega, x, n)}{n} = \hat{\mu}_\beta(S).$$

Proof. We define the function $f : \left[0, \dfrac{\lfloor\beta\rfloor}{\beta - 1}\right] \times \{0, 1\}^{\mathbb{N}} \to \mathbb{R}$ by

$$f(\omega, x) = \chi_S(\omega, x) = \begin{cases} 0 & (\omega, x) \notin S \\ 1 & (\omega, x) \in S \end{cases},$$

and see that $f^n(\omega, x) = h(\omega, x, n)$. Then the Birkhoff ergodic theorem gives that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(K_\beta^k(\omega, x)) = \lim_{n\to\infty} \frac{h(\omega, x, n)}{n} = \hat{\mu}_\beta(S),$$

for $\hat{\mu}_\beta$-a.e. pair $(x, \omega) \in \{0,1\}^{\mathbb{N}} \times \left[0, \frac{\lfloor\beta\rfloor}{\beta-1}\right]$.

Now since $\hat{\mu}_\beta = \mu_\beta \times m_p$ is a product measure, statements which are true for almost every pair (ω, x) with respect to $\hat{\mu}_\beta$ are also true for almost every x with respect to μ_β and almost every ω with respect to $m_{\frac{1}{2}}$.

We recall that μ_β is absolutely continuous with respect to λ. Then for λ-a.e. x we have that for $m_{\frac{1}{2}}$-a.e. ω,

$$\lim_{n\to\infty} \frac{h(\omega, x, n)}{n} = \hat{\mu}_\beta(S).$$

□

We now complete the proof of theorem 3.1.1. Since almost everywhere convergence implies convergence in probability, we have from the previous lemma that, for λ-a.e. x:

$\forall \epsilon, \delta > 0, \exists N_{\epsilon\delta}$ such that $\forall n > N_{\epsilon\delta}$,

$$m_{\frac{1}{2}}\left(\{\omega \in \{0,1\}^{\mathbb{N}} : \left|\frac{h(\omega, x, n)}{n} - \hat{\mu}_\beta(S)\right| \geq \epsilon\}\right) < \delta.$$

We define the good set

$$\begin{aligned} G(n, x, \epsilon) &= \{\omega \in \{0,1\}^{\mathbb{N}} : \left|\frac{h(\omega, x, n)}{n} - \hat{\mu}_\beta(S)\right| < \epsilon\} \\ &= \{\omega \in \{0,1\}^{\mathbb{N}} : n(\hat{\mu}_\beta(S) - \epsilon) < h(\omega, x, n) < n(\hat{\mu}_\beta(S) + \epsilon)\}. \end{aligned}$$

Now for λ-a.e. x and all $n > N_{\epsilon\delta}$

$$m_{\frac{1}{2}}(G(n,x,\epsilon)) > 1-\delta,$$

and so

$$\begin{aligned}\int_{\{0,1\}^{\mathbb{N}}} 2^{h(\omega,x,n)} dm_{\frac{1}{2}}(\omega) &\geq \int_{G(n,x,\epsilon)} 2^{h(\omega,x,n)} dm_{\frac{1}{2}}(\omega) \\ &\geq (1-\delta)2^{n(\hat{\mu}_\beta(S)-\epsilon)}.\end{aligned}$$

Then for λ-a.e. x and all $n > N_{\epsilon\delta}$

$$\mathcal{N}_n(x;\beta) \geq (1-\delta)(2^{n(\hat{\mu}_\beta(S)-\epsilon)}),$$

and since ϵ and δ were arbitrary, we have that

$$\liminf_{n\to\infty} \frac{\log(\mathcal{N}_n(x;\beta))}{n} \geq \log(2)\hat{\mu}_\beta(S).$$

This completes the proof of theorem 3.1.1.

3.4 Does There Exist a Growth Rate?

The natural question to ask is whether the growth rate $\lim_{n\to\infty} \dfrac{\log(\mathcal{N}_n(x;\beta))}{n}$ exists. This has been done for the following class of β.

Definition 3.4.1. *A Pisot-Vijayaraghavan number, or PV number, is a real algebraic integer greater than one such that all of its Galois conjugates have absolute value less than one.*

Feng and Sidorov showed in [SF] that if β is a PV number then there exists a constant $k(\beta) > 0$ such that

$$\liminf_{n\to\infty} \frac{\log(\mathcal{N}_n(x;\beta))}{n} = k(\beta),$$

for almost every $x \in \left[0, \dfrac{\lfloor\beta\rfloor}{\beta-1}\right]$.

Furthermore, the value of $k(\beta)$ was computed for $\beta = \frac{1+\sqrt{5}}{2}$ in [SF], and shown to be strictly greater than $\log(2)\hat{\mu}_\beta(S)$, showing that our lower bound for the growth rate is not always sharp.

The chief limitation of our technique is that we cannot say much about the behaviour of

$$\int_{G^c_{n,x,\epsilon}} 2^{h(\omega,x,n)} dm_{\frac{1}{2}}(\omega).$$

Even though $m_{\frac{1}{2}}(G^c_{n,x,\epsilon})$ is tending to zero, $2^{h(\omega,x,n)}$ could potentially be growing as fast as 2^n on this set, and so we cannot discount it. New ideas will be required to consider the possible existence of a growth rate.

Chapter 4

Topological Pressure for Suspension Flows over Countable Markov Shifts

4.1 Introduction

Suspension flows over Markov shifts are useful models for a number of interesting dynamical systems. For example, geodesic flows on compact surfaces of constant negative curvature, and more generally Axiom A flows on compact manifolds, can be modelled by suspension flows over finite alphabet Markov shifts. Through the study of the thermodynamic formalism of these suspension flows, which is well understood due to Sinai, Ruelle, Bowen and others, it has been possible to prove many interesting results about the related flows.

A much larger class of flows on non-compact spaces can be modelled by suspension flows over countable (non-compact) Markov shifts, such as the geodesic flow on the modular surface (see [Ser85]) and the Teichmüller flow (see [BG08]). Recently two models for the thermodynamic formalism of such suspension flows have been suggested. In [Sav98], Savchenko gave a definition of topological entropy for suspension flows with roof function locally constant on each cylinder of length one. In [BI06], Barreira and Iommi gave a definition of topological pressure for suspension flows with Hölder continuous roof functions which do not approach zero. They were shown to be equivalent when they are both defined.

In this chapter we demonstrate that the definition of [BI06] can be extended to suspension flows where the roof function approaches zero, and that crucially the variational principle and relation to pressure on compact invariant subsets still hold. This extended definition coincides with the definition of [Sav98] everywhere that theirs is defined. Furthermore, we prove a relation with the growth rate of weighted sums of periodic orbits, allowing an equivalent definition of pressure as a much more natural analogue of Gurevic pressure for a Markov shift.

We stress that there is a large volume of recent work using various different ideas for topological entropy or pressure of countable alphabet suspension flows (see [BI06], [BG08], [GK01], [Ham10], [Iom10] and [Sav98]). This provides our motivation for seeking a fuller understanding of the relationship between the various definitions.

We now state our main result. All of the definitions will be made precise in the next section.

Theorem 4.1.1. *Let (Σ, σ) be a topologically mixing Markov shift with countable alphabet $\mathcal{A}$ and $f : \Sigma \to \mathbb{R}^+$ a roof function with summable variation giving rise to a suspension flow ϕ on space Σ_f. For any function $g : \Sigma_f \to \mathbb{R}$ for which the function $\Delta_g : \Sigma \to \mathbb{R}$ defined by $\Delta_g(x) := \int_0^{f(x)} g(x,k)dk$ has summable variation,* *the following notions of topological pressure are equivalent.*

$$P_\phi(g) := \lim_{t\to\infty} \frac{1}{t} \log \left(\sum_{\phi_s(\underline{x},0)=(\underline{x},0), s\le t} \exp\left(\int_0^s g(\phi_k(\underline{x},0))dk \right) \chi_{[a]}(\underline{x}) \right) \tag{4.1}$$

$$= \sup_{K\in\mathcal{K}_{\Sigma_f}} P_\phi(g|K) \tag{4.2}$$

$$= \inf\{t\in\mathbb{R} : P_\sigma(\Delta_g - tf) \le 0\} = \sup\{t\in\mathbb{R} : P_\sigma(\Delta_g - tf) \ge 0\} \tag{4.3}$$

$$= \sup\{h_\nu(\phi) + \int g d\nu : \nu \in \mathcal{E}_\phi, \int g d\nu > -\infty\} \tag{4.4}$$

where a is any element of $\mathcal{A}$, $\mathcal{E}_\phi$ is the set of ergodic flow invariant probability measures on Σ_f, $\mathcal{K}_{\Sigma_f}$ is the set of compact flow invariant subsets of Σ_f and P_σ is Gurevic pressure on Σ.

$P_\phi(g)$ takes values in $(-\infty,\infty]$. Proofs will be given for two sided Markov shifts, but, as explained in chapter 2, these proofs transfer over to the case of one sided shifts and the results are valid in either setting.

We have stated our regularity condition on g in terms of the summable variation of Δ_g, this is to avoid having to define a metric on Σ_f. Our variational principle is stated in terms of ergodic invariant measures, we comment on this in section 4.6.

In [Sav98], Savchenko proved that (4.2) and (4.4) are equivalent if f is uniformly locally constant and $g = 0$. In [BI06], Barreira and Iommi proved that (4.2),(4.3) and (4.4) are equivalent in the case that f is bounded away from zero and Hölder continuous. The definition (4.1) and the equivalence of the four definitions in the more general setting of roof functions which are allowed to approach zero are new.

In section 4.3 we prove that the definition (4.1) of pressure is well defined. In section 4.4 we show that the definitions (4.1) and (4.2) are equivalent. In section 4.5 we recall lemmas from [BI06] giving that the definitions (4.2) and (4.3) are equivalent and giving an inequality between the quantities defined by (4.3) and (4.4). Finally, in section 4.6 we show the definitions (4.3) and (4.4) are equivalent.

4.2 Preliminaries

4.2.1 Suspension Flows

In this section we define suspension flows and give definitions of metric entropy and topological pressure for flows on compact spaces analogous to those for transformations in chapter 2.

Given a Markov shift Σ and a function $f : \Sigma \to \mathbb{R}^+$ which we call the roof function we define the space

$$\Sigma_f := \{(x,t) : x \in \Sigma, 0 \leq t \leq f(x)\}$$

with the identification $(x, f(x)) = (\sigma(x), 0)$. We further define the suspension flow ϕ on Σ_f by

$$\phi_t(x,s) = (x, s+t)$$

for $0 \leq t \leq f(x) - s$, and extend this to a flow for all time t using the identification $(x, f(x)) = (\sigma(x), 0)$. We let $\mathcal{M}_\phi$ denote the set of ϕ invariant probability measures on Σ_f.

In order to be a well defined flow we require that $\phi_t(x,s)$ is defined for all $t \in \mathbb{R}$, and hence that $\sum_{n=1}^{\infty} f(\sigma^n(x)) = \sum_{n=1}^{\infty} f(\sigma^{-n}(x)) = \infty$ for all $x \in \Sigma$. A discussion of how this affects the class of allowable roof functions is included on page 39.

Thermodynamics for Suspension Flows over finite Markov shifts:

For a flow $\phi : X \to X$ on a conservative measure space (X, μ), we define the entropy $h_\mu(\phi)$ to be the entropy $h_\mu(\phi_1)$ of the time one transformation $\phi_1 : X \to X$.

In the case of suspension flows Σ_f over finite alphabet Markov shifts with Hölder continuous roof functions, topological pressure for a function $g : \Sigma_f \to \mathbb{R}$ can be defined using dynamical balls, as was done in chapter 2 for transformations. However, since this definition does not generalise well to non-compact spaces, we focus instead on the following three formulations of pressure which are equivalent to the classical definition in the compact case. We define

$$P_\phi(g) := \lim_{t\to\infty} \frac{1}{t} \log \left(\sum_{\phi_s(\underline{x},0)=(\underline{x},0), s\leq t} \exp\left(\int_0^s g(\phi_k(\underline{x},0))dk \right) \right).$$

It was proved by Bowen and Ruelle in [BR75] that

$$P_\sigma(\Delta_g - P_\phi(g).f) = 0,$$

where $\Delta_g(\underline{x}) := \int_0^{f(\underline{x})} g(\phi_k(\underline{x}))dk$. This allows the study of properties of the pressure function P_ϕ on suspension flows over finite Markov shifts to be reduced to the study of the pressure function P_σ on the base. In particular, using the variational principle on a finite Markov shift, it was proved in [BR75] that

$$P_\phi(g) = \sup \left\{ h_\mu(\phi) + \int g d\mu : \mu \in \mathcal{M}_\phi \right\}.$$

Invariant Measures for the Base and for the Flow:

We now let Σ be a countable Markov shift. Given a measure μ on Σ for which $\int_\Sigma f d\mu$ is finite, we can lift the measure to Σ_f by defining

$$\mu_f := \mathcal{L}(\mu) := \frac{(\mu \times m)|_{\Sigma_f}}{\int_\Sigma f d\mu}$$

where m is Lebesgue measure. Measures on Σ which are invariant under σ lift to measures which are invariant under ϕ, see [Abr59].

We have

$$h_{\mu_f}(\phi) = \frac{h_\mu(\sigma)}{\int f d\mu}.$$

This was proved in the case that μ is finite by Abramov in [Abr59] in the general case by Savchenko in [Sav98].

In the case that there exist $c_1, c_2 > 0$ with $c_1 < f < c_2$, $\mathcal{L} : \mathcal{M}_\sigma \to \mathcal{M}_\phi$ is a bijection, where we recall that $\mathcal{M}_\sigma$ (resp. $\mathcal{M}_\phi$) were defined as the sets of σ-invariant (resp. ϕ-invariant) probability measures. However, if f is not bounded away from zero then members of $\mathcal{M}_\phi$ may be the lift of σ-finite invariant measures μ with $\mu(\Sigma) = \infty$, an example of this is given at the end of this section.

Thermodynamic formalism for infinite measure spaces is not as well developed as for finite measure spaces. The fact that members of $\mathcal{M}_\phi$ may be the lift of infinite measures lessens our ability to use the thermodynamic formalism on the base to prove results about the thermodynamic formalism for the flow. In particular, this makes our proof of the variational principle significantly more technical.

Topological Entropy:

As discussed in chapter 2, topological entropy for a transformation T can be defined as the supremum over all ergodic invariant probability measures μ of the metric entropy $h_\mu(T)$. Similarly for a flow ϕ topological entropy can be defined as the supremum of $h_\mu(\phi)$. Then putting $g = 0$ into definition (4.2) of pressure gives topological entropy. Thus we have the following corollary to theorem 4.1.1

Corollary 4.2.1. *Let (Σ, σ) be a topologically mixing countable state Markov shift and $f : \Sigma \to \mathbb{R}^+$ a roof function with summable variation giving rise to a suspension flow ϕ on space Σ_f. The following notions of topological entropy of the flow ϕ are equivalent.*

$$\begin{aligned}
h_{top}(\phi) &:= \sup\{h_\nu(\phi) : \nu \in \mathcal{E}_\phi\} \\
&= \lim_{t\to\infty} \frac{1}{t} \log \left(\sum_{\phi_s(\underline{x},0)=(\underline{x},0), s\leq t} \chi_{[a]}(\underline{x}) \right) \\
&= \sup_{K\in\mathcal{K}_{\Sigma_f}} h_{top}(\phi|K) \\
&= \inf\{t \in \mathbb{R} : P_\sigma(-tf) \leq 0\} = \sup\{t \in \mathbb{R} : P_\sigma(-tf) \geq 0\}
\end{aligned}$$

where a is any element of $\mathcal{A}$.

Compact Subsets:

In order to discuss compact subsets of Σ_f, as in the formulation of definition (4.3), we need a topology on Σ_f. When modelling different systems as suspension flows we may wish to consider various metrics on Σ_f which may induce different topologies. For maximum generality we do not specify precisely what metric or topology we give Σ_f, however we shall assume that set of compact subsets of Σ_f includes all restrictions of Σ_f to suspension flows over finite alphabet Markov shifts $\Sigma' \subset \Sigma$. Natural choices of metric on Σ_f, such as the generalisation of the Bowen-Walters distance considered in [BI06], satisfy this property.

We now state formally our hypotheses.

Hypotheses:

The following hypotheses will be used throughout the chapter. We let Σ be a topologically mixing countable Markov shift with shift operator σ. We assume that the roof function $f : \Sigma \to \mathbb{R}^+$ satisfies $\sum_{n=1}^{\infty} f(\sigma^n(\underline{x})) = \sum_{n=1}^{\infty} f(\sigma^{-n}(\underline{x})) = \infty$ for all $\underline{x} \in \Sigma$, and let ϕ be the corresponding flow on Σ_f. We consider topological pressure of functions $g : \Sigma \to \mathbb{R}$. We assume that both f and $\Delta_g(\underline{x}) = \int_0^{f(\underline{x})} g(\underline{x}, k)dk$ have summable variation, recalling that we do not include var_0 in our definition of summable variation and hence do not require f or Δ_g to be bounded.

Furthermore we assume that $f(\underline{x})$ and $\Delta_g(\underline{x})$ each depend only on the non-negative coordinates $x_0 x_1 \cdots$, this is purely to make the folowing analysis more simple. It was explained in chapter 2 that, for f and Δ_g depending on both positive and negative coordinates, we can add coboundaries such that they depend only on the positive coordinates without affecting any of the thermodynamic properties.

We define $\mathcal{K}_\Sigma$ to be the set of compact shift invariant subsets of Σ upon which σ is topologically mixing, and $\mathcal{K}_{\Sigma_f}$ to be the set of compact flow invariant subsets of Σ_f upon which ϕ is topologically mixing.

An Example:

The requirement that $\sum_{n=1}^{\infty} f(\sigma^n(\underline{x})) = \sum_{n=1}^{\infty} f(\sigma^{-n}(\underline{x})) = \infty$ for all $\underline{x} \in \Sigma$ places some restriction on the systems that we can study. For example, suppose that Σ

is a full shift and $f(\underline{x}) = f(x_0)$ is not bounded away from zero. Then there must exist a sequence of symbols $a_n \in \mathcal{A}$ such that $\lim_{n\to\infty} f(a_n) = 0$, and hence a subsequence b_n for which $f(b_n) < 2^{-n}$. But then for $\underline{x} \in \Sigma$ with $x_i = b_i$ we would have $\sum_{n=1}^{\infty} f(\sigma^n(\underline{x})) < \infty$, contradicting our assumptions. Hence if Σ is a full shift and $f(\underline{x}) = f(x_0)$ then f must be bounded away from zero.

One might wonder whether roof functions approaching zero are just a result of badly chosen codings of flows, and that any flow satisfying our hypotheses can be modeled as a suspension flow with roof bounded away from zero. To dispel these concerns, we give an example of a flow (Σ_f, ϕ) satisfying our conditions that has arbitrarily short closed orbits. Such a flow cannot be recoded to have roof function bounded away from zero without losing some of the orbits in the recoding process. This shows that the class of flows that we consider is genuinely wider than the class of flows considered in [BI06]. We also give an example of an invariant probability measure μ_f on Σ_f which is the lift of an infinite invariant measure on Σ.

Example 4.2.1. *We let $\mathcal{A} = \mathbb{N}$ and (Σ, σ) be a Markov shift over $\mathcal{A}$ corresponding to the incidence matrix M given by*

$$M_{ij} = \begin{cases} 1 & \textit{if i=1, j =1 or } i = j \\ 0 & \textit{otherwise} \end{cases},$$

i.e.

$$M = \begin{pmatrix} 1 & 1 & 1 & 1 & \cdots \\ 1 & 1 & 0 & 0 & \\ 1 & 0 & 1 & 0 & \\ 1 & 0 & 0 & 1 & \\ \vdots & & & & \ddots \end{pmatrix}.$$

We define the roof function f by

$$f(\underline{x}) = 2^{-x_0}.$$

We see that any $\underline{x} \in \Sigma$ must either have $x_n = 1$ for infinitely many $n \geq 1$, or have that there exist $j \in \mathcal{A}$ and $N \in \mathbb{N}$ such that $x_n = j$ for all $n > N \in \mathbb{N}$. In either case $\sum_{n=1}^{\infty} f(\sigma^n(\underline{x})) = \infty$, and using the same arguments with σ^{-1} gives $\sum_{n=1}^{\infty} f(\sigma^{-n}(\underline{x})) = \infty$. Hence the suspension flow Σ_f satisfies our hypotheses.

We see that the periodic orbit of (Σ_f, ϕ) corresponding to the fixed point $\underline{x}^j$ of (Σ, σ) with $x_n = j \ \forall\ n \in \mathbb{Z}$ has period 2^{-j}, and thus that there are periodic orbits of (Σ_f, ϕ) of arbitrarily small period.

Now we define δ_j to be the Dirac measure of mass 1 on the fixed point $\underline{x}^j$ of (Σ, σ). We define $\mu = \sum_{j=1}^{\infty} \delta_j$, and see that μ is invariant and that $\mu(\Sigma) = \sum_{j=1}^{\infty} \delta_j(\Sigma) = \infty$. However, the invariant measure μ_f on Σ_f defined by $\mu_f := \mathcal{L}(\mu)$ has total mass

$$\mu_f(\Sigma_f) = \sum_{j=1}^{\infty} \mathcal{L}(\delta_j)(\Sigma_f) = \sum_{j=1}^{\infty} 2^{-j} = 1.$$

Hence we have shown that there exists an invariant probability measure on Σ_f which is the lift of an infinite invariant measure on Σ.

4.3 An Analogue of Gurevich Pressure for Suspension Flows

We define

$$P_\phi(g) := \lim_{t\to\infty} \frac{1}{t} \log \left(\sum_{\phi_s(\underline{x},0)=(\underline{x},0), s\le t} \exp\left(\int_0^s g(\phi_k(\underline{x},0))dk \right) \chi_{[a]}(\underline{x}) \right)$$

for $a \in \mathcal{A}$, where the summation is over all $\underline{x} \in \Sigma$ such that the orbit of ϕ based at $(\underline{x}, 0)$ is periodic with period s for some $s \le t$. In order to simplify the following arguments, we allow the multiple counting of points, so if the periodic orbit based at $(\underline{x}, 0)$ has prime period $s \le \frac{t}{2}$ we include both the orbit of length s and the orbit of length $2s$ based at $(x, 0)$ in the above summation. If we were to restrict to orbits of prime period s this would have no affect on the quantity P_ϕ.

This definition of pressure is a weighted growth rate as t tends to infinity of the number of periodic orbits of length less than t passing through $[a]$, and is a natural analogue of the Gurevich pressure of [Sar99]. As with Gurevich pressure, we restrict ourselves to counting orbits which pass through some symbol $a \in \mathcal{A}$. Pressure is a measure of the complexity of a transformation, and the growth rate of the number of periodic orbits provides an effective notion of complexity. The actual number of periodic orbits however is unimportant since, for example, the identity transformation on Σ has infinitely many periodic points of any period but would not be regarded as having high complexity.

In this section we show that the limit in the above definition exists for any choice of $a \in \mathcal{A}$, and further that it is independent of any such choice, making P_ϕ well defined. We begin with a slight variant on the 'almost subadditivity' lemma of [Sar99], which is itself a variant on a classical lemma.

Lemma 4.3.1. *If* (a_t) *is a sequence for which there exist constants* c_1, c_2 *such that*

$$a_{s+t+c_2} + c_1 \geq a_s + a_t \tag{4.5}$$

for all $s, t \in \mathbb{R}^+$*, then* $\lim_{t\to\infty} \frac{a_t}{t}$ *exists, taking a value in* $(-\infty, \infty]$*.*

Furthermore, for any $\epsilon, \delta > 0$ *there exists* $T > 0$ *depending only on* c_1, c_2, ϵ *and* δ *such that for all* $t > T$*,*

$$\frac{a_t}{t} \leq \frac{1}{1-\delta} \lim_{t\to\infty} \frac{a_t}{t} + \epsilon.$$

We stress that T is chosen in such a way as to be independent of the sequence a_t, depending only on the constants in equation (4.5). This allows us to prove that certain quantities converge uniformly over subsets of Σ_f, which in turn allows us to prove later that $P_\phi(g)$ can be approximated by the pressure of g on compact invariant subsets of Σ_f.

Proof. Let $\epsilon, \delta > 0$. We will prove that

$$\underline{\lim}\frac{a_t}{t} > (1-\delta)\overline{\lim}\frac{a_t}{t} - \epsilon,$$

and, since ϵ and δ are arbitrary, this will prove the lemma.

We can choose a real number T large enough such that $\frac{T}{T+c_2} > 1 - \delta$ and $\frac{c_1}{T+c_2} < \epsilon$. We fix $p > T$. Then for any $t > c_2$ we can write $t = k(p+c_2) + i + c_2$ where $k \in \mathbb{Z}_+$ and $i \in [0, p+c_2]$. Then we can rewrite

$$\begin{aligned} a_t &= a_{k(p+c_2)+i+c_2} \\ &\geq a_{k(p+c_2)} + a_i - c_1, \end{aligned}$$

using $a_{s+t+c_2} + c_1 \geq a_s + a_t$. Repeated application of this allows us to rewrite

$$a_{k(p+c_2)} \geq ka_p - kc_1.$$

Then

$$\frac{a_t}{t} \geq \frac{ka_p + a_i - (k+1)c_1}{k(p+c_2)+i+c_2}.$$

Keeping p fixed we let t (and hence k) tend to infinity. Since i remains bounded above by $p + c_2$, a_i is also bounded above. We see that

$$\begin{aligned}
\underline{\lim}_{t\to\infty}\frac{a_t}{t} &\geq \underline{\lim}_{k\to\infty}\frac{ka_p}{k(p+c_2)+i+c_2} + \underline{\lim}_{k\to\infty}\frac{a_i}{k(p+c_2)+i+c_2} \\
&- \overline{\lim_{k\to\infty}}\frac{(k+1)c_1}{k(p+c_2)+i+c_2} \\
&= \frac{a_p}{p+c_2} + 0 - \frac{c_1}{p+c_2}.
\end{aligned}$$

This gives that $\underline{\lim}\frac{a_t}{t} > -\infty$. Rearranging we see that

$$\begin{aligned}
\underline{\lim}_{t\to\infty}\frac{a_t}{t} &\geq \frac{p}{p+c_2}\frac{a_p}{p} - \frac{c_1}{p+c_2} \\
&= (1-\delta)\frac{a_p}{p} - \epsilon
\end{aligned}$$

for all $p > T$, proving the second half of the lemma. We can let p tend to infinity and we see that

$$\underline{\lim}_{t\to\infty}\frac{a_t}{t} \geq (1-\delta)\overline{\lim}_{p\to\infty}\frac{a_p}{p} - \epsilon.$$

Then since ϵ and δ were arbitrary we have $\underline{\lim}\frac{a_t}{t} \geq \overline{\lim}\frac{a_t}{t}$, and hence $\lim_{t\to\infty}\frac{a_t}{t}$ is well defined. □

We are now able to show that, for any choice of $a \in \mathcal{A}$, the limit in the definition of $P_\phi(g)$ exists.

Lemma 4.3.2. *Given $a \in \mathcal{A}$, the limit*

$$P_{\phi,a}(g) := \lim_{t\to\infty} \frac{1}{t} \log \left(\sum_{\phi_s(\underline{x},0)=(\underline{x},0), s\le t} \exp\left(\int_0^s g(\phi_k(\underline{x},0))dk \right) \chi_{[a]}(\underline{x}) \right)$$

exists, taking a value in $(-\infty, \infty]$.

Proof. We consider the following sequence.

$$a_t := \log \left(\sum_{\phi_s(\underline{x},0)=(\underline{x},0), s\le t} \exp\left(\int_0^s g(\phi_k(\underline{x},0))dk \right) \chi_{[a]}(\underline{x}) \right).$$

To specify a periodic orbit for the flow it is enough to specify a point $\underline{x}$ on the base through which it passes. Given a word $x_1 \cdots x_n \in \Sigma$ such that $x_n x_1 \in \Sigma$, we let $\overline{(x_1 \cdots x_n)}$ denote the sequence $(y_i)_{i=-\infty}^{\infty} \in \Sigma$ where $y_i = x_{i(\text{mod } n)}$.

Now suppose that we have a periodic orbit $\gamma_1 = \overline{(x_1 \cdots x_n)}$ of period $t = f^n(\overline{(x_1 \cdots x_n)})$, and a periodic orbit $\gamma_2 = \overline{(y_1 \cdots y_m)}$ of period $s = f^m(\overline{(y_1 \cdots y_m)})$, with $x_1 = y_1 = a$. Then the periodic orbit $\gamma_1\gamma_2 := \overline{(x_1 \cdots x_n y_1 \cdots y_m)}$ has period

$$\begin{aligned} f^{n+m}\overline{(x_1 \cdots x_n y_1 \cdots y_m)} &= f^n\overline{(x_1 \cdots x_n y_1 \cdots y_m)} + f^m\overline{(y_1 \cdots y_m x_1 \cdots x_n)} \\ &\le f^n\overline{(x_1 \cdots x_n)} + \sum_{k=1}^{n} var_k(f) \\ &+ f^m\overline{(y_1 \cdots y_m)} + \sum_{k=1}^{m} var_k(f) \\ &\le s + t + 2\sum_{k=1}^{\infty} var_k(f) \end{aligned}$$

We define $c_2 := 2\sum_{k=1}^{\infty} var_k(f) < \infty$, and observe that it is independent of the lengths n and m. Thus any two periodic orbits γ_1 and γ_2 sharing a common base point can be interwoven to give the periodic orbit $\gamma_1\gamma_2$ of period less than or equal to $l(\gamma_1) + l(\gamma_2) + c_2$.

For a periodic orbit γ of period t passing through point $(\overline{x_1 \cdots x_n}, 0)$ we write

$$\int_\gamma g := \int_0^t g(\phi_k(\overline{x_1 \cdots x_n}, 0))dk = \sum_{k=1}^{n} \Delta_g(\sigma^k \overline{(x_1 \cdots x_n)}).$$

Now Δ_g has summable variation, and so letting $c_1 := 2\sum_{n=1}^{\infty} var_k(\Delta_g) < \infty$ and using the same arguments given above for f, we have

$$\int_{\gamma_1\gamma_2} g + c_1 \geq \int_{\gamma_1} g + \int_{\gamma_2} g.$$

So for any γ_1, γ_2 in the summations for a_s and a_t, their concatenation $\gamma_1\gamma_2$ is in the summation for a_{s+t+c_2}, and the evaluation of g over this orbit differs by at most c_1. We may have extra orbits in the summation for a_{s+t+c_2} but these cannot contribute negatively. Thus we get the inequality,

$$a_{s+t+c_2} + c_1 \geq a_s + a_t.$$

Then lemma 4.3.1 gives that $\lim_{t\to\infty} \frac{a_t}{t}$ exists, and since $\lim_{t\to\infty} \frac{a_t}{t}$ is $P_{\phi,a}(g)$, we have $P_{\phi,a}(g)$ is well defined. □

We now show that $P_{\phi,a}(g)$ does not depend on the choice of $a \in \mathcal{A}$.

Lemma 4.3.3. *$P_{\phi,a}(g)$ is independent of a, and hence $P_\phi(g)$ is well defined.*

Proof. Let $a, b \in \mathcal{A}$. We let a_t be defined as in the previous lemma and let b_t be defined analogously:

$$b_t := \log\left(\sum_{\phi_s(\underline{x},0)=(\underline{x},0), s\leq t} \exp\left(\int_0^s g(\phi_k(\underline{x}, 0))dk\right) \chi_{[b]}(\underline{x})\right).$$

We choose and fix finite words $x_1 \cdots x_m$ and $y_1 \cdots y_n$, where $x_1 = a$ and $x_m b$ is an admissible word, and where $y_1 = b$ and $y_n a$ is an admissible word. These should be thought of as paths in Σ linking a to b and b to a respectively. We define

$$\begin{aligned} T_{a,b} &= \sup\{f^m(x) : x \in [x_1 \cdots x_m b]\} + \sup\{f^n(y) : y \in [y_1 \cdots y_n a]\} \\ G_{a,b} &= \inf\{\Delta_g^m(x) : x \in [x_1 \cdots x_m b]\} + \inf\{\Delta_g^n(y) : y \in [y_1 \cdots y_n a]\}. \end{aligned}$$

These are finite since f and Δ_g have summable variation, even though f and Δ_g may be unbounded.

Then any periodic orbit γ_1 of length t based at $(\overline{z_1 \cdots z_p}, 0)$ with $z_1 = a$ can be extended to a periodic orbit γ_2 based at $(\overline{y_1 \cdots y_n z_1 \cdots z_p x_1 \cdots x_m}, 0)$. This orbit passes through $([b], 0)$ and so is included in the summation for b_t.

We see that

$$\begin{aligned} l(\gamma_2) &= f^{n+p+m}(\overline{y_1 \cdots y_n z_1 \cdots z_p x_1 \cdots x_m}) \\ &\leq \sup\{f^n(y) : y \in [y_1 \cdots y_n a]\} + f^p(\overline{z_1 \cdots z_p}) + \sum_{n=1}^{\infty} var_n(f) \\ &+ \sup\{f^m(x) : x \in [x_1 \cdots x_m b]\} \\ &\leq t + c_2 + T_{a,b}. \end{aligned}$$

Similarly

$$\int_{\gamma_1} g - c_1 + G_{a,b} \leq \int_{\gamma_2} g.$$

So we see that

$$\log\left(\sum_{\phi_s(\underline{x},0)=(\underline{x},0),s\leq t+T_{a,b}+c_2}\exp\left(\int_0^s g(\phi_k(\underline{x},0))dk\right)\chi_{[b]}(\underline{x})\right)\geq$$
$$\log\left(\sum_{\phi_s(\underline{x},0)=(\underline{x},0),s\leq t}\exp\left(G_{a,b}-c_1+\int_0^s g(\phi_k(\underline{x},0))dk\right)\chi_{[a]}(\underline{x})\right),$$

i.e. that

$$b_{t+T_{a,b}+c_2}-G_{a,b}+c_1\geq a_t.$$

Dividing by t, taking the limit as t tends to infinity we see that

$$\lim_{t\to\infty}\frac{b_t}{t}\geq\lim_{t\to\infty}\frac{a_t}{t}.$$

But since $a,b\in\mathcal{A}$ were arbitrary, this gives us that $\lim_{t\to\infty}\frac{b_t}{t}=\lim_{t\to\infty}\frac{a_t}{t}$, and that our definition of pressure is independent of a. □

4.4 Compact Invariant Subsets

We want to prove that our definition of the topological pressure of g on Σ_f is the supremum over all compact invariant subsets $J_f\subset\Sigma_f$ of $P_\phi(g|J_f)$.

We define Σ_f^{fin} to be the set of suspension flows $\Sigma_f'\subset\Sigma_f$ for which Σ' is the restriction of Σ to sequences in $\mathcal{A}'^{\mathbb{Z}}$, for some finite subalphabet $\mathcal{A}'$ of $\mathcal{A}$. We recall that our set of compact invariant subsets of Σ_f includes Σ_f^{fin}. Given two compact invariant subsets $A,B\subset\Sigma_f$ for which $A\subset B$ we have that $P_\phi(g|A)\leq P_\phi(g|B)\leq P_\phi(g)$. Thus in order to prove that $P_\phi(g)$ is the supremum over all compact invariant subsets $K_f \subset \Sigma_f$ of $P_\phi(g|K_f)$, it is enough to prove that

$$P_\phi(g) = \sup_{K_f \in \Sigma_f^{fin}} P_\phi(g|K_f).$$

In [Sar99], Sarig proved that for $\Delta_g : \Sigma \to \mathbb{R}$ we have $P_\sigma(\Delta_g) = \sup_{K \in \mathcal{K}_\Sigma} P_\sigma(\Delta_g|K)$. We adapt the statement and proof to our setting.

Lemma 4.4.1. *$P_\phi(g) = \sup_{K_f \in \mathcal{K}_{\Sigma_f}} P_\phi(g|K_f)$.*

Proof. We have already argued that it is enough to prove this for compact invariant sets $K_f \subset \Sigma_f$ which are suspension flows over finite Markov shifts.

We define

$$a_t(K_f) := \log \left(\sum_{(\underline{x},0) \in K_f : \phi_s(\underline{x},0) = (\underline{x},0), s \leq t} \exp \left(\int_0^s g(\phi_k(\underline{x},0)) dk \right) \chi_{[a]}(\underline{x}) \right).$$

where a is any member of the alphabet upon which K_f is supported. We recall that

$$P_\phi(g|K_f) := \lim_{t \to \infty} \frac{a_t(K_f)}{t}$$

and

$$P_\phi(g) := \lim_{t \to \infty} \frac{a_t}{t}.$$

Then since the summation in the definition of $a_t(K_f)$ is over a smaller set than the corresponding summation in the definition of a_t, we have that $a_t(K_f) \leq a_t$ and hence

$$P_\phi(g) \geq \sup\{P_\phi(g|K_f) : K_f \in \mathcal{K}_{\Sigma_f}\}.$$

We will prove the reverse inequality. Let us assume that $P_\phi(g) < \infty$, the infinite case is similar. Fix $\epsilon, \delta > 0$. We recall that lemma 4.3.2 gives that

$$a_{s+t+c_2}(K_f) + c_1 \geq a_s(K_f) + a_t(K_f)$$

where c_1 and c_2 do not depend on K_f, and hence by lemma 4.3.1 there exists $T > 0$ independent of K_f such that for all $t > T$,

$$(1+\delta)P_\phi(g|K_f) + \epsilon \geq \frac{a_t(K_f)}{t}.$$

We choose and fix $t > T$ large enough so that

$$P_\phi(g) \leq \frac{1}{t}a_t + \epsilon.$$

Now the summation in the definition of a_t is a summation over the countable set of loops in Σ from a to a of length less than or equal to t. But countable summation is just the limit of summations over finite subsets, and any finite set of loops of length less than or equal to t must pass through only finitely many elements of $\mathcal{A}$.

Then we can choose $M \in \mathbb{N}$ big enough so that

$$\frac{1}{t}a_t \leq \frac{1}{t}a_t((\{1, \ldots, M\}^{\mathbb{Z}} \cap \Sigma)_f) + \epsilon,$$

where by $(\{1, \ldots, M\}^{\mathbb{Z}} \cap \Sigma)_f$ we mean the suspension flow over the restriction of Σ to the alphabet $\{1, \cdots, M\}$. By adding a finite number of symbols we can extend $(\{1, \ldots, M\}^{\mathbb{N}} \cap \Sigma)_f$ to a space K_f which intersects $[a] \times \{0\}$, is still compact, and on which the shift transformation on the base is topologically mixing. We still have

$$\frac{1}{t}a_t \leq \frac{1}{t}a_t(K_f) + \epsilon.$$

We have argued that

$$\begin{aligned} P_\phi(g) &\leq \frac{a_t}{t} + \epsilon \\ \frac{a_t}{t} &\leq \frac{a_t(K_f)}{t} + \epsilon \text{ and} \\ \frac{a_t(K_f)}{t} &\leq (1+\delta)P_\phi(g|K_f) + \epsilon. \end{aligned}$$

Then we have

$$P_\phi(g) \leq (1+\delta)P_\phi(g|K_f) + 3\epsilon,$$

and since ϵ and δ were arbitrary this gives that

$$P_\phi(g) \leq \sup\{P_\phi(g|K_f) : K_f \in \mathcal{K}_{\Sigma_f}\}.$$

Combining with the reverse inequality given earlier, we have that $P_\phi(g)$ is indeed the supremum of the topological pressures of suspension flows over compact flow invariant subsets.

□

4.5 The Definition of Barreira and Iommi

In this section we restate a lemma from [BI06] which 4.5.1 proves that the notion of topological pressure used in [BI06], is equal to the supremum over compact invariant subsets of the classical notion of pressure for ϕ restricted to that subset. It has been extended from a lemma in [BI06] to deal with roof functions that approach zero without altering the proof. This gives us that our definition of topological pressure and the definition used in [BI06] are equivalent.

We recall that, in [BI06], Barreira and Iommi defined topological pressure by the equation

$$P_{BI}(g) := \inf\{t \in \mathbb{R} : P_\sigma(\Delta_g - tf) \leq 0\} = \sup\{t \in \mathbb{R} : P_\sigma(\Delta_g - tf) \geq 0\}.$$

We will shortly prove that P_{BI} and P_ϕ are equivalent, after which we will no longer use the notation P_{BI}.

Lemma 4.5.1. *$P_{BI}(g) = \sup_{K \in \mathcal{K}_{\Sigma_f}} P_\phi(g|K)$*

Proof. We have that

$$\begin{aligned} P_{BI}(g) &:= \inf\{t \in \mathbb{R} : P_\sigma(\Delta_g - tf) \leq 0\} \\ &= \inf\{t \in \mathbb{R} : \sup_{K \in \mathcal{K}_\Sigma} \{P_\sigma((\Delta_g - tf)|_K)\} \leq 0\} \\ &= \inf\{t \in \mathbb{R} : P_\sigma((\Delta_g - tf)|_K) \leq 0 \ \forall K \in \mathcal{K}_\Sigma\}. \end{aligned}$$

The second line uses the fact that Gurevich pressure can be approximated by topological pressure on compact invariant subsets. Given $K \in \mathcal{K}_\Sigma$ we denote K_f the element of $\mathcal{K}_{\Sigma_f}$ given by $\{(\underline{x}, y) : \underline{x} \in K, 0 \leq y \leq f(x)\}$. This gives a one to one correspondence between members of $\mathcal{K}_\Sigma$ and $\mathcal{K}_{\Sigma_f}$. But

$$\begin{aligned} P_\sigma((\Delta_g - tf)|_K) \leq 0 &\implies t \geq \text{the unique } t_0 \in \mathbb{R} \text{ satisfying } P_\sigma((\Delta_g - t_0 f)|_K) = 0 \\ &\implies t \geq P_\phi(g|K_f). \end{aligned}$$

Combining this with the previous equation gives

$$\begin{aligned} P_{BI}(g) &= \inf\{t \in \mathbb{R} : P_\phi(g|K) \leq t \ \forall K \in \mathcal{K}_{\Sigma_f}\} \\ &= \sup_{K \in \mathcal{K}_{\Sigma_f}} P_\phi(g|K) \end{aligned}$$

as required. $\square$

Since we also have that $P_\phi(g) = \sup_{K \in \mathcal{K}_{\Sigma_f}} P_\phi(g|K)$, we have the equivalence of P_ϕ and P_{BI}.

4.6 A Variational Principle for P_ϕ

We wish to state a variational principle for our notion of topological pressure. In order to do so, we first define some relevant spaces of measures.

We recall that $\mathcal{M}_\phi$ was defined as the set of all flow invariant probability measures on the space Σ_f. We further define the spaces

$$\begin{aligned} \mathcal{M}_{\phi,g} &:= \left\{ \nu \in \mathcal{M}_\phi : \int g d\nu > -\infty \right\}, \\ \mathcal{M}^p_{\phi,g} &:= \left\{ \nu \in \mathcal{M}_\phi : \int g d\nu > -\infty, \nu = \mathcal{L}(\mu) \text{ for some } \mu \in \mathcal{M}_\sigma \right\}. \end{aligned}$$

We recall that any measure $\nu \in \mathcal{M}_\phi$ is the lift of some σ-invariant measure on Σ, but that this measure may not always be finite. We let $\mathcal{E}_\phi$ denote the restriction of $\mathcal{M}_\phi$ to ergodic measures, and do the same for $\mathcal{E}_{\phi,g}, \mathcal{E}^p_{\phi,g}$ etc.

We begin with lemma 4.6.1, which proves that P_ϕ satisfies a limited variational principle, the statement has been altered from that in [BI06] to avoid the complications with infinite measures that arise in our wider setting of roof functions which are allowed to approach zero but the proof remains essentially the same.

Lemma 4.6.1. $P_\phi(g) = \sup \left\{ h_{\mu_f}(\phi) + \int g d\mu_f : \mu_f \in \mathcal{M}^p_{\phi,g} \right\}$

Proof. We let $K \in \mathcal{K}_\Sigma$ and let K_f be the corresponding element of $\mathcal{K}_{\Sigma_f}$. We have that

$$P_\phi(g|K_f) \;=\; \sup\left\{h_{\mu_f}(\phi) + \int_{\Sigma_f} g d\mu_f : \mu_f \in \mathcal{M}_\phi(K_f)\right\},$$

where $\mathcal{M}_\phi(K_f)$ is the restriction of $\mathcal{M}_\phi$ to measures fully supported on K_f. This is the statement of the variational principle for flows on compact spaces.

Furthermore, $g : K_f \to \mathbb{R}$ must be bounded below, since g is continuous and K_f is compact. Then for $\mu_f \in \mathcal{M}_\phi(K_f)$ we have that $\int_{K_f} g d\mu_f > -\infty$.

For $\mu_f \in \mathcal{M}_\phi(K_f)$ we have $\mu_f = \mathcal{L}(\mu)$ where μ is an invariant measure on Σ with $\int_K f d\mu < \infty$. But $f > 0$ must be bounded away from zero on the compact set K, and if $\int_K f d\mu < \infty$ then we must also have that $\mu(K) < \infty$. So we can restate the variational principle for ϕ on K_f as follows.

$$P_\phi(g|K_f) \;=\; \sup\left\{h_{\mu_f}(\phi) + \int_{\Sigma_f} g d\mu_f : \mu_f \in \mathcal{M}^p_{\phi,g}(K_f)\right\}$$

Now using lemma 4.5.1, we take the supremum over compact subsets and get that

$$\begin{aligned} P_\phi(g) &= \sup_{K_f \in K_{\Sigma_f}} \sup\left\{h_{\mu_f}(\phi) + \int_{\Sigma_f} g d\mu_f : \mu_f \in \mathcal{M}^p_{\phi,g}(K_f)\right\} \\ &\leq \sup\left\{h_{\mu_f}(\phi) + \int_{\Sigma_f} g d\mu_f : \mu_f \in \mathcal{M}^p_{\phi,g}\right\}. \end{aligned}$$

We now prove the reverse inequality. For $t > P_\phi(g) = \inf\{t : P_\sigma(\Delta_g - tf) \leq 0\}$ we have that

$$\begin{aligned}
0 &\geq P_\sigma(\Delta_g - tf) \\
&\geq \sup\left\{h_\mu(\sigma) + \int_\Sigma \Delta_g d\mu - t\int_\Sigma f d\mu : \mu \in \mathcal{M}_\sigma, \int_\Sigma \Delta_g d\mu > -\infty\right\} \\
&= \sup\left\{\int_\Sigma f d\mu \left(\frac{h_\mu(\sigma)}{\int_\Sigma f d\mu} + \frac{\int_\Sigma \Delta_g d\mu}{\int_\Sigma f d\mu} - t\right) : \mu \in \mathcal{M}_\sigma, \int_\Sigma \Delta_g d\mu > -\infty\right\}.
\end{aligned}$$

The second line is the statement of the variational principle for Gurevich pressure, and the third line is just rearrangement, using that $0 < \int_\Sigma f d\mu < \infty$.

Then dividing by $\int_\Sigma f d\mu$ we see that

$$0 \geq \sup\left\{h_{\mu_f}(\phi) + \int g d\mu_f - t : \mu_f \in \mathcal{M}^p_{\phi,g}\right\},$$

giving

$$t \geq \sup\left\{h_{\mu_f}(\phi) + \int g d\mu_f : \mu_f \in \mathcal{M}^p_{\phi,g}\right\}.$$

We have proved that

$$t > P_\phi(g) \implies t \geq \sup\left\{h_{\mu_f}(\phi) + \int g d\mu_f : \mu_f \in \mathcal{M}^p_{\phi,g}\right\},$$

i.e. that

$$P_\phi(g) \geq \sup\left\{h_{\mu_f}(\phi) + \int g d\mu_f : \mu_f \in \mathcal{M}^p_{\phi,g}\right\}.$$

We have now proved the inequality in both directions, and hence have that

$$P_\phi(g) = \sup\left\{h_{\mu_f}(\phi) + \int g d\mu_f : \mu_f \in \mathcal{M}^p_{\phi,g}\right\}.$$

□

Furthermore, we recall that, on compact sets, the variational principle can be phrased in terms of ergodic measures, that is

$$P_\phi(g|K_f) \quad = \quad \sup\left\{h_{\mu_f}(\phi) + \int_{\Sigma_f} gd\mu_f : \mu_f \in \mathcal{E}^p_{\phi,g}(K_f)\right\}.$$

Using this observation one can follow the method of the previous proof exactly to obtain the following corollary.

Corollary 4.6.1.

$$P_\phi(g) = \sup\left\{h_{\mu_f}(\phi) + \int gd\mu_f : \mu_f \in \mathcal{E}^p_{\phi,g}\right\}.$$

We now extend this variational principle to include ergodic measures which are the lift of infinite invariant measures on Σ:

Lemma 4.6.2. $P_\phi(g) = V(g) := \sup\{h_\nu(\phi) + \int gd\nu : \nu \in \mathcal{E}_{\phi,g}\}$.

Using lemma 4.6.1, it remains only to prove that

$$\sup\{h_\nu(\phi) + \int gd\nu : \nu \in \mathcal{E}^p_{\phi,g}\} = \sup\{h_\nu(\phi) + \int gd\nu : \nu \in \mathcal{E}_{\phi,g}\}.$$

Plan of Proof:

While the details are slightly technical, the principle behind the proof here is simple. The proof follows the following three steps.

Step 1: Prove that for $\epsilon > 0$ there exists an ergodic conservative σ-finite measure μ on Σ such that $h_{\mu_f}(\phi) + \int gd\mu_f > V(g) - \epsilon$.

Step 2: Define a sequence of finite measures μ^n on Σ and show that they are well defined.

Step 3: Furthermore, show that μ^n also satisfy

$$\begin{aligned} \int f d\mu^n &\to \int f d\mu > 0, \\ \int \Delta_g d\mu^n &\to \int \Delta_g d\mu, \text{ and} \\ h_{\mu^n}(\sigma) &\to h_\mu(\sigma). \end{aligned}$$

We then have that

$$h_{\mu_f^n}(\phi) + \int g d\mu_f^n = \frac{h_{\mu^n}(\sigma)}{\int f d\mu^n} + \frac{\int \Delta_g d\mu^n}{\int f d\mu^n} \to \frac{h_\mu(\sigma)}{\int f d\mu} + \frac{\int \Delta_g d\mu}{\int f d\mu} = h_{\mu_f}(\phi) + \int g d\mu_f.$$

Thus we have a sequence of measures $\mu_f^n \in \mathcal{M}_{\phi,g}^p$ which come arbitrarily close to achieving the supremum $V(g)$. This proof is an extension of the one given by Savchenko in [Sav98], which dealt with the case that f is locally constant and $g = 0$. We begin by selecting an appropriate sigma finite measure on the base.

Proof. **Step 1:** We identify a suitable measure μ on the base.

Lemma 4.6.3. *For any $\epsilon > 0$ there exists an ergodic conservative σ-finite measure μ on Σ with $\int_\Sigma f d\mu < \infty$, $\int_\Sigma \Delta_g d\mu > -\infty$ and*

$$h_{\mu_f}(\phi) + \int g d\mu_f + \epsilon > V(g),$$

where $\mu_f = \mathcal{L}(\mu)$.

Proof. We let μ_f be an ergodic probability measure on Σ_f with $h_{\mu_f} + \int g d\mu_f + \epsilon > V(g)$, and recall that μ_f is automatically the lift of some σ-finite measure μ on Σ with $\int_\Sigma f d\mu < \infty$ under the map $\mathcal{L}$. We want to show that μ must be ergodic and conservative.

We note that any set $A_f \subset \Sigma_f$ which is invariant under ϕ is necessarily of the form $(A \times [0, \infty))|_{\Sigma_f}$ where A is a subset of Σ which is invariant under σ. An invariant set on the space of the suspension flow is uniquely defined by its intersection with the base.

We suppose for a contradiction that μ is not ergodic, i.e. that there exists some invariant set $A \subset \Sigma$ with $\mu(A) > 0$ and $\mu(A^c) > 0$. But then the corresponding set on the space of the suspension flow, $A_f := (A \times [0, \infty))|_{\Sigma_f}$, is also invariant.

Since $f > 0$ is continuous we can measurably partition A by $A = \cup_{n=0}^{\infty} A^n$ where $A^n := \{x \in A : \frac{1}{n+1} \leq f(x) < \frac{1}{n}\}$. Since $\mu(A) > 0$ it follows that at least one of the sets A^i has positive measure. But since $f \in [\frac{1}{i+1}, \frac{1}{i})$ on A^i it follows further that $\mu_f(A_f^i) > \frac{\mu(A) \times \frac{1}{i+1}}{\int f d\mu} > 0$. Hence $\mu_f(A_f) > 0$, and identical arguments show that $\mu_f(A_f^c) > 0$. But then A_f is an invariant set with $\mu_f(A_f) > 0$ and $\mu_f(A_f^c) > 0$, contradicting the assumption that μ_f is an ergodic measure. Hence μ must be ergodic.

Now we recall that a measure μ is called conservative if for every measurable set A with $\mu(A) > 0$, almost every point of A will return to A. Finite measures are necessarily conservative by the Poincaré recurrence theorem. So since μ_f is finite, it is a conservative invariant measure on Σ_f.

We suppose that μ is not conservative, and let measurable sets $A, B \subset \Sigma$ have that $\mu(A) > 0, \mu(B) > 0, B \subset A$ and that no point of B returns to the set A under the action of σ. Then no point of B_f returns to A_f under the action of ϕ. But, as argued above in the case of ergodicity, $\mu_f(B_f) > 0$, and hence μ_f cannot be conservative. This contradiction proves that μ must be conservative. $\square$

We have shown that there exists a conservative ergodic invariant measure μ on Σ with $h_{\mu_f}(\phi) + \int g d\mu_f + \epsilon > V(g)$, completing step 1 of the proof of lemma 4.6.2. We will now prove that there exists a sequence of finite measures μ^n for which

$$h_{\mu_f^n}(\phi) + \int g d\mu_f^n \to h_{\mu_f}(\phi) + \int g d\mu_f,$$

and then using step 1 this shows that

$$\lim_{n\to\infty} h_{\mu_f^n}(\phi) + \int g d\mu_f^n \geq V(g) - \epsilon.$$

Since ϵ was arbitrary, this will complete the proof of the variational principle.

Step 2: We define a sequence of finite measures μ^n on Σ and show that they are well defined. In step 3 we will use these measures to complete the proof of lemma 4.6.2.

Let $\delta > 0$. We choose $m \in \mathbb{N}$ such that $\sum_{k=m-1}^{\infty} var_k(f) < \delta$.

Given a set $A \subset \Sigma$, we let the set A_∞ be the set of sequences $\underline{x}$ for which $\sigma^n(\underline{x})$ intersects A for infinitely many positive and negative values of n. Since μ is conservative and ergodic we have that, for every set A such that $0 < \mu(A) < \infty$, $\mu(\Sigma \backslash A_\infty) = 0$ and $h_\mu(\sigma) = h_{\mu|A}(\sigma|A)$ (see the earlier section on metric entropy).

We choose A to be a cylinder set $[a_1 \cdots a_m]$ for which $\mu[a_1 \cdots a_m] > 0$ and for which there doesn't exist a $k < m$ such that $a_1 \cdots a_k = a_{m-k+1} \cdots a_m$. This technical restriction just avoids two occurrences of the word $a_1 \cdots a_m$ overlapping, preventing the need for further combinatorial arguments later. Since multiplying μ by a constant has no effect on the lifted measure μ_f, we replace μ with $\frac{1}{\mu[a_1 \cdots a_m]}\mu$. The new measure μ continues to satisfy the requirements of step 1, and we have that $\mu[a_1 \cdots a_m] = 1$.

The set of all finite words in Σ in which $a_1 \cdots a_m$ appears at the start and end but nowhere else is countable. We number the elements arbitrarily $(\gamma_i)_{i=1}^{\infty}$, and say that word γ_i has length $l(\gamma_i)$. The cylinder $[\gamma_i]$ is the set of sequences in Σ whose first $l(\gamma_i)$ coordinates coincide with those of γ_i. The set $\Sigma \cap [a_1 \cdots a_m]_\infty$ is partitioned by $\{\sigma^k[\gamma_i] : i \in \mathbb{N}, 0 \leq k \leq l(\gamma_i) - m\}$. Finally we let $q_n := \sum_{i=1}^n \mu[\gamma_i]$, and observe that q_n increases to $\mu[a_1 \cdots a_m] = 1$ as n tends to infinity.

Lemma 4.6.4. *For each $n \in \mathbb{N}$ there exists a shift invariant measure μ^n on Σ such that*

1. $\mu^n(\Sigma \backslash [a_1 \cdots a_m]_\infty) = 0$
2. $\mu^n[\gamma_i] = \frac{\mu[\gamma_i]}{q_n}$ *for* $i \in \{1, \cdots n\}$
3. $\mu^n[\gamma_i] = 0$ *for* $i > n$
4. *The induced measure* $\mu^n|_{[a_1 \cdots a_m]}$ *is a Bernoulli measure on choices of* $[\gamma_i]$*,*
5. μ^n *is invariant under* σ.
6. $\mu^n(\Sigma) < \infty$

Proof. A point x in A which returns to A infinitely many times uniquely determines, and is uniquely determined by, the sequence of loops in Σ corresponding to successive excursions from A. We have already enumerated these paths $(\gamma_i)_{i=1}^\infty$, and so we can code points $x \in A_\infty$ with the doubly infinite sequence of members of $(\gamma_i)_{i=1}^\infty$ corresponding to excursions from A of x under σ and σ^{-1}. We write this as a sequence in $\mathbb{N}^\mathbb{Z}$. Since (Σ, σ) is a Markov shift, the history of a point x before it entered A places no restriction on its future trajectory, and for each sequence in $\mathbb{N}^\mathbb{Z}$ there exists a corresponding point in A_∞. We let $\Sigma' := \mathbb{N}^\mathbb{Z}$, and see that the shift transformation σ on Σ' models the action of the induced transformation $\sigma|_A$ on A.

For $i \in \{1, \cdots n\}$ we define $\nu[i] = \frac{\mu[\gamma_i]}{q_n}$ and for $i > n$ we define $\nu[i] = 0$. Then

$$\sum_{i=1}^{\infty} \nu[i] = \sum_{i=1}^{n} \frac{\mu[\gamma_i]}{\sum_{i=1}^{n} \mu[\gamma_i]} = 1.$$

Indeed, the reason that we divided by q_n in the definition of ν was that it gave us the above property, which allows us to extend ν to a Bernoulli measure on Σ' by defining

$$\nu[i_m i_{m+1} \cdots i_n] = \Pi_{k=m}^{n} \nu[i_k].$$

This measure extends naturally to an invariant measure μ^n on the subspace A_∞ of Σ by defining $\mu^n[\gamma_{i_1} \cdots \gamma_{i_k}] = \nu[i_1 \cdots i_k]$, and then using additivity to extend this to cylinders $[x_1 \cdots x_n]$ in Σ which are not closed loops based at $[a_1 \cdots a_m]$. By defining $\mu^n(\Sigma \backslash [a_1 \cdots a_m]_\infty) = 0$ this extends to a measure on Σ satisfying properties 1-5 above. To prove that μ^n is a finite measure, we note that

$$\mu^n(\Sigma) = \sum_{i=1}^{n} \sum_{k=0}^{l(\gamma_i)-m} \mu^n(\sigma^k[\gamma_i]) < \sum_{i=1}^{n} \sum_{k=0}^{l(\gamma_i)-m} \mu^n[a_1 \cdots a_m] < \infty,$$

since each γ_i contains some occurrence of $a_1 \cdots a_m$, and $\mu^n[a_1 \cdots a_m] = 1$.

□

Step 3: We now show that the sequence of measures μ_f^n have

$$h_{\mu_f^n}(\phi) + \int_{\Sigma_f} g d\mu_f^n \to_{n\to\infty} h_{\mu_f}(\phi) + \int_{\Sigma_f} g d\mu_f.$$

We begin by investigating the integral $\int f d\mu^n$. For $\underline{x}$ and $\underline{y}$ in $\sigma^k[\gamma_i]$, $|f(\underline{x}) - f(\underline{y})| \le var_{l(\gamma_i)-k}(f)$, because f has summable variation and depends only on future coordinates. For each $i \in \mathbb{N}$ we choose $\underline{x_i}$ in $[\gamma_i]$. Then

$$\begin{aligned} \int_{\sigma^k[\gamma_i]} f d\mu^n &\leq (f(\sigma^k(\underline{x_i})) + var_{l(\gamma_i)-k}(f)).\mu^n(\sigma^k[\gamma_i]) \\ &= f(\sigma^k(\underline{x_i}))\mu^n[\gamma_i] + var_{l(\gamma_i)-k}(f)\mu^n[\gamma_i] \end{aligned}$$

since $\mu^n(\sigma^k[\gamma_i]) = \mu^n[\gamma_i]$. The same argument works replacing μ^n with μ and approximating from below rather than above, giving

$$\int_{\sigma^k[\gamma_i]} f d\mu \geq f(\sigma^k(\underline{x_i}))\mu[\gamma_i] - var_{l(\gamma_i)-k}(f)\mu[\gamma_i].$$

Then summing we get

$$\sum_{k=0}^{l(\gamma_i)-m} \int_{\sigma^k[\gamma_i]} f d\mu^n \leq \left(\sum_{k=0}^{l(\gamma_i)-m} f(\sigma^k(\underline{x_i}))\mu^n[\gamma_i] \right) + \mu^n[\gamma_i].\sum_{j=m}^{\infty} var_j(f), \tag{4.6}$$

and

$$\begin{aligned} \sum_{k=0}^{l(\gamma_i)-m} \int_{\sigma^k[\gamma_i]} f d\mu &\geq \left(\sum_{k=0}^{l(\gamma_i)-m} f(\sigma^k(\underline{x_i}))\mu[\gamma_i] \right) - \mu[\gamma_i].\sum_{j=m}^{\infty} var_j(f) \\ &= q_n \left(\left(\sum_{k=0}^{l(\gamma_i)-m} f(\sigma^k(\underline{x_i}))\mu^n[\gamma_i] \right) - \mu^n[\gamma_i].\sum_{j=m}^{\infty} var_j(f) \right), \end{aligned}$$

giving

$$\sum_{k=0}^{l(\gamma_i)-m} f(\sigma^k(\underline{x_i}))\mu^n[\gamma_i] \leq \frac{1}{q_n} \sum_{k=0}^{l(\gamma_i)-m} \int_{\sigma^k[\gamma_i]} f d\mu + \mu^n[\gamma_i].\sum_{j=m}^{\infty} var_j(f). \tag{4.7}$$

Then, recalling that Σ can be partitioned by $\{\sigma^k[\gamma_i] : i \in \mathbb{N}, 0 \leq k \leq l(\gamma_i) - m\}$, we have that

$$\begin{aligned}\int_\Sigma f d\mu^n &= \sum_{i=1}^{n} \sum_{k=0}^{l(\gamma_i)-m} \left(\int_{\sigma^k[\gamma_i]} f d\mu^n \right) \\ &\leq \sum_{i=1}^{n} \left(\left(\sum_{k=0}^{l(\gamma_i)-m} f(\sigma^k(\underline{x_i}))\mu^n[\gamma_i] \right) + \mu^n[\gamma_i]. \sum_{j=m}^{\infty} var_j(f) \right).\end{aligned}$$

The second line here came from equation (4.6). Substituting in (4.7), we have that

$$\begin{aligned}\int_\Sigma f d\mu^n &\leq \sum_{i=1}^{n} \left(\left(\frac{1}{q_n} \sum_{k=0}^{l(\gamma_i)-m} \int_{\sigma^k[\gamma_i]} f d\mu + \mu^n[\gamma_i]. \sum_{j=m}^{\infty} var_j(f) \right) + \mu^n[\gamma_i]. \sum_{j=m}^{\infty} var_j(f) \right) \\ &= \frac{1}{q_n} \left(\sum_{i=1}^{n} \sum_{k=0}^{l(\gamma_i)-m} \int_{\sigma^k[\gamma_i]} f d\mu \right) + 2\mu^n[a_1 \cdots a_m] \sum_{j=m}^{\infty} var_j(f).\end{aligned}$$

Then, since $q_n \to 1$, $\sum_{k=m}^{\infty} var_k(f) < \delta$ and $\mu^n[a_1 \cdots a_m] = 1$, we can take limits as n tends to infinity in the above equation to get

$$\lim_{n\to\infty} \int_\Sigma f d\mu^n \leq \int_\Sigma f d\mu + 2\delta.$$

Repeating the argument but approximating $\int f d\mu^n$ from below and $\int f d\mu$ from above we get

$$\lim_{n\to\infty} \int_\Sigma f d\mu^n \geq \int_\Sigma f d\mu - 2\delta,$$

and an identical argument shows that

$$\int_\Sigma \Delta_g d\mu - 2\delta \leq \lim_{n\to\infty} \int_\Sigma \Delta_g d\mu^n \leq \int_\Sigma \Delta_g d\mu + 2\delta.$$

We now consider the entropy. Since $[a_1 \cdots a_m]$ is a sweep out set for σ, we have that

$$h_{\mu^n}(\sigma) = h_{\mu^n|_{[a_1\cdots a_m]}}(\sigma|_{[a_1\cdots a_m]}).$$

But $\mu^n|_{[a_1\cdots a_m]}$ is a Bernoulli measure on choices of γ_i, and so

$$\begin{aligned} h_{\mu^n}(\sigma) &= -\sum_{i=1}^{n} \mu^n[\gamma_i]\log(\mu^n[\gamma_i]) \\ &= -\frac{1}{q_n}\sum_{i=1}^{n} \mu[\gamma_i]\log(\mu[\gamma_i]) + \log(q_n). \end{aligned}$$

Hence, since $\lim_{n\to\infty} q_n = 1$, we have that

$$\lim_{n\to\infty} h_{\mu^n}(\sigma) = -\sum_{i=1}^{\infty} \mu[\gamma_i]\log(\mu[\gamma_i]).$$

Now $-\sum_{i=1}^{\infty} \mu[\gamma_i]\log(\mu[\gamma_i]) < \infty$, since otherwise $\lim_{n\to\infty} h_{\mu^n}(\sigma)$ would be infinite, giving $\lim_{n\to\infty} h_{\mu_f^n}(\phi) = \infty$ and contradicting the finiteness of $P_\phi(g)$. Then since

$$\begin{aligned} 0 < \int f d\mu - 2\delta \le \lim_{n\to\infty}\int f d\mu^n &\le \int f d\mu + 2\delta \\ \int \Delta_g d\mu - 2\delta \le \lim_{n\to\infty}\int \Delta_g d\mu^n &\le \int \Delta_g d\mu + 2\delta \text{ and} \\ \lim_{n\to\infty} h_{\mu^n}(\sigma) &= -\sum_{i=1}^{\infty} \mu[\gamma_i]\log(\mu[\gamma_i]), \end{aligned}$$

we can choose n and δ (and hence m) such that

$$\left| \frac{-\sum_{i=1}^{\infty} \mu[\gamma_i]\log(\mu[\gamma_i])}{\int_\Sigma f d\mu} + \frac{\int_\Sigma \Delta_g d\mu}{\int_\Sigma f d\mu} - \frac{h_{\mu^n}(\sigma)}{\int_\Sigma f d\mu^n} - \frac{\int_\Sigma \Delta_g d\mu^n}{\int_\Sigma f d\mu^n} \right| < \epsilon.$$

Now we recall that for a finite partition ζ, $\frac{1}{n}H_\mu(\sigma, \vee_{i=0}^{n}\sigma^{-i}\zeta)$ decreases to $h_\mu(\sigma,\zeta)$ (see theorem 4.10 of [Wal82]). Furthermore, for a generating partition ζ, $H_\mu(\sigma,\zeta) = h_\mu(\sigma)$, and the partition of cylinder sets of length one is a generating partition of Σ'. Then

$$-\sum_{i=1}^{\infty} \mu[\gamma_i]\log(\mu[\gamma_i]) \ge h_{\mu|_{[a_1\cdots a_m]}}(\sigma|_{[a_1\cdots a_m]}) = h_\mu(\sigma).$$

So we have

$$\frac{h_\mu(\sigma)}{\int f d\mu} + \frac{\int \Delta_g d\mu}{\int f d\mu} - \frac{h_{\mu^n}(\sigma)}{\int f d\mu^n} - \frac{\int \Delta_g d\mu^n}{\int f d\mu^n} < \epsilon,$$

giving

$$\left(h_{\mu_f}(\phi) + \int g d\mu_f\right) - \left(h_{\mu_f^n}(\phi) + \int g d\mu_f^n\right) < \epsilon$$

and hence

$$V(g) - \left(h_{\mu_f^n}(\phi) + \int g d\mu_f^n\right) < 2\epsilon$$

as required. Each of our μ^n are finite measures, so we scale them to be probability measures without affecting μ_f^n. This makes each μ_f^n an element of $\mathcal{E}^p_{\phi,g}$, and completes the proof. □

Lemmas 4.6.1 and 4.6.2 prove two different variational principles. Lemma 4.6.1 relates P_ϕ to a supremum taken over the set of flow invariant probability measures which are the lift of finite shift invariant measures, whereas Lemma 4.6.2 relates P_ϕ to the set of ergodic flow invariant measures. It is natural to ask whether we can state the variational principle as a supremum over flow invariant probability measures μ_f without the requirement that μ_f should be ergodic or the lift of some finite measure μ. Unfortunately, because Σ_f is non-compact, we have been unable to do this. We note that, in the case that the roof function f is bounded away from zero, flow invariant probability measures are automatically the lift of finite invariant measures on the base, and so we can state our variational principle in terms of $\mathcal{M}_\phi$.

4.7 Equilibrium States

Now that we have a coherent idea of topological pressure, it is natural to ask about equilibrium states, which we recall are measures $\mu \in \mathcal{M}_{\phi,g}$ satisfying

$$h_\mu(\phi) + \int g d\mu = \sup\{h_\nu(\phi) + \int g d\nu : \nu \in \mathcal{M}_{\phi,g}\}.$$

In the case of suspension flows with roof functions f bounded away from zero, the study of equilibrium states on Σ_f has been reduced to the study of equilibrium states on the base by the following result of Barreira and Iommi [BI06], which is a generalisation of an earlier result of Bowen and Ruelle in [BR75] for finite shifts.

Theorem 4.7.1. *Let $\Sigma, f : \Sigma \to \mathbb{R}^+$ and $g : \Sigma_f \to \mathbb{R}$ be as before, with the added assumption that f is bounded away from zero. Then the following two statements are equivalent*

1. *There exists an equilibrium state $\mu_f \in \mathcal{M}_{\phi,g}$ associated to g.*
2. *$P_\sigma(\Delta_g - P_\phi(g).f) = 0$ and there exists an equilibrium state $\mu \in \mathcal{M}_{\Sigma,f}$ associated to $\Delta_g - P_\phi(g).f$.*

In the case that these conditions hold, $\mu_f = \mathcal{L}(\mu)$.

This theorem no longer holds in the case of suspension flows where the roof function approaches zero. While the second condition still implies the first, it may be the case that an equilibrium state for the flow is the lift of an infinite invariant measure on the base. An example of this was given in section 4.2. In seeking a theory of equilibrium states for suspension flows whose roof functions may approach zero, we ask the following two questions.

Question 1: Is there a way of recognising whether a measure μ_f on Σ_f is an equilibrium state for some potential g by considering μ, f and Δ_g on Σ, even if $\mu(\Sigma) = \infty$?

Question 2: Is there a way of recognising whether there exists an equilibrium state μ_f on Σ_f for some potential g without using the base transformation?

Regarding question 2, we recall that for a suspension flow for a finite Markov shift there exists such a method. In [Bow72], Bowen showed that, for a Hölder continuous potential $g : \Sigma_f \to \mathbb{R}$, the sequence of measures $\mu_{t,g}$ defined below converges in the weak* topology to the equilibrium state associated to g. The measures $\mu_{t,g}$ are defined by

$$\mu_{t,g} := \frac{\sum_{(\underline{x},0)\in PO(t)} \delta_{\gamma(\underline{x})} \exp(g(\gamma(\underline{x})))\chi_{[a]}(\underline{x})}{\sum_{(\underline{x},0)\in PO(t)} \exp(g(\gamma(\underline{x})))\chi_{[a]}(\underline{x})},$$

where $PO(t)$ is the set of periodic orbits of period less than or equal to t and $\delta_{\gamma(\underline{x})}$ is the invariant measure on the periodic orbit $\gamma(\underline{x})$ passing through $\underline{x}$, with total mass $l(\gamma)$.

Furthermore, Hamenstädt proved in [Ham10] that the Teichmüller flow, which can be modelled as a suspension flow over a countable Markov shift, has a measure of maximal entropy which is equal to the weak star limit of the measures $\mu_{t,0}$. In future work I plan to investigate the relationship between the sequence of measures $\mu_{t,g}$ and equilibrium states associated to g. In particular, it seems reasonable to make the following conjecture.

Conjecture: Let Σ, f and g be as above. Then there exists an equilibrium state μ associated to g if and only if the sequence $\mu_{t,g}$ converges in the weak* topology, in which case the measures $\mu_{t,g}$ will converge to μ.

This would provide both new information about the way that periodic orbits are distributed and a new criterion for the existence of equilibrium states. We mention that putting $f = 1$ gives a corresponding conjecture for Markov shifts, which to the best of our knowledge is also new. So far equilibrium states for Markov shifts are only understood in the case that the potential is bounded and has summable variation, a positive answer to the above conjecture would give a significant new criterion for the existence of equilibrium states.

4.8 Applications to the Positive Geodesic Flow

In this section we explain how our results can be used to significantly simplify estimates of the topological entropy of the positive geodesic flow.

Let $\mathcal{H} = \{z \in \mathbb{C} : Im(z) > 0\}$ be the upper half plane equipped with the hyperbolic metric. The modular surface is defined by $M = \mathcal{H}/SL(2,\mathbb{Z})$. Coding methods for the geodesic flow on $\mathcal{M}$ have been the subject of much interest. One method of generating a code for a geodesic γ, the geometric code, involves tiling $\mathcal{H}$ with copies of the fundamental domain

$$F := \{z \in \mathbb{C} : -\frac{1}{2} \leq Re(z) \leq \frac{1}{2}, |z| > 1\},$$

and studying the sequence of edges of F crossed by γ. Alternatively geodesics can be coded by writing down the backwards continued fraction code of their attracting fixed point, the so called arithmetic code. Each of these coding methods allows us to model the geodesic flow as a suspension flow over a countable Markov shift. A survey of these coding methods is given by S. Katok in [Kat96]. In that paper, the set of geodesics for which the arithmetic and geometric codes are the same was studied. This is also the set of geodesics which are always clockwise oriented when mapped back in to the fundamental domain F. The geodesic flow restricted to this set is called the positive geodesic flow.

In [GK01], Gurevich and Katok modelled the positive geodesic flow as a suspension flow over a countable Markov shift with Hölder continuous roof function. By replacing f with the locally constant function $g(\underline{x}) := \sup\{f(\underline{y}) : \underline{y} \in [x_0]\}$ the authors were able to use the results of Savchenko [Sav98] and Polyakov [Pol01] to get a lower bound for the topological entropy of the flow (which they defined as the supremum of the metric entropies, and hence coincides with our notion of topological entropy). Similarly they used the infimum of the roof function on cylinders of length one to get an upper bound. This method was generalised by Ahmadi Dastjerdi and Lamei in [AL11] to give arbitrarily close approximation to the entropy. Their method was to give a sequence of representations of the geodesic flow as suspension flows in which the roof function becomes progressively more flat, and so g becomes progressively better as an approximation of f and the method of Gurevich and Katok gives increasingly good estimates to the entropy of the flow.

Our main result gives a simple way of estimating the topological entropy of the positive geodesic flow by measuring the growth rate of the number of periodic orbits. Unlike the method of [AL11], our method does not involve recoding the flow, because we do not need to approximate the flow by suspension flows with locally constant roof functions.

Chapter 5

Zero Temperature Limit Laws

5.1 Introduction

Given a dynamical system (X, T) and a function $f : X \to \mathbb{R}$, an equilibrium state associated to f is an invariant measure μ_f for which

$$h_{\mu_f} + \int_X f d\mu_f = \sup_{\nu \in \mathcal{M}_T} \left\{ h_\nu + \int f d\nu \right\}.$$

Under certain conditions on X, T and f, equilibrium states exist and are unique. If for some particular choices of X, T and f there exist unique equilibrium states μ_{tf} associated to the function $t.f$ for all $t > 0$, we can ask what happens to the measures μ_{tf} as t tends to infinity. Answers to such questions are broadly termed 'zero temperature limit laws', because of the following application to statistical mechanics.

If (X, T) is a model for a system of particles in which interactions between particles at temperature k are given by the potential f, then replacing f by $t.f$ corresponds to studying the same system at temperature $\frac{k}{t}$. Thus studying the equilibrium states μ_{tf} as t tends to infinity corresponds to studying the system of particles as temperature tends to absolute zero, and the existence of a limit point of μ_{tf} as t tends to infinity corresponds to the existence of a ground state for the system of particles.

The first occurrence of questions relating to zero temperature limit laws in the context of dynamical systems seems to be in the thesis of Coelho, [Coe90]. Here they were used as a way of finding maximising measures, that is measures μ for which the integral $\int f d\mu$ is as large as possible. Given two measure μ and ν, the metric entropies h_μ and h_ν do not depend on f or t, and so if $\int_X f d\mu > \int_X f d\nu$ then there will exist a T such that for all $t > T$ we have

$$h_\mu + \int_X t f d\mu > h_\nu + \int_X t f d\nu.$$

If the function $\mu \to \int_X f d\mu$ is upper semicontinuous with respect to the weak* topology on $\mathcal{M}_T$, as is the case for many systems including countable Markov shifts (see [JMU05]), any limit point of μ_{tf} must be a maximising measure for f. Ergodic optimisation, which is the study of maximising measures, is an active field of research and is one of our motivations for studying zero temperature limit laws. A good introduction to ergodic optimisation is given by Oliver Jenkinson's survey article [Jen06].

The study of zero temperature limit laws tends to focus around the following three questions.

1. Does μ_{tf} converge in the weak* topology as t tends to infinity?
2. If so, can the limit be identified?
3. What are the properties of the limit points of μ_{tf}?

In [Bré03], Brémont proved the convergence as t tends to infinity of the equilibrium states μ_{tf} associated to a locally constant potential f on a finite topologically mixing Markov shift. This proof used techniques from analytic geometry. The results of Brémont were extended by Leplaideur in [Lep05], using dynamical systems techniques to prove the convergence of the equilibrium states μ_{tf+g}, where f is locally constant and g Hölder continuous. However Chazottes and Hochman showed in [CH10] that if f is Hölder continuous then μ_{tf} need not converge.

There has also been interest in zero temperature limit laws for countable Markov shifts. In [Iom07], Iommi proved the convergence of equilibrium states μ_{tf} for a locally constant potential f on a countable renewal type shift. In [JMU05], Jenkinson, Mauldin and Urbański considered the equilibrium states μ_{tf} associated to Hölder continuous f on a countable Markov shift with suitable conditions to ensure the existence of equilibrium states, given below. They proved that μ_{tf} has at least one limit point.

If μ_{tf} does converge then finding the limit can be useful. In [CGU09], Chazottes, Gambaudo and Ugalde gave a simple algorithm to find the zero temperature limit of μ_{tf} for f locally constant on a finite Markov shift. However in [BLL10], Baraviera, Leplaideur and Lopes gave an example to show that, in the case of Hölder continuous functions on a finite Markov shift for which the zero temperature limit exists, the limit can behave counterintuitively as f varies.

In [JMU05], zero temperature limits were described as the most 'physically relevant' maximising measures. Further weight was given to this statement when, in [Mor07], Morris proved that any limit point of μ_{tf} has maximal entropy among the maximising measures of f. So in the case that there is a unique maximising measure, or that among maximising measures there is one with greater entropy than all the others, the sequence μ_{tf} will converge to this measure.

In this chapter we consider uniformly locally constant potentials on a countable

Markov shift under suitable conditions as given in [JMU05] to ensure the existence of equilibrium measures μ_{tf} for all t. We prove that the equilibrium states μ_{tf} converge as t tends to infinity and that their limit can be found by first reducing to a finite Markov shift and then using the algorithm given in [CGU09].

5.2 Set Up

We let (Σ, σ) be a two sided Markov shift over a countable alphabet $\mathcal{A}$ satisfying the big images and preimages property (BIP), as defined in chapter 2. Given a function $f : \Sigma \to \mathbb{R}$, we let $P(f)$ denote the Gurevich pressure of f, and recall that, for countable Markov shifts, a measure μ is called an equilibrium state if it satisfies

$$h_\mu + \int f d\mu = \sup\{h_\nu + \int f d\nu : \nu \in \mathcal{M}_\sigma, \int f d\nu > -\infty\},$$

noting that, in order to be well defined, the supremum is taken over a smaller class of measures than $\mathcal{M}_\sigma$.

The potential f is called uniformly locally constant if there exists an n for which $var_n(f) = 0$, giving $f(\underline{x}) = f(x_{-n} \cdots x_n)$ for all $\underline{x} \in \Sigma$. By recoding the shift and adding a coboundary if necessary, it is possible to assume that for a uniformly locally constant potential f we have $f(\underline{x}) = f(x_0 x_1)$.

We let $h(\mu)$ denote the metric entropy of μ and $\mathcal{M}_\sigma$ the set of σ invariant Borel probability measures on Σ. We define the weak* topology on $\mathcal{M}_\sigma$ by letting $\mu_n \to \mu$ if and only if for every bounded continuous function $f : \Sigma \to \mathbb{R}$ we have $\int_\Sigma f d\mu_n \to \int_\Sigma f d\mu$, as in Billingsley [Bil99].

We let Gibbs measures be defined as in chapter 2 and denote μ_f the Gibbs measure associated to potential f. For countable Markov shifts it is possible that for a

Gibbs measure μ_f we can have $h_\mu = \infty$ and $\int f d\mu = -\infty$, in which case the sum $h_\mu + \int f d\mu$ is not defined. Our conditions will ensure that for all $t \geq 2$, μ_{tf} is both the invariant Gibbs measure and the equilibrium state for tf. Such measures are sometimes termed Gibbs equilibrium states.

We assume that f has summable variation and finite topological pressure and that Σ satisfies BIP. This implies that tf has summable variation and finite topological pressure for all $t \geq 1$, and therefore that Gibbs measures μ_{tf} exist for all $t \geq 1$.

In our case that $f(\underline{x}) = f(x_0 x_1)$, f has summable variation if and only if

$$\sup\{|f(\underline{x}) - f(\underline{y})| : x_0 = y_0\} < \infty.$$

The following lemma will allow us to prove that the Gibbs measures μ_{tf} are also equilibrium states for $t \geq 2$.

Lemma 5.2.1. *Let (Σ, σ) be a topologically mixing Markov shift satisfying BIP and $f : \Sigma \to \mathbb{R}$ be uniformly locally constant and have summable variation and finite topological pressure. Then $\sum_{i\in\mathcal{A}} \exp(\sup f|_{[i]}) < \infty$.*

It was shown by Morris in [Mor07] that this implies that for all $t \geq 2$ we have

$$\sum_{i\in\mathcal{A}} \sup(tf|_{[i]}) \exp(\sup tf|_{[i]}) < \infty,$$

which we recall from chapter 2 gives us that μ_{tf} is also an equilibrium state. We now have enough conditions to ensure the existence of μ_{tf} for all t and to state our theorem.

Theorem 5.2.1. *Let (Σ, σ) be a topologically mixing Markov shift satisfying BIP and let $f : \Sigma \to \mathbb{R}$ be uniformly locally constant with summable variation and finite* *topological pressure. Then the equilibrium measures μ_{tf} exist for all $t \geq 2$ and converge in the weak* topology as t tends to infinity.*

It is known that in the case of a finite alphabet Markov shift and locally constant potential the zero temperature limit exists. Our method will be to relate μ_{tf} to the equilibrium states ν_{tf} of f on some finite subshift $\Sigma' \subset \Sigma$ and argue that, for any bounded continuous $g : \Sigma \to \mathbb{R}$, $\int_{\Sigma} g d\mu_{tf} - \int_{\Sigma'} \nu_{tf}(g) \to 0$ as $t \to \infty$, thus allowing us to use the convergence of the ν_{tf} on the finite subshift to imply the convergence of the μ_{tf} on the countable shift.

We now prove lemma 5.2.1

Proof. We know that Σ satisfies BIP, which we recall means that there exists a finite set $\mathcal{K} = \{k_1, \cdots, k_n\}$ such that for each $a \in \mathcal{A}$ there exist $i, j \in \{1, \cdots, n\}$ such that $k_i a k_j$ is an admissible word in Σ. For each pair $(i, j) \in \{1, \cdots, n\}^2$ we define the set $\mathcal{A}(k_i, k_j)$ to be the set of $a \in \mathcal{A}$ such that $k_i a k_j$ is an admissible word.

Since Σ is topologically mixing, there exists some finite word $x_1 \cdots x_n$ linking k_j to k_i, which gives that $k_i a k_j x_1 \cdots x_n k_i$ is an admissible word and can be extended to a periodic sequence $\underline{x}$ of period $n + 3$.

Then since $P(f)$ is finite, we have that $\sum_{\sigma^{n+3}(\underline{x})=\underline{x}} \exp(f^{n+3}(\underline{x}))\chi_{[k_i]}(\underline{x}) < \infty$. This implies that

$$\begin{aligned}\sum_{a\in\mathcal{A}(k_i,k_j)} \exp(f^{n+3}(k_i a k_j x_1 \cdots x_n k_i)) &= \exp(f^{n+1}(k_j x_1 \cdots x_n k_i)) \\ &\times \sum_{a\in\mathcal{A}(k_i,k_j)} \exp(f(k_i a) + f(a k_j)) \\ &< \infty,\end{aligned}$$

where we have used the fact that f is locally constant to split the summation.

Now $\{f(k_i a) : a \in \mathcal{A}(k_i, k_j)\}$ is bounded above and below since f has summable variation, and $\exp(f^{n+1}(k_j x_1 \cdots x_n k_i))$ is independent of $a \in \mathcal{A}(k_i, k_j)$. So the above line gives that

$$\sum_{a \in \mathcal{A}(k_i, k_j)} \exp(f(ak_j)) < \infty,$$

and so multiplying by $\exp(var_1(f))$ we see that

$$\sum_{a \in \mathcal{A}(k_i, k_j)} \exp(\sup f|_{[a]}) < \infty.$$

Each $i \in \mathcal{A}$ appears in at least one of the sets $A(k_i, k_j)$, and so summing over the finite set of pairs (k_i, k_j), we have that

$$\sum_{i \in \mathcal{A}} \exp(\sup f|_{[i]}) < \infty,$$

as required. □

5.3 Recasting the Question

In this section we recast the question as one about the convergence of ratios of certain sums. It was proved by Jenkinson, Mauldin and Urbański in [JMU06] that, given any pair (Σ, f) for which f has an equilibrium state μ_f, there exists at least one measure μ for which

$$\int f d\mu = \alpha(f) := \sup \left\{ \int f dm : m \in \mathcal{M}_\sigma \right\}.$$

Such a measure is called a maximising measure and the set of maximising measures is denoted $\mathcal{M}_{max}(f)$. Corollary 2.2.1 gives that there exists some $f' \sim f$ with $f'(\underline{x}) = f'(x_0 x_1)$ and $f' \leq \alpha(f)$. For ease of computation we replace f with $f' - \alpha(f')$, without affecting the equilibrium states of f. We now have that $\alpha(tf) = 0$ and $tf \leq 0$ for all $t \in \mathbb{R}$. We use this to identify a finite set of symbols such that any limit point of μ_{tf} must be supported on Σ restricted to this finite set of symbols.

Lemma 5.3.1. *There exists a finite subset* $I = \{i_1, \cdots, i_k\}$ *of* $\mathcal{A}$ *upon which* $\sup f|_{[i]} = 0$*, and a constant* $d > 0$ *such that* $\sup f|_{[i]} \leq -d$ *for all* $i \in \mathcal{A} \setminus I$*.*

Proof. There exists at least one maximising measure μ. This measure must give positive measure to the cylinder $[i]$ for at least one $i \in \mathcal{A}$. But since $f \leq 0$, any measure ν giving positive measure to a cylinder $[j]$ for which $\sup f|_{[j]} < 0$ must have $\int_\Sigma f d\nu < 0$. Then since $f \leq 0, \int f d\mu = 0$ and $\mu[i] > 0$, we must have $\sup f|_{[i]} = 0$.

The set of states I upon which $\sup f|_{[i]} = 0$ must be a finite set, otherwise $\sum_{i \in \mathcal{A}} \exp(\sup f|_{[i]})$ would be infinite, contradicting lemma 5.2.1. Indeed the same argument gives that for any $c \in \mathbb{R}$, the set $\{i \in \mathcal{A} : \sup f|_{[i]} > c\}$ must be finite.

We choose a constant $c < 0$ such that $\{\sup f|_{[i]} : i \in \mathcal{A}\} \cap (c, 0)$ is non-empty. As argued above, this set must be finite, and hence there exists a largest such value which we call $-d$. This completes the proof of the lemma. □

The next lemma helps us to reduce the task of showing that μ_{tf} converges to one of understanding the behaviour of μ_{tf} on I.

Lemma 5.3.2. $\lim_{t \to \infty} \mu_{tf}([i_1] \cup \cdots \cup [i_k]) = 1$

Proof. Recalling that $P(f) < \infty$, the variational principle gives the inequality

$$\sup \left\{ h_\mu : \mu \in \mathcal{M}_\phi, \int f d\mu = k \right\} \leq P(f) - k.$$

Then

$$\sup \left\{ h_\mu + t \int f d\mu : \mu \in \mathcal{M}_\phi, \int f d\mu = k \right\} \leq P(f) + (t-1)k.$$

But since there exists a maximising measure μ for which $\int f d\mu = 0$, and $h_\mu \geq 0$, we know

$$\sup\left\{h_\mu + t\int f d\mu : \mu \in \mathcal{M}_\phi, \int f d\mu > -\infty\right\} \geq 0.$$

Combining these equations gives

$$P(f) + (t-1)\int f d\mu_{tf} \geq 0.$$

Hence

$$\int f d\mu_{tf} \geq \frac{-P(f)}{t-1}.$$

Then since $f \leq -d$ on $([i_1] \cup \cdots \cup [i_k])^c$, we have

$$\mu_{tf}(([i_1] \cup \cdots \cup [i_k])^c) \leq \frac{P(f)}{(t-1)d}.$$

Letting t tend to infinity, this proves the lemma. □

The next lemma further simplifies the question of whether μ_{tf} converges.

Lemma 5.3.3. *If μ_{tf} fails to converge, then there must exist a finite word $a_1 \cdots a_n$ with $a_n = a_1$ such that $\mu_{tf}[a_1 \cdots a_n]$ fails to converge*

Proof. By the definition of weak star convergence, the sequence μ_{tf} fails to converge if and only if there exists some bounded continuous $g : \Sigma \to \mathbb{R}$ such that $\int g d\mu_{tf}$ fails to converge. This in turn must imply that there exists a set B for which $\mu_{tf}(B)$ fails to converge, and since our topology is generated by the cylinder sets, there must exist a set $[b_1 \cdots b_m]$ for which $\mu_{tf}[b_1 \cdots b_m]$ fails to converge.

Now we can take any symbol $a_1 \in \mathcal{A}$ and write $\mu_{tf}[b_1 \cdots b_m]$ as a countable summation of the measure of periodic words from a_1 to a_1 (this technique is explained in detail when it is used again at the end of this section). Then the non-convergence of $\mu_{tf}[b_1 \cdots b_m]$ implies that there exists at least one periodic word $[a_1 \cdots a_n]$ for which $\mu_{tf}[a_1 \cdots a_n]$ does not converge, proving the lemma. $\square$

We now show that the convergence of μ_{tf} is equivalent to a simpler condition.

Lemma 5.3.4. *The measures μ_{tf} converge if and only if $\lim_{t\to\infty} \mu_{tf}[a]$ exists for all $a \in I$.*

Proof. Given a word $\alpha = a_1 \cdots a_n$ we write as shorthand $\mu(\alpha) := \mu[a_1 \cdots a_n]$ and $f(\alpha) := \sum_{k=0}^{n-2} f(\sigma^k(a_1 \cdots a_n))$. We write

$$(tf - P(tf))(\underline{x}) := t.f(\underline{x}) - P(tf).$$

Then since f is locally constant the Gibbs inequality guarantees that, for a closed loop $a_1 \cdots a_n = \gamma$, we have

$$\mu_{tf}[\gamma] = \mu_{tf}[a_1] \exp((tf - P(tf))(\gamma)).$$

Since μ_{tf} is a probability measure, the above equation can be phrased as saying points in $[a]$ follow the path γ with probability $\exp((tf - P(tf))(\gamma))$.

It was proved by Morris in [Mor07] that $P(tf)$ is monotone and decreases to h, the maximal entropy of any maximising measure, and so $P(tf) - P((t+1)f) \to 0$. Then if $f(\gamma) = 0$, $\exp((tf - P(tf))(\gamma)$ will increase as $P(tf)$ decreases to h. If $f(\gamma) < 0$ then $\exp((tf - P(tf))(\gamma)$ will eventually decrease. In either case, $\exp((tf - P(tf))(\gamma)$ is eventually monotone, and so if $\mu_{tf}[a_1]$ converges then $\mu_{tf}[\gamma]$ must also converge.

It remains to prove that $\mu_{tf}[a]$ converges for all $a \in \mathcal{A}$. We know that $\mu_{tf}[a] \to 0$ for $a \notin I$, so we need only to prove the convergence of $\mu_{tf}[a]$ for $a \in I$. $\square$

Furthermore, since I is a finite set it is only necessary to check the convergence of the ratios $\frac{\mu_{tf}[a]}{\mu_{tf}[b]}$ for $a, b \in I$. The limit is allowed to be infinite.

The rest of this section is dedicated to proving that $\frac{\mu_{tf}[a]}{\mu_{tf}[b]}$ can be rewritten in such a way as to make the study of the limit as t tends to infinity comparatively straightforward. We fix $a, b \in I$.

Lemma 5.3.5. *The set*

$$A := \{\underline{x} \in \Sigma : x_n = a \text{ for infinitely many positive and negative } n\}$$

has $\mu_{tf}(A) = 1$ *for all* $t > 0$.

Proof. For all $t > 0$ we have $\mu_{tf}[a] > 0$, since μ_{tf} is a Gibbs measure and hence fully supported. Then by the Poincaré recurrence theorem, almost every point of $[a]$ returns to $[a]$ infinitely often under the actions of σ and σ^{-1}, and so $\mu_{tf}(A) \geq \mu_{tf}[a]$. Now A is a σ-invariant set of positive measure, and hence since μ_{tf} is ergodic we have $\mu_{tf}(A) = 1$. $\square$

In particular, this gives us that $\mu_{tf}[b] = \mu_{tf}\left([b] \cap A\right)$.

We refer to a finite word $x_0 \cdots x_n$ as a path from x_0 to x_n. If $x_0 = x_n$ we call $x_0 \cdots x_n$ a loop. We enumerate $(\gamma_i)_{i=1}^{\infty}$ the set of loops $\{\gamma = x_0 \cdots x_n, x_j = a \text{ iff } j \in \{0, n\}\}$, and for $\gamma_i = x_0 \cdots x_n$ we define $l(\gamma_i) = n$. Then A is partitioned by the set $\{\sigma^k[\gamma_i] : i \in \mathbb{N}, 1 \leq k \leq l(\gamma_i)\}$, and so $[b] \cap A$ is partitioned by the set

$$\left\{\sigma^k[\gamma_i] : i \in \mathbb{N}, 1 \leq k \leq l(\gamma_i), \sigma^k[\gamma_i] \in [b]\right\}$$

We let $N(b,\gamma_i)$ denote the number of occurrences of the symbol b in loop γ_i. Then

$$\begin{aligned}
\mu_{tf}[b] &= \sum_{i=1}^{\infty}\sum_{k=1}^{l(\gamma_i)} \mu_{tf}(\sigma^k[\gamma_i]).\chi_{[b]}(\sigma^k[\gamma_i]) \\
&= \sum_{i=1}^{\infty} \mu_{tf}[\gamma_i] N(b,\gamma_i) \\
&= \sum_{i=1}^{\infty} \mu_{tf}[a] \exp((tf - P(tf))(\gamma_i)) N(b,\gamma_i),
\end{aligned}$$

and hence

$$\frac{\mu_{tf}[b]}{\mu_{tf}[a]} = \sum_{i=1}^{\infty} \exp((tf - P(tf))(\gamma_i)) N(b,\gamma_i).$$

Now $(\gamma_i)_{i=1}^{\infty}$ is the set of all loops from a to a which have no intermediate occurrence of a. Loops γ_i which do not pass through b do not affect the above equation and we disregard them. All other loops γ_i can be split into three pieces, a path from a to b with no intermediate occurrence of a or b, $N(b,\gamma_i) - 1$ loops from b to b with no intermediate occurrence of a or b, and a path from b to a with no intermediate occurrence of a or b.

For $i,j \in \{a,b\}$ we denote by $\{\alpha : i \to j\}$ the set of paths $\alpha = \alpha_1 \cdots \alpha_m$ with $\alpha_1 = i, \alpha_m = j, \alpha_i \notin \{a,b\}$ for $i \in \{2,\cdots,m-1\}$. Then

$$\begin{aligned}
\frac{\mu_{tf}[b]}{\mu_{tf}[a]} &= \sum_{n=0}^{\infty}(n+1)\left(\sum_{\alpha:a\to b}\exp((tf-P(tf))(\alpha))\right)\left(\sum_{\alpha:b\to b}\exp((tf-P(tf))(\alpha))\right)^n \\
&\times \left(\sum_{\alpha:b\to a}\exp((tf-P(tf))(\alpha))\right). \qquad (5.1)
\end{aligned}$$

Each of these summations is a sum of positive terms. The finiteness of $\dfrac{\mu_{tf}[b]}{\mu_{tf}[a]}$

guarantees the finiteness of each summation, and so the sums must converge. This ensures that the following is well defined, for $i, j \in \{a, b\}$ we define

$$p_{ij}^t \quad = \sum_{\alpha : i \to j \in \Sigma} \exp((tf - P(tf))(\alpha)).$$

Rewriting equation (5.1) with this new notation we get

$$\begin{aligned} \frac{\mu_{tf}[b]}{\mu_{tf}[a]} &= p_{ab}^t p_{ba}^t \left(\sum_{n=0}^{\infty} (n+1)(p_{bb}^t)^n \right) \\ &= \frac{p_{ab}^t p_{ba}^t}{(1 - p_{bb}^t)^2}. \end{aligned} \tag{5.2}$$

Similarly

$$\frac{\mu_{tf}[a]}{\mu_{tf}[b]} = \frac{p_{ab}^t p_{ba}^t}{(1 - p_{aa}^t)^2}, \tag{5.3}$$

and so dividing equation (5.2) by equation (5.3) we see that

$$\frac{(\mu_{tf}[b])^2}{(\mu_{tf}[a])^2} = \frac{p_{ab}^t p_{ba}^t}{(1 - p_{bb}^t)^2} \times \frac{(1 - p_{aa}^t)^2}{p_{ab}^t p_{ba}^t} = \frac{(1 - p_{aa}^t)^2}{(1 - p_{bb}^t)^2}.$$

This gives us that

$$\frac{\mu_{tf}[b]}{\mu_{tf}[a]} = \frac{1 - p_{aa}^t}{1 - p_{bb}^t}.$$

Thus we have reduced the problem of showing that μ_{tf} converges to showing that the ratio $\dfrac{1 - p_{aa}^t}{1 - p_{bb}^t}$ converges for each $a, b \in I$. In the next section we define a relevant finite Markov shift and use the above form for the measure to compare equilibrium states on the finite shift with those on the countable shift.

5.4 A Related Finite Markov Shift

We have already seen that any limit points of μ_{tf} are fully supported on $\Sigma|_I$, which is a finite Markov shift. Our goal is to define a suitable finite Markov shift Σ' upon which equilibrium states for tf approximate μ_{tf}, and use the convergence of the equilibrium states restricted to Σ', which is guaranteed by the theorem of Brémont, to prove the convergence of μ_{tf}. The following three definitions give such a Σ'.

Definition 5.4.1. *Given Σ, f and $k \in \mathbb{R}$ we define Σ_k to be the subshift of finite type Σ restricted to the set of sequences $(x_n)_{n=-\infty}^{\infty}$ for which each $x_n \in \{i \in \mathcal{A} : \sup_{\underline{x}\in[i]}\{f(\underline{x})\} \geq k\}$.*

Definition 5.4.2. *We let $c > 0$ and $N \in \mathbb{N}$ be constants such that for each $a, b \in I$:*

1. *There exists a loop γ in Σ passing through a and b with $f(\gamma) \geq -c$ and $l(\gamma) \leq N$.*
2. *If there exists a loop γ in Σ passing through a and b and avoiding some set $I' \subset I$, then there exists such a loop with $f(\gamma) > -c$ and $l(\gamma) \leq N$.*
3. *The transitive component of Σ_{-c} containing all of I is topologically mixing.*

We choose for each I' a loop passing through a and b and avoiding I', should such a loop exist. We let $-c$ be the minimum value of $f(\gamma)$ for any of these loops and N be the maximum length of any of the loops. Since the set of subsets I' of I is finite, both c and N are finite. In considering p_{aa}^t we are interested in loops from a to a which avoid b, part 2 of the above definition is necessary in considering paths which avoid various subsets of I.

We say that a and b are in the same component of Σ_k if there exists some $n \in \mathbb{N}$ such that $\sigma^{-n}[a] \cap [b] \neq \phi$, where $[a]$ and $[b]$ are cylinder sets in Σ_k. For any pair $a, b \in I$

there exists a loop $\gamma \in \Sigma$ passing through a and b, and hence by the definition of c there exists such a path with $f(\gamma) > -c$. Therefore for any $k < -c$ there exists a component of Σ_k containing all of the elements of I.

Definition 5.4.3. *We define Σ' to be Σ_{-7c} restricted to the transitive component containing I.*

(Σ', σ) is a finite topologically mixing Markov shift, and hence there exist equilibrium states ν_{tf} associated to $tf|_{\Sigma'}$. The term $P(tf)$ is now ambiguous, we let P_{tf} denote the topological pressure of the potential tf on Σ and Q_{tf} denote the pressure of tf on Σ'. We have that $P_{tf} \geq Q_{tf}$.

Our choice of $-7c$ is purely to make the following analysis more simple. In fact, choosing $-c$ would be sufficient, we show that the behaviour of μ_{tf} is mirrored by the behaviour of ν_{tf} on the finite shift Σ', but once we have reduced to the finite case we can use the algorithm of Chazottes, Gambaudo and Ugalde in [CGU09], which confirms that in order to find the zero temperature limit of the equilibrium states on Σ' it is enough to look at Σ_{-c}.

Defining $q^t_{ij} = \sum_{\alpha: i \to j \in \Sigma'} \exp((tf - Q_{tf})(\alpha))$, the analysis of the previous section yields

$$\frac{\nu_{tf}[b]}{\nu_{tf}[a]} = \frac{1 - q^t_{aa}}{1 - q^t_{bb}}.$$

Now the convergence of $\dfrac{\nu_{tf}[b]}{\nu_{tf}[a]}$, which is the main result of [Bré03], ensures the convergence of $\dfrac{1 - q^t_{aa}}{1 - q^t_{bb}}$. We will use this to prove the convergence of $\dfrac{1 - p^t_{aa}}{1 - p^t_{bb}}$ and hence of $\dfrac{\mu_{tf}[b]}{\mu_{tf}[a]}$.

5.4.1 Convergence

We let $a(t) \sim b(t)$ mean that $\lim_{t\to\infty} \dfrac{a(t)}{b(t)} = 1$.

We introduce a third term, r_{ij}^t, for summation over paths between i and j, and recall the earlier definitions for comparison. There are two causes for the difference between p_{ij}^t and q_{ij}^t, that we change the pressure from P_{tf} to Q_{tf}, and that we are summing over a different set of paths. While r_{ij}^t has no physical or probabilistic interpretation, it allows us to separate out these two effects.

$$\begin{aligned}
p_{ij}^t &= \sum_{\alpha:i\to j\in\Sigma} \exp((tf - P_{tf})(\alpha)) \\
q_{ij}^t &= \sum_{\alpha:i\to j\in\Sigma'} \exp(tf - Q_{tf})(\alpha)) \\
r_{ij}^t &= \sum_{\alpha:i\to j\in\Sigma'} \exp(tf - P_{tf})(\alpha)).
\end{aligned}$$

We have $r_{ij}^t \le p_{ij}^t$, since the set of paths from i to j that lie entirely in Σ' is a subset of those in Σ. Furthermore, $r_{ij}^t \le q_{ij}^t$ because $Q_{tf} \le P_{tf}$.

Since a and b were arbitrary, statements for p_{aa}^t, q_{aa}^t and r_{aa}^t automatically carry over to p_{bb}^t, q_{bb}^t and r_{bb}^t. This rest of the section is structured as follows.

1. Find a lower bound for $1 - p_{aa}^t$ and $1 - q_{aa}^t$. (Lemma 5.4.1).

2. Show that p_{aa}^t and r_{aa}^t are close. (Lemma 5.4.2, proof deferred until the next section).

3. Combining the results of steps 1 and 2, infer that $1 - p_{aa}^t \sim 1 - r_{aa}^t$ (Lemma 5.4.3).

4. Prove that the sum $r_{aa}^t + \dfrac{r_{ab}^t r_{ba}^t}{1 - r_{bb}^t}$ is close to 1 (Lemma 5.4.4). This is done by observing that the corresponding sum for p_{ij}^t is equal to one, and then using step 2.

5. Combining steps 1 and 4, infer that $1 - q^t_{aa} \sim 1 - r^t_{aa}$ (Lemma 5.4.5).

6. Combining steps 3 and 5, conclude that $1 - p^t_{aa} \sim 1 - q^t_{aa}$, and show that this proves our main theorem.

To prove that $p^t_{aa} \sim q^t_{aa}$ is relatively straightforward, but for our purposes we need to prove that $1 - p^t_{aa} \sim 1 - q^t_{aa}$. To this end, it is necessary to find lower bounds on $1 - p^t_{aa}$ and $1 - q^t_{aa}$.

Lemma 5.4.1. $1 - p^t_{aa} \geq \exp(-tc - NP_{tf})$

Proof. By the Poincaré recurrence theorem, the probability, with respect to μ_{tf}, that a path from a returns to a eventually is one. We split this into p^t_{aa}, the probability that a path from a goes to a without passing through b, and $p^t_{ab}p^t_{ba}\sum_{n=0}^{\infty}(p^t_{bb})^n$, the probability that a path returns to a passing through b at least once. So

$$p^t_{aa} + p^t_{ab}p^t_{ba}\sum_{n=0}^{\infty}(p^t_{bb})^n = p^t_{aa} + \frac{p^t_{ab}p^t_{ba}}{1 - p^t_{bb}} = 1.$$

We recall that by the definition of c there exists some path γ from a to a passing through b with $f(\gamma) \geq -c$ and $l(\gamma) \leq N$. This path is included in the summation $p^t_{ab}p^t_{ba}\sum_{n=0}^{\infty} p^t_{bb}$, so

$$1 - p^t_{aa} = p^t_{ab}p^t_{ba}\sum_{n=0}^{\infty}(p^t_{bb})^n \geq \exp(-tc - NP_{tf}).$$

The same arguments work for $1 - q^t_{aa}$, $1 - q^t_{bb}$ and $1 - p^t_{bb}$. □

We also note that, by the definition of Σ', $p^t_{aa} = 0$ if and only if there is no path $\alpha : a \to a$ which avoids b. If there does exist such a path, then there exists one of length less than or equal to N and with $\exp(tf - P_{tf})(\alpha) \geq \exp(-tc - NP_{tf})$. In this case $p^t_{aa} \geq \exp(-tc - NP_{tf})$.

We need to use the following technical lemma, the proof of which is deferred to the final section.

Lemma 5.4.2. *There exists a $K \in \mathbb{R}$ such that for all $t > 0$ we have*

$$\begin{aligned} p^t_{aa} &\leq r^t_{aa} + K(\exp(-3ct)), \\ p^t_{bb} &\leq r^t_{bb} + K(\exp(-3ct)) \text{ and} \\ p^t_{ab}p^t_{ba} &\leq r^t_{ab}r^t_{ba} + K(\exp(-3ct)). \end{aligned}$$

This allows us to prove in a very simple manner the asymptotic convergence of the ratio $\frac{1-p^t_{aa}}{1-r^t_{aa}}$ to 1.

Lemma 5.4.3. $1 - p^t_{aa} \sim 1 - r^t_{aa}$, $1 - p^t_{bb} \sim 1 - r^t_{bb}$

Proof. We have that $1 - p^t_{aa} \geq \exp(-ct - NP_{tf})$. Then

$$1 \leq \frac{1-r^t_{aa}}{1-p^t_{aa}} = 1 + \frac{p^t_{aa} - r^t_{aa}}{1-p^t_{aa}} \leq 1 + \frac{K(\exp(-3ct)}{\exp(-ct - NP_{tf})} \to 1$$

giving $1 - p^t_{aa} \sim 1 - r^t_{aa}$. Identical arguments work for p^t_{bb}. □

The following two lemmas prove that $1 - r^t_{aa} \sim 1 - q^t_{aa}$.

Lemma 5.4.4.

$$r^t_{aa} + \frac{r^t_{ab}r^t_{ba}}{1-r^t_{bb}} \geq 1 - o(\exp(-ct))$$

Proof. We assume that $p^t_{aa} > 0$. Using lemma 5.4.2 we have

$$\begin{aligned} r^t_{aa} + \frac{r^t_{ab}r^t_{ba}}{1-r^t_{bb}} &\geq p^t_{aa} - K(\exp(-3ct)) + \frac{p^t_{ab}p^t_{ba} - K(\exp(-3ct))}{1-p^t_{bb} + K(\exp(-3ct))} \\ &= p^t_{aa}\left(1 - \frac{K(\exp(-3ct))}{p^t_{aa}}\right) + \frac{p^t_{ab}p^t_{ba}}{1-p^t_{bb}}\left(\frac{1 - \frac{K(\exp(-3ct)}{p^t_{ab}p^t_{ba}}}{1 + \frac{K(\exp(-3ct))}{1-p^t_{bb}}}\right) \end{aligned}$$

We have that p^t_{aa}, $p^t_{ab}p^t_{ba}$ and $1-p^t_{bb}$ are all greater than $\exp(-ct-NP_{tf})$. So using the above line we have

$$\begin{aligned}
r^t_{aa} + \frac{r^t_{ab}r^t_{ba}}{1-r^t_{bb}} &\geq p^t_{aa}\left(1-K\exp(-3ct+ct+NP_{tf})\right) \\
&+ \left(\frac{p^t_{ab}p^t_{ba}}{1-p^t_{bb}}\right)\left(\frac{1-K(\exp(-3ct+ct+NP_{tf}))}{1+K(\exp(-3ct+ct+NP_{tf}))}\right) \\
&\geq \left(p^t_{aa} + \frac{p^t_{ab}p^t_{ba}}{1-p^t_{bb}}\right)\left(\frac{1-K(\exp(-3ct+ct+NP_{tf}))}{1+K(\exp(-3ct+ct+NP_{tf}))}\right) \\
&= 1-o(\exp(-ct)).
\end{aligned} \tag{5.4}$$

If $p^t_{aa}=0$ then $r^t_{aa}=0$ and we have that $r^t_{aa} \geq p^t_{aa}\left(1-K\exp(-3ct+ct+NP_{tf})\right)$, and so equation (5.4) still holds and we complete the proof as in the case that $p^t_{aa}>0$. □

Finally, the following lemma gives $1-q^t_{aa} \sim 1-r^t_{aa}$. Combining this with lemma 5.4.3, which gives $1-r^t_{aa} \sim 1-p^t_{aa}$, we have the required asymptotic relation between $1-q^t_{aa}$ and $1-p^t_{aa}$.

Lemma 5.4.5. $1-r^t_{aa} \sim 1-q^t_{aa}$, $1-r^t_{bb} \sim 1-q^t_{bb}$.

Proof. Since $P_{tf} \geq Q_{tf}$ we have immediately that $1-r^t_{aa} \geq 1-q^t_{aa}$. We consider the other direction.

Substituting $q^t_{aa} + \dfrac{q^t_{ab}q^t_{ba}}{1-q^t_{bb}} = 1$ into the result of lemma 5.4.4 gives

$$r^t_{aa} + \frac{r^t_{ab}r^t_{ba}}{1-r^t_{bb}} \geq q^t_{aa} + \frac{q^t_{ab}q^t_{ba}}{1-q^t_{bb}} - o(\exp(-ct)),$$

and so

$$1 - r_{aa}^t \;\; \leq \;\; 1 - q_{aa}^t - \frac{q_{ab}^t q_{ba}^t}{1 - q_{bb}^t} + \frac{r_{ab}^t r_{ba}^t}{1 - r_{bb}^t} + o(\exp(-ct)).$$

Then since $r_{ij}^t \leq q_{ij}^t$ we have $\frac{r_{ab}^t r_{ba}^t}{1 - r_{bb}^t} \leq \frac{q_{ab}^t q_{ba}^t}{1 - q_{bb}^t}$, and hence

$$1 - q_{aa}^t \leq 1 - r_{aa}^t \leq 1 - q_{aa}^t + o(\exp(-ct)).$$

Finally using $1 - q_{aa}^t \geq \exp(-ct - NP_{tf})$ we conclude that

$$1 - r_{aa}^t \sim 1 - q_{aa}^t.$$

□

Then combining lemmas 5.4.3 and 5.4.5 we have

$$\frac{\mu_{tf}[b]}{\mu_{tf}[a]} = \frac{1 - p_{aa}^t}{1 - p_{bb}^t} \sim \frac{1 - r_{aa}^t}{1 - r_{bb}^t} \sim \frac{1 - q_{aa}^t}{1 - q_{bb}^t} = \frac{\nu_{tf}[b]}{\nu_{tf}[a]}$$

which converges, and so the limit $\lim_{t\to\infty} \frac{\mu_{tf}[b]}{\mu_{tf}[a]}$ exists, giving us finally that $\lim_{t\to\infty} \mu_{tf}$ exists and proving theorem 5.2.1.

The exact value of $\lim_{t\to\infty} \nu_{tf}$ is given by an algorithm in [CGU09] which terminates after finitely many steps. Then since $\lim_{t\to\infty} \mu_{tf}$ gives the same measure, we have that the zero temperature limit for the countable case can be given by reducing Σ to Σ' and then following the same algorithm.

5.5 Proof of Technical Lemma

The following technical lemma was stated earlier and used in the proof of theorem 5.2.1.

Lemma 5.4.2. *There exists a* $K \in \mathbb{R}$ *such that* $p^t_{aa} \leq r^t_{aa} + K(\exp(-3ct))$, $p^t_{bb} \leq r^t_{bb} + K(\exp(-3ct))$ *and* $p^t_{ab}p^t_{ba} \leq r^t_{ab}r^t_{ba} + K(\exp(-3ct))$ *for all* t.

Before proving the lemma, we explain why the proof is relatively technical. Hopefully this will also go some way to explaining the direction taken in the proof.

The difference between p^t_{aa} and r^t_{aa} is that p^t_{aa} is a summation over a set of paths in Σ, whereas r^t_{aa} sums only over the intersection of that set of paths with Σ'. But by the definition of Σ', any path $\alpha : a \to a$ which exits Σ' must necessarily have $f(\alpha) \leq -7c$, while there exists a path $\alpha_0 : a \to a$ contained in Σ' with $f(\alpha_0) = 0$.

This might tempt one into thinking that, for any $\alpha : a \to a \in \Sigma \backslash \Sigma'$, $\dfrac{\exp((tf - P_{tf})(\alpha))}{\exp((tf - P_{tf})(\alpha_0))}$ decreases as t increases. However this is not always true. P_{tf} is decreasing, and so $-P_{tf}(\alpha) = -(l(\alpha) - 1)P_{tf}$ is increasing, and the rate of increase depends on the length of the loop α.

If the loop α is much longer than α_0 then the effect of the increase of $-P_{tf}(\alpha)$ as t increases may be large enough to compensate for the decrease of $tf(\alpha)$ for sufficiently small t. This effect forces us to pay careful attention to the length of loops that we are dealing with and is the reason that the following proof becomes technical.

Proof. We prove the inequality for p^t_{aa}, which extends to the case of p^t_{bb} since a and b were arbitrary. We explain at the end of the proof why the same argument also works for the product $p^t_{ab}p^t_{ba}$.

For any path $\alpha : a \to a$ we let $n(\alpha)$ be the number of occurrences of elements of I

in α. We define the set X^n_{aa} to be the set of possible sequences $a = i_1, i_2, \cdots, i_n = a$ of elements of I in paths $\alpha : a \to a$ with $n(\alpha) = n$. Then, writing $\alpha : i_k \hookrightarrow i_{k+1}$ for paths α from i_k to i_{k+1} not passing through any other element of I, we have

$$\begin{aligned} p^t_{aa} &= \sum_{n=2}^{\infty} \sum_{i_1,\cdots,i_n \in X^n_{aa}} \prod_{k=1}^{n-1} \left(\sum_{\alpha : i_k \hookrightarrow i_{k+1} \in \Sigma} \exp((tf - P_{tf})(\alpha)) \right) \text{ and} \\ r^t_{aa} &= \sum_{n=2}^{\infty} \sum_{i_1,\cdots,i_n \in X^n_{aa}} \prod_{k=1}^{n-1} \left(\sum_{\alpha : i_k \hookrightarrow i_{k+1} \in \Sigma'} \exp((tf - P_{tf})(\alpha)) \right). \end{aligned}$$

We define

$$p^t_{aa}(n) = \sum_{i_1,\cdots,i_n \in X^n_{aa}} \prod_{k=1}^{n-1} \left(\sum_{\alpha : i_k \hookrightarrow i_{k+1} \in \Sigma} \exp((tf - P_{tf})(\alpha)) \right),$$

this is the summation p^t_{aa} restricted to those paths α with $n(\alpha) = n$. We let $r^t_{aa}(n)$ be defined similarly. We define

$$\epsilon(n) = \frac{p^t_{aa}(n)}{r^t_{aa}(n)} \geq 1.$$

Then

$$0 \leq p^t_{aa} - r^t_{aa} = \sum_{n=1}^{\infty} p^t_{aa}(n) \left(1 - \frac{1}{\epsilon(n)} \right).$$

We now require another technical lemma, which is proved immediately after this proof is completed. This allows us to prove that $p^t_{aa}(n)$ decreases in n with a certain rate, giving return time statistics for the set $[a]$.

Lemma 5.5.1. *There exists a K_2 such that, for $r \geq 0$ and $n \in \{r|I|, r|I| + 1, \cdots, (r+1)|I| - 1\}$,*

1. $p^t_{aa}(n) \leq (1 - \exp(-ct - NP_{tf}))^r$

2. $\epsilon(n) \le (1 + K_2 \exp(-5ct))^r$ *if* $r \ge 1$

3. $\epsilon(n) \le (1 + K_2 \exp(-5ct))$ *for* $n \in \{1, \cdots, |I| - 1\}$

Then using this lemma we have

$$\begin{aligned} p^t_{aa} - r^t_{aa} &= \sum_{n=1}^{\infty} p^t_{aa}(n) \left(1 - \frac{1}{\epsilon(n)}\right) \\ &\le \sum_{n=1}^{\infty} p^t_{aa}(n) \left(\epsilon(n) - 1\right) \\ &\le |I| \sum_{r=0}^{\infty} (1 - \exp(-ct - NP_{tf}))^r \left((1 + K_2 \exp(-5ct))^r - 1\right) \\ &= \frac{|I|}{1 - (1 - \exp(-ct - NP_{tf}))(1 + K_2 \exp(-5ct))} - \frac{|I|}{\exp(-ct - NP_{tf})} \\ &\le \frac{|I|}{\exp(-ct - NP_{tf}) - K_2 \exp(-5ct))} - \frac{|I|}{\exp(-ct - NP_{tf})} \\ &\le K \exp(-3ct) \end{aligned}$$

for sufficiently large K, completing the proof of the technical lemma. □

It remains only to prove the three claims of lemma 5.5.1.

Proof. **Claim 1:**

To find an upper bound on $p^t_{aa}(n)$, we in fact find an upper bound on $\sum_{j=n}^{\infty} p^t_{aa}(j)$. By the definition of c, there exists for any i_k a closed loop γ based at i_k passing through a, avoiding b, and with $f(\gamma) \ge -c, l(\gamma) \le N$. We can remove any subloops from γ without decreasing $f(\gamma) \ge -c$, since $f \le 0$. Then γ contains a path from i_k to a passing through at most $|I|$ elements of I.

Elements of $[i_k]$ follow path γ with (μ_{tf}) probability $\exp((tf - P_{tf})(\gamma)) \ge \exp(-ct - NP_{tf})$. Then in particular, the probability that an element of $[i_k]$ passes through at most $|I|$ elements of I before returning to $[a]$ is greater than or equal to $\exp(-ct -$

$NP_{tf})$. So $\sum_{k=1}^{\infty} p_{aa}^t(k) \leq 1$ and

$$\sum_{k=m+|I|}^{\infty} p_{aa}^t(k) \leq (1 - \exp(-tc - NP_{tf})) \sum_{k=m}^{\infty} p_{aa}^t(k),$$

giving that for $n \in \{r|I|, r|I|+1, \cdots, (r+1)|I|-1\}$,

$$p_{aa}^t(n) \leq \sum_{k=r|I|}^{\infty} p_{aa}^t(k) \leq (1 - \exp(-tc - NP_{tf}))^r.$$

Claim 2: We recall that P_{tf} decreases to h, the maximum entropy of any maximising measure, and that $d > 0$ is such that $\sup f|_{[i]} \leq -d$ for all $i \in \mathcal{A} \setminus I$. We let T be such that $P_{Tf} < h + d$. Then we have that for all $t > T$,

$$-d < P_{(t+1)f} - P_{tf} < 0.$$

We consider $(tf - P_{tf})$ evaluated along a path $\alpha = \alpha_0 \cdots \alpha_m : i_k \hookrightarrow i_{k+1} \in \Sigma \backslash \Sigma', m \geq 2$. We have $f(\alpha_0\alpha_1) \leq 0$ and $f(\alpha_n\alpha_{n+1}) < -d$ for $1 \leq n < m$, because α_n is not an element of I for $1 \leq n < m$. Furthermore, since $\alpha \in \Sigma \setminus \Sigma'$, there exists at least one n for which $f(\alpha_n\alpha_{n+1}) < -7c$. So

$$\begin{aligned}
((t+1)f - P_{(t+1)f})(\alpha) - (tf - P_{tf})(\alpha) &= (f - P_{(t+1)f} + P_{tf})(\alpha) \\
&= m(P_{tf} - P_{(t+1)f}) + f(\alpha) \\
&\leq md - 7c - (m-1)(d) \\
&= -7c + d \\
&\leq -6c
\end{aligned}$$

for $t > T$. We define

$$K_1 := \exp(6cT) \sup_{i_k, i_{k+1}} \sum_{\alpha : i_k \hookrightarrow i_{k+1} \in \Sigma \setminus \Sigma'} \exp((Tf - P_{Tf})(\alpha))$$

and see that for any choices of i_k, i_{k+1} and for any $t > T$,

$$\sum_{\alpha : i_k \hookrightarrow i_{k+1} \in \Sigma \setminus \Sigma'} \exp((tf - P_{tf})(\alpha)) \leq K_1 \exp(-6ct).$$

Now by the definition of c there exists some path $\beta : i_k \hookrightarrow i_{k+1} \in \Sigma'$ with $f(\beta) \geq -c$. So for $t > T$,

$$\sum_{\alpha : i_k \hookrightarrow i_{k+1} \in \Sigma \setminus \Sigma'} \exp((tf - P_{tf})(\alpha)) \leq K_1 \exp(-5ct) \left(\sum_{\alpha : i_k \hookrightarrow i_{k+1} \in \Sigma'} \exp((tf - P_{tf})(\alpha)) \right)$$

giving

$$\sum_{\alpha : i_k \hookrightarrow i_{k+1} \in \Sigma} \exp((tf - P_{tf})(\alpha)) \leq (1 + K_1 \exp(-5ct)) \left(\sum_{\alpha : i_k \hookrightarrow i_{k+1} \in \Sigma'} \exp((tf - P_{tf})(\alpha)) \right).$$

So for any $n \in \{r|I|, r|I|+1, \cdots, (r+1)|I|-1\}$ we have

$$\begin{aligned} \frac{p^t_{aa}(n)}{r^t_{aa}(n)} &\leq (1 + K_1 \exp(-5ct))^n \\ &\leq \left((1 + K_1 \exp(-5ct))^{2|I|}\right)^r \end{aligned}$$

for $r \geq 1$.

Now expanding $(1 + K_1 \exp(-5ct))^{2|I|}$ we get $2^{2|I|}$ terms. One of these terms is 1, and the rest are of the form $(K_1 \exp(-5ct))^j$, $j \in \{1, \cdots, 2|I|\}$.

If $K_1 \leq 1$ then we put $K_2 = 2^{2|I|-1}$. If $K_1 > 1$ we put $K_2 = 2^{2|I|-1} K_1^{2|I|}$.

In either case we have $(1 + K_1 \exp(-5ct))^{2|I|} \leq 1 + K_2 \exp(-5ct)$, and hence

$$\frac{p^t_{aa}(n)}{r^t_{aa}(n)} \leq (1 + K_2 \exp(-5ct))^r$$

for $r \geq 1$, completing the proof of claim 2.

Claim 3:

Following the above analysis, we have that, for $n \in \{1, \cdots, |I| - 1\}$,

$$\begin{aligned} \frac{p^t_{aa}(n)}{r^t_{aa}(n)} &\leq (1 + K_1 \exp(-5ct))^n \\ &\leq (1 + K_1 \exp(-5ct))^{|I|} \\ &\leq 1 + K_2 \exp(-5ct). \end{aligned}$$

This completes the proof of claim 3 and hence of lemma 5.5.1. □

We also mention that the above proof for p^t_{aa} and p^t_{bb} extends to the product $p^t_{ab}p^t_{ba}$. Where statements were made about loops passing through a and avoiding b we replace them with statements about loops passing through a and b, allowing us to replace p^t_{aa} with $p^t_{ab}p^t_{ba}$. This was required to prove the statements involving p^t_{ab} in lemma 5.4.2.

Chapter 6

Factors of Gibbs Measures for Subshifts of Finite Type

In this chapter we consider a map Π from a subshift of finite type Σ_1 to another subshift. Given a Gibbs measure μ on Σ_1 we ask what can be said about the image measure $\nu := \mu \circ \Pi^{-1}$. We give sufficient conditions to ensure that the image measure ν is a Gibbs measure. We also give an example of a map Π which does not satisfy our conditions and for which the resulting measure ν is not a Gibbs measure.

Factors of Markov shifts appear in many natural situations. For example, if we observe a system with Markov dynamics, but our observation is imperfect and two states in the system are indistinguishable, then we do not see the true transformation. Instead we see some factor transformation on the set of equivalence classes of indistinguishable states. This observed transformation may not be Markov even if the original transformation is, and it is for this reason that such factor transformations are referred to as hidden Markov processes. Further examples of hidden Markov processes include the transmission of codes down a noisy channel which corrupts information, mutations in DNA sequences, and reductions of the number of colours or pixels in digital images. Recent survey articles by Boyle and Petersen [BP11] and by Verbitskiy [Ver11] give a good introduction to hidden Markov processes and their thermodynamic formalism.

Gibbs measures play an important role in the study of symbolic dynamical systems, but the image of a Gibbs measure need not always be a Gibbs measure. In studying a dynamical system with respect to some invariant measure μ, the knowledge that μ is a Gibbs measure allows one to use a variety of techniques to study the dynamical system. For that reason, knowledge of whether the image of a Gibbs measure is still a Gibbs measure is useful in the study of factors of Markov shifts.

Further motivation for studying this problem is drawn from questions of renormalizations of Gibbs measures in statistical mechanics. This is discussed in section 6.6.

In the case that μ is a Markov measure, sufficient conditions for ν to be a Gibbs measure were given by Chazottes and Ugalde in [CU03]. These results were extended to deal with the case that μ is a Gibbs measure and Σ_1 is a full shift in work by Chazottes and Ugalde [CU11], Verbitskiy [Ver11] and myself and Pollicott [KP11]. In this chapter we discuss the most general case of a Gibbs measure supported on a subshift of finite type.

6.0.1 Preliminaries and Technical Hypotheses

Recalling definitions from chapter 2, we let Σ_1 be a subshift of finite type. We let μ be a Gibbs measure supported on Σ_1 associated to potential ψ_1. The potential ψ_1 must be continuous in order to satisfy inequality (2.1), but we do not impose any extra requirements on its regularity. By replacing ψ_1 with $\hat{\psi}_1 = \psi_1 - P(\psi_1)$ we have that μ is a Gibbs measure associated to $\hat{\psi}_1$ with $P(\hat{\psi}_1) = 0$. This allows us to consider only potentials ψ_1 for which the pressure $P(\psi_1)$ equals zero. We now define our map Π.

Definition 6.0.1. *Suppose we have a map* Π *from alphabet* $A = \{1, \cdots, k_1\}$ *to a smaller alphabet* $\{1, \cdots, k_2\}$*. This can be extended to a map from subshift of finite type* Σ_1 *over* $\{1, \cdots, k_1\}$ *to subshift* Σ_2 *over* $\{1, \cdots, k_2\}$ *by defining* $\Pi((x_i)_{i=0}^{\infty}) := (\Pi(x_i))_{i=0}^{\infty}$*. We call* Π *a one block factor map.*

Similarly, given a map $\Pi : \{1, \cdots, k_1\}^n \to \{1, \cdots, k_2\}$, we call the corresponding map $\Pi : \Sigma_1 \to \Sigma_2$ an n-block factor map. Any continuous factor can be represented as an n-block factor map for some natural number n, see [BP11] for a proof of this fact along with a detailed introduction to factor maps in symbolic dynamics. Then by recoding words of length n as letters in a new alphabet, we can reduce our problem to the study of 1-block factor maps. The images of Markov shifts under one block factor maps are also referred to as fuzzy, lumped or amalgamated Markov chains.

Σ_2 is not necessarily a subshift of finite type, it will be a subspace of $\{1, \cdots, k_2\}^{\mathbb{N}}$ but it need not be the case that the set of admissible sequences can be defined by a finite number of forbidden words.

It was shown by Chazottes and Ugalde in [CU03] that, in general, the image of a Markov measure on a subshift of finite type need not have a potential defined at all points, and hence need not be a Gibbs measure. We give a further simple example in Section 4. This motivates the following conditions.

The first condition is a mixing condition on fibres $\Pi^{-1}(\underline{z})$. Loosely, it says that if there exist sequences $\underline{x}, \underline{x}' \in \Pi^{-1}(\underline{z})$ then for any n and $m > N$ there exists a sequence $\underline{y} \in \Pi^{-1}(\underline{z})$ with $y_1 \cdots y_n = x_1 \cdots x_n$ and $y_{n+m} \cdots = x'_{n+m} \cdots$.

The second condition says that, in order to verify that a symbol x_n mapping onto z_n can be extended to a sequence $\underline{x}$ mapping onto $\underline{z}$, one needs only to check that x_n can be extended to a word $x_{n-N} \cdots x_{n+N}$ mapping on to $z_{n-N} \cdots z_{n+N}$.

The following definition allows us to state our hypotheses more formally.

Definition 6.0.2. *Given a set $B \subset \Sigma_1$ we let $\mathcal{A}_n(B)$ be the set of values of x_n for sequences $\underline{x}$ in B.*

Hypothesis 6.0.1. *We assume that for $\Pi : \Sigma_1 \to \Sigma_2$ there exists a natural number N such that for any $\underline{z} \in \Sigma_2$, $j \in \Sigma_1$,*

1. *Given $m, n \in \mathbb{N}$ with $m > N$, if there exists $\underline{x}$ in $\Pi^{-1}(\underline{z})$ with $x_{n+m} = j$ then $\mathcal{A}_n\{\underline{x} : x_{n+m} = j, \Pi(\underline{x}) = \underline{z}\} = \mathcal{A}_n\{\Pi(\underline{x}) = \underline{z}\}$.*

2. *$\mathcal{A}_n\{\underline{x} : \Pi(x_{n-N} \cdots x_{n+N}) = z_{n-N} \cdots z_{n+N}\} = \mathcal{A}_n\{\underline{x} : \Pi(\underline{x}) = \underline{z}\}$.*

Up to recoding of the alphabet A we can assume that $N = 1$, and thus the hypothesis implies that Σ_2 is a subshift of finite type, and that specifying some digit x_n in the set of sequences projecting to a word $\underline{z}$ only places restrictions on x_{n-1} and x_{n+1}.

These hypotheses are trivially satisfied for full shifts. Our conditions include cases not covered by [CU03], for example if $\underline{z}$ is periodic we do not require that N should be the period of $\underline{z}$. In section 6.5 we explain these conditions further and put them in the context of the technical conditions of Fan and Pollicott in [FP00], and explain how our conditions are weaker than those of [CU03].

Example 6.0.1. *Consider the shift space $\sigma : \Sigma_1 \to \Sigma_1$ associated to the transition matrix*

$$M = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 \end{pmatrix}$$

and the map Π from Σ_1 to the full shift on two symbols given by

$$\Pi(1) = a, \Pi(2) = \Pi(3) = \Pi(4) = b.$$

Then $\{2\} = \mathcal{A}_1\{\underline{x} : \Pi(x_1x_2) = ba\} \neq \mathcal{A}_1\{\underline{x} : \Pi(x_1) = b\} = \{2,3,4\}$. *We also see that putting* $x_1 = 3$ *makes it impossible that* x_2=*3, but places no restriction on possible values of* x_3x_4... *Thus the hypothesis fails on both counts with* $N = 0$, *but putting* $N = 1$ *it is satisfied.*

6.0.2 Results

The following is our main theorem.

Theorem 6.0.1. *Suppose that* $\Pi : \Sigma_1 \to \Sigma_2$ *satisfies hypothesis 6.0.1. If* μ *is a Gibbs measure on* Σ_1 *then* $\nu = \mu \circ \Pi^{-1}$ *is a Gibbs measure on* Σ_2. *If* ψ_1 *is a potential for* μ *and* ψ_2 *a potential for* ν *then*

1. *If* $var_n(\psi_1) < c_1\theta_1^{\sqrt{n}}$ *for some* $c_1 > 0$ *and* $\theta_1 \in (0,1)$, *then* $var_n(\psi_2) < c_2\theta_2^{\sqrt{n}}$ *for some* $c_2 > 0$ *and* $\theta_2 \in (0,1)$.

2. *If* $\sum_{n=0}^{\infty} n^k var_n(\psi_1) < \infty$ *for some* $k \geq 1$ *then* $\sum_{n=0}^{\infty} n^{k-1} var_n(\psi_2) < \infty$.

This generalises results in the papers [CU03] and [CU11] by Chazottes and Ugalde and [Ver11] by Verbitskiy. In [CU03] it was shown that the image of a Markov measure is a Gibbs measure with Hölder continuous potential provided the map Π satisfied two topological conditions. In [CU11] and [Ver11] it was assumed that Σ_1 is a full shift, and bounds on the regularity of ψ_2 were given in terms of the regularity of ψ_1. The results of [CU03] and [CU11] follow as corollaries to theorem 6.0.1. [Ver11] gives sharper bounds than theorem 6.0.1 on the regularity of ψ_2 in the case that Σ_1 is a full shift and ψ_1 is Hölder continuous.

It was conjectured by Chazottes and Ugalde in [CU11] that for any Gibbs measure μ on a subshift of finite type Σ_1, there would exist constants C_1, C_2 such that the image of μ under any one-block factor map would satisfy the inequality in Definition 2.2.3

almost everywhere. An example in section 6.4 shows this to be false. We believe that, while Hypothesis 6.0.1 could potentially be weakened, the principle that a choice of z_0 cannot affect potential choices of z_n for arbitrarily large n is crucial to the validity of the theorem, and thus that the theorem probably cannot be extended to more general factors of subshifts of finite type.

It was demonstrated in [Ver11] that Hölder continuity of the potential is preserved under factor maps in the case that Σ_1 is a full shift. The question of whether the same is true for subshifts of finite type remains open. The regularity conditions on ψ_1 in theorem 6.0.1 (i) are weaker than Hölder continuity, but we have been unable to show that requiring ψ_1 to be Hölder continuous improves the estimates on the regularity of ψ_2, except in the special case given in example 6.4.2.

In section 6.1 we will define a function ψ_2 and show that, should it be well defined, it is a potential for ν. Section 6.2 is dedicated to demonstrating that ψ_2 is well defined. In section 6.3 we prove that the variation behaves as in theorem 6.0.1. In section 6.4 we give an example and define a class of potentials for which Hölder continuity is preserved under Π.

6.1 Defining the Potential ψ_2

In this section we define a sequence of functions which are potentials for measures which approximate ν. The most technical part of the chapter involves demonstrating that the limit of this sequence of potentials converges and satisfies certain regularity conditions, this is deferred until the next section. Here we assume that the limit is well defined and show that it is indeed a potential for ν.

Definition 6.1.1. *We define our projected measure ν in terms of μ by*

$$\nu[z_0 \cdots z_n] = \sum_{x_0 \cdots x_n} \mu[x_0 \cdots x_n],$$

where the summation is over all words $x_0 \cdots x_n$ in Σ_1 projecting to $z_0 \cdots z_n$.

Since μ is a Gibbs measure, there exist by definition a potential ψ_1 and constants C_1, C_2 such that

$$\begin{aligned} C_1 \left(\sum_{\underline{x}=x_0 \cdots x_n} \exp(\psi_1^{n+1}(\underline{xw}(x_n))) \right) &\leq \nu[z_0 \cdots z_n] \\ &\leq C_2 \left(\sum_{\underline{x}=x_0 \cdots x_n} \exp(\psi_1^{n+1}(\underline{xw}(x_n))) \right) \end{aligned}$$

for any sequence $\underline{w}(x_n)$ in Σ_1 which can follow x_n. If we can find constants k_1, k_2 independent of n and a function ψ_2 such that

$$\begin{aligned} k_1 \left(\sum_{\underline{x}=x_0 \cdots x_n} \exp(\psi_1^{n+1}(\underline{xw}(x_n))) \right) &\leq \exp(\psi_2^{n+1}(\underline{z})) \\ &\leq k_2 \left(\sum_{\underline{x}=x_0 \cdots x_n} \exp(\psi_1^{n+1}(\underline{xw}(x_n))) \right) \end{aligned}$$

for all $\underline{t} \in \Sigma_2$ and some $\underline{w}(x_n)$ in Σ_1, then combining the previous two inequalities will give

$$\frac{C_1}{k_2} \leq \frac{\nu[z_0 \cdots z_n]}{\exp(\psi_2^{n+1}(\underline{z}))} \leq \frac{C_2}{k_1}.$$

This would make ψ_2 a potential for ν. Dividing by $\exp(\psi_2^n(\sigma(\underline{z})))$, we see that such a ψ_2 would also have to satisfy

$$\begin{aligned}\frac{k_1}{k_2}.\frac{\sum_{\underline{x}=x_0\cdots x_n}\exp(\psi_1^{n+1}(\underline{x}\underline{w}(x_n)))}{\sum_{\underline{x}'=x_1\cdots x_n}\exp(\psi_1^{n}(\underline{x}'\underline{w}(x_n)))} &\leq \exp(\psi_2(\underline{z})) \\ &\leq \frac{k_2}{k_1}.\frac{\sum_{\underline{x}=x_0\cdots x_n}\exp(\psi_1^{n+1}(\underline{x}\underline{w}(x_n)))}{\sum_{\underline{x}'=x_1\cdots x_n}\exp(\psi_1^{n}(\underline{x}'\underline{w}(x_n)))}\end{aligned}$$

Our aim is to use these equations, letting n tend to infinity, to define a potential ψ_2.

In [KP11], where we restricted our attention to factors of full shifts, we investigated the sequence

$$u_{\underline{w},n}(\underline{z}) := \frac{\sum_{\underline{x}=x_0\cdots x_n}\exp(\psi_1^{n+1}(\underline{x}\underline{w}(x_n)))}{\sum_{\underline{x}'=x_1\cdots x_n}\exp(\psi_1^{n}(\underline{x}'\underline{w}(x_n)))},$$

and showed that letting n tend to infinity led to a definition of ψ_2. Because we are dealing with the factors of subshifts rather than full shifts in this work the concatenation of sequences is more difficult, and in particular there is no simple expression for $u_{\underline{w},n+1}(\underline{z})$ as a function of terms $u_{\underline{w}',n}(\underline{z})$. This motivates the following refinement of the definition.

Definition 6.1.2. *For $n \in \mathbb{N}$, $j \in A$ and $\underline{w} = \underline{w}(j)$ a sequence in Σ_1 such that $j\underline{w}$ is admissible, we define $u_{j,\underline{w},n} : \Sigma_2 \to \mathbb{R}$:*

$$u_{j,\underline{w},n}(\underline{z}) = \frac{\sum_{\underline{x}=x_0\cdots x_{n-1}j}\exp(\psi_1^{n+1}(\underline{x}\underline{w}))}{\sum_{\underline{x}'=x_1\cdots x_{n-1}j}\exp(\psi_1^{n}(\underline{x}'\underline{w}))}$$

where

1. *$\psi_1^n = \sum_{i=0}^{n} \psi \circ \sigma^i$ and $\psi_1^{n+1} = \sum_{i=0}^{n+1} \psi_1 \circ \sigma^i$;*

2. *each summation is over finite strings from Σ_1 for which x_i projects to z_i, for $i \in \{0, \dots, n\}$;*

3. *$\underline{x}j\underline{w} \in \Sigma_1$ denotes the concatenation of words to give the sequence*

$(x_0 \cdots x_{n-1} j w_0 w_1 \cdots)$.

We note that there is an explicit dependence on the choices of j and $\underline{w}$ here, and that technical reasons make the introduction of j and $\underline{w}$ necessary. However, we will show that the limit $u := \lim_{n\to\infty} u_{j,\underline{w},n}$ is a well defined function depending only on $\underline{z} \in \Sigma_2$ and that $\psi_2 := \log u$ is a potential for ν.

Given a finite word $x_1 \cdots x_n$, $\psi_1(x_1 \cdots x_n)$ is not defined, and so we have introduced $\underline{w}$ in order that we can consider $\psi_1(x_1 \cdots x_n \underline{w})$. It will often be necessary to consider tail sequences after various different words, and since Σ_1 may not be a full shift we may have to consider different tail sequences $\underline{w}$ for different choices of x_n. We define $\underline{w} : \{1, \cdots, k_1\} \to \Sigma_1$, this assigns a tail sequence $\underline{w}(x_n)$ to follow $x_1 \cdots x_n$ for each possible value of x_n. As shorthand we write $x_1 \cdots x_n \underline{w}$ for the concatenation $x_1 \cdots x_n \underline{w}(x_n)$.

Proposition 6.1.1. *$u(\underline{z}) := \lim_{n\to\infty} u_{j,\underline{w},n}(\underline{z})$ is well defined and independent of j and $\underline{w}$.*

We return to the proof in the next section. We work towards showing that, should $u(\underline{z})$ be well defined, $\log(u)$ is a potential for ν.

Lemma 6.1.1. *There is a constant C depending only on ψ_1 such that for each $n, s \in \mathbb{N}$ with $s > n$ and for each $\underline{z} \in \Sigma_2$,*

$$\frac{1}{C} \le \frac{\sum_{\underline{x}=x_0\cdots x_n} \exp(\psi_1^{n+1}(\underline{xw})) \sum_{\overline{x}=x_{n+1}\cdots x_s} \exp(\psi_1^{s-n}(\overline{x}\underline{w}))}{\sum_{\underline{x}=x_0\cdots x_s} \exp(\psi_1^{s+1}(\underline{xw}))} \le C$$

Proof. We split the expression into two fractions. Hypothesis 6.0.1 gives us that a choice x_n cannot affect choices of x_{n+2}, since for any x_n there exists an x_{n+1} projecting to z_{n+1} such that $x_n x_{n+1} x_{n+2}$ is an admissible word. Given x_n, we shall use the notation $\sum_{\overline{x}=x_{n+1}\cdots x_s}^{x_n}$ to mean the summation over all words $\overline{x} = x_{n+1} \cdots x_s$

in Σ_1 projecting to $z_{n+1}\cdots z_s$ with the added restriction that $x_n x_{n+1}$ must be an admissible word in Σ_1. Then given x_n we have

$$\begin{aligned} 1 &\leq \frac{\sum_{\overline{x}=x_{n+1}\cdots x_s} \exp(\psi_1^{s-n}(\overline{x}\underline{w}))}{\sum_{\overline{x}=x_{n+1}\cdots x_s}^{x_n} \exp(\psi_1^{s-n}(\overline{x}\underline{w}))} \\ &= \frac{\sum_{\hat{x}=x_{n+2}\cdots x_s} \sum_{\overline{x}=x_{n+1}}^{x_{n+2}} \exp(\psi_1^{s-n}(\overline{x}\hat{x}\underline{w}))}{\sum_{\hat{x}=x_{n+2}\cdots x_s} \sum_{\overline{x}=x_{n+1}}^{x_n,x_{n+2}} \exp(\psi_1^{s-n}(\overline{x}\hat{x}\underline{w}))} \end{aligned}$$

Here all we have done is to split the summation into two pieces. But this can be further rewritten

$$\begin{aligned} &= \frac{\sum_{\hat{x}=x_{n+2}\cdots x_s} \exp(\psi_1^{s-n-1}(\hat{x}\underline{w})) \sum_{\overline{x}=x_{n+1}}^{x_{n+2}} \exp(\psi_1(\overline{x}\hat{x}\underline{w}))}{\sum_{\hat{x}=x_{n+2}\cdots x_s} \exp(\psi_1^{s-n-1}(\hat{x}\underline{w})) \sum_{\overline{x}=x_{n+1}}^{x_n,x_{n+2}} \exp(\psi_1(\overline{x}\hat{x}\underline{w}))} \\ &\leq \exp(var_0(\psi_1)).|A|. \end{aligned}$$

The final line follows because, given any $x_n, \hat{x}$, hypothesis 6.0.1 guarantees the existence of at least one choice of $\overline{x}$ linking x_n to x_{n+2} and there can be at most $|A|$, thus the ratio of the number of terms can be at most $|A|$, and for any $\overline{x} = x_{n+1}$,

$$\frac{\exp(\psi_1(\overline{x'}\hat{x}\underline{w}))}{\exp(\psi_1(\overline{x}\hat{x}\underline{w}))} \leq \exp(var_0(\psi_1)).$$

We note that from the definition of a Gibbs measure we have that, for any choices of $\underline{x} = x_0 \cdots x_n$, $\underline{w}$ and $\underline{w}'$,

$$C_1 \exp(\psi_1^{n+1}(\underline{x}\underline{w})) \leq \mu[x_0 \cdots x_n] \leq C_2 \exp(\psi_1^{n+1}(\underline{x}\underline{w}')),$$

which gives in particular that for any $\overline{x}$,

$$\frac{\sum_{\underline{x}=x_0\cdots x_n} \exp(\psi_1^{n+1}(\underline{x}\underline{w}))}{\sum_{\underline{x}=x_0\cdots x_n} \exp(\psi_1^{n+1}(\underline{x}\overline{x}\underline{w}))} \leq \frac{C_2}{C_1}.$$

Returning to our original expression, we have that

$$\begin{aligned}
&\frac{\sum_{\underline{x}=x_0\cdots x_n}\exp(\psi_1^{n+1}(\underline{xw}))\sum_{\overline{x}=x_{n+1}\cdots x_s}\exp(\psi_1^{s-n}(\overline{x}\underline{w}))}{\sum_{\underline{x}=x_0\cdots x_s}\exp(\psi_1^{s+1}(\underline{xw}))}\\
=\;&\frac{\sum_{\underline{x}=x_0\cdots x_n}\exp(\psi_1^{n+1}(\underline{xw}))}{\sum_{\underline{x}=x_0\cdots x_n}\exp(\psi_1^{n+1}(\underline{x}\overline{x}\underline{w}))}\frac{\sum_{\overline{x}=x_{n+1}\cdots x_s}\exp(\psi_1^{s-n}(\overline{x}\underline{w}))}{\sum_{\overline{x}=x_{n+1}\cdots x_s}^{x_n}\exp(\psi_1^{s-n}(\overline{x}\underline{w}))}\\
\leq\;&\frac{C_2}{C_1}|A|\exp(var_0(\psi_1)),
\end{aligned}$$

and so putting $C=|A|\exp(var_0(\psi_1))\frac{C_2}{C_1}$ we are done. □

We define $\psi_2 := \log u$. The following lemma gives that ψ_2 is a potential for ν.

Lemma 6.1.2. *If u is well defined then ψ_2 is a potential for ν.*

Proof. Fix $n \geq 1$. We can write

$\psi_2^{n+1}(\underline{z}) = \lim_{m\to+\infty}\log u_{j,\underline{w},m}(\underline{z}) + \cdots + \lim_{m\to+\infty}\log u_{j,\underline{w},m}(\sigma^n\underline{z})$, giving

$$\psi_2^{n+1}(\underline{w}) = \lim_{m\to+\infty}\log\left(\frac{\sum_{\underline{x}=x_0\cdots x_{m-1}j}\exp(\psi_1^{m+1}(\underline{xw}))}{\sum_{\overline{x}=x_{n+1}\cdots x_{m-1}j}\exp(\psi_1^{m+1}(\overline{x}\underline{w}))}\right).$$

Moreover, by lemma 6.1.1

$$\sum_{\underline{x}=x_0\cdots x_{m-1}j}\exp(\psi_1^{m+1}(\underline{xw})) \leq C\sum_{\underline{x}'=x_0\cdots x_n}\exp(\psi_1^{n+1}(\underline{x}'\underline{w}))\sum_{\overline{x}=x_{n+1}\cdots x_{m-1}j}\exp(\psi_1^{m-n}(\overline{x}\underline{w}))$$

so we can bound

$$\begin{aligned}
\frac{1}{C}\sum_{\underline{x}'=x_0\cdots x_n}\exp(\psi_1^{n+1}(\underline{x}'\underline{w})) &\leq \underbrace{\frac{\sum_{\underline{x}=x_0\cdots x_{m-1}j}\exp(\psi_1^{m+1}(\underline{xw}))}{\sum_{\overline{x}=x_{n+1}\cdots x_{m-1}j}\exp(\psi_1^{m-n}(\overline{x}\underline{w}))}}_{=\exp(\sum_{i=0}^{n}\log u_{j,\underline{w},(m-i)}(\sigma^i\underline{z}))}\\
&\leq C\sum_{\underline{x}'=x_0\cdots x_n}\exp(\psi_1^{n+1}(\underline{x}'\underline{w})).
\end{aligned}$$

Since μ is a Gibbs measure for ψ_1, there exist constants $C_1, C_2 > 0$ such that for any $\pi(\underline{x}) = \underline{z}$ and $n \geq 1$:

$$C_1 \exp(\psi_1^{n+1}(\underline{x})) \leq \mu[x_0 \cdots x_n] \leq C_2 \exp(\psi_1^{n+1}(\underline{x})).$$

Summing over strings $x_0 \cdots x_n$ corresponding to $\pi(\underline{x}) = \underline{z}$ gives

$$C_1 \leq \frac{\nu[z_0 \cdots z_n]}{\sum_{\underline{x}'=x_0\cdots x_n} \exp(\psi_1^{n+1}(\underline{x}'\underline{w}))} \leq C_2,$$

and hence

$$\frac{C_1}{C} \leq \frac{\nu[z_0 \cdots z_n]}{\exp(\psi_2^{n+1}(\underline{z}))} \leq C_2 C.$$

Therefore ν is a Gibbs measure for ψ_2. □

6.2 Proof that ψ_2 is Well Defined

In this section we will demonstrate that ψ_2 is well defined and prove properties of the variation of ψ_2. While the details are quite technical, the underlying principles are straightforward. We explain them in terms of the following three definitions, which help us quantify how accurate an approximation the function $u_{j,\underline{w},n}(\underline{z})$ is to the limit $u(\underline{z})$.

Definition 6.2.1. $\Lambda_n(\underline{z}) := [\min_{j,\underline{w}} u_{j,\underline{w},n}(\underline{z}), \max_{j',\underline{w}'} u_{j',\underline{w}',n}(\underline{z})]$

We will show that, for any $\underline{z} \in \Sigma_2$, $\Lambda_n(\underline{z})$ is a nested sequence. Earlier we defined $u(\underline{z})$ to be the limit as n tends to infinity of the sequence $u_{j,\underline{w},n}(\underline{z})$, claiming that the limit is independent of j and $\underline{w}$. Once we have shown that $\Lambda_n(\underline{z})$ is nested, $u(\underline{z})$ being well defined will follow from the diameter of the intervals $\Lambda_n(\underline{z})$ tending to zero. For technical reasons it is easier to study $\lambda_n(\underline{z})$, defined as follows.

Definition 6.2.2. $\lambda_n(\underline{z}) := \sup \left\{ \frac{u_{j,\underline{w},n}(\underline{z})}{u_{j',\underline{w}',n}(\underline{z})} : \underline{w}, \underline{w}' \in \Sigma, j, j' \in A \right\}.$

If we can show that $\lambda_n(\underline{z})$ converges to 1 for any point $\underline{z}$, this will imply that the diameter of $\Lambda_n(\underline{z})$ tends to zero, giving the existence of $u(\underline{z})$. In fact we prove that the following quantity tends to one, giving the existence of $u(\underline{z})$ for all $z \in \Sigma_2$.

Definition 6.2.3. $\lambda_n := \sup_{\underline{z}\in\Sigma_2} \lambda_n(\underline{z})$

In order to prove properties of the regularity of ψ_2, we note from the definition that $u_{j,\underline{w},n}(\underline{z})$ actually depends only on $z_0 \cdots z_n$. So $\Lambda_n(\underline{z})$ only depends on $z_0 \cdots z_n$, and if $\underline{z}$ and $\underline{z}'$ agree to $n+1$ places then $\Lambda_n(\underline{z}) = \Lambda_n(\underline{z}')$. Then the nestedness of Λ_n ensures that $u(\underline{z})$ and $u(\underline{z}')$ are both contained in the interval $\Lambda_n(\underline{z})$. Hence $|\psi_2(\underline{z}) - \psi_2(\underline{z}')| \leq \log(\lambda_n(\underline{z}))$ and so $var_{n+1}(\psi_2) \leq \log(\lambda_n)$.

This section is dedicated to proving that $\lambda_n \to 1$, and hence that ψ_2 is well defined. A key lemma in the proof, lemma 6.2.3, will be used later to obtain rates of convergence of λ_n which give the variation of ψ_2.

Lemma 6.2.1. *The sequence of intervals $\Lambda_n(\underline{z})$ is nested.*

Proof. From the definitions of $u_{j,\underline{w},n}(\underline{z})$ and $u_{j,\underline{w},n+1}(\underline{z})$ we observe that

$$\begin{aligned} u_{j,\underline{w},n+1}(\underline{z}) &= \frac{\sum_{x_n}^{j} \text{numerator}(u_{x_n,j\underline{w},n}(\underline{z})).\exp(\psi_1(j\underline{w}))}{\sum_{x_n}^{j} \text{denominator}(u_{x_n,j\underline{w},n}(\underline{z})).\exp(\psi_1(j\underline{w}))} \\ &\leq \max_{x_n,\underline{w}'} u_{x_n,\underline{w}',n}(\underline{z}) \end{aligned}$$

where the second line follows because $\dfrac{\sum_{k=1}^{n} a_k}{\sum_{k=1}^{n} b_k} \leq \max_{k=1,\cdots,n}\left\{\dfrac{a_k}{b_k}\right\}$.

The same observation works for the minimum. ☐

To demonstrate that $\lambda_n \to 1$ we define a probability vector which allows us to express the function $u_{j,\underline{w},s}$ in terms of functions $u_{j',\underline{w}',n}$ for $n < s$.

Definition 6.2.4. *Let $0 < n < s$, $j, x_n \in A$ be fixed. Let $\overline{x}$ be some choice of word*

$x_{n+1}\cdots x_s$ *compatible with* x_n. *We then define*

$$P^{(s+2,n)}(x_n,\overline{x},j,\underline{w}) = \frac{\sum_{\underline{x}=x_1\cdots x_n}^{x_n}\sum_{\hat{x}=x_{s+1}j}^{x_s}\exp(\psi_1^{s+2}(\underline{x}\overline{x}\hat{x}\underline{w}))}{\sum_{\underline{x}'=x_1\cdots x_{s+1}j}\exp(\psi_1^{s+2}(\underline{x}'\underline{w}))}.$$

By construction, this is a probability vector over choices of x_n and $\overline{x}$, it can be seen as the probability of that a word $y_1\cdots y_{s+1}j \in \Pi^{-1}(z_1\cdots z_{s+2})$ has $y_n\cdots y_s = x_n\overline{x}$ for some notion of probability arising from ψ_1. In the limit as s tends to infinity this probability is in terms of measure the μ. Note that by Hypothesis 6.0.1 there is always some choice of $\hat{x}$ linking x_s to j and so $P^{(s+2,n)}(\overline{x},j,\underline{w})$ is never zero for $\overline{x}$ compatible with x_n.

Definition 6.2.5. *Given* x_n, $\overline{x} = x_{n+1}\cdots x_s$, $\underline{w}$ *and* j *we let* $\underline{w}^{max}$ *be the concatenation* $\hat{x}\underline{w}$ *for the value of* x_{s+1} *which maximises* $u_{x_n,\overline{x}\hat{x}\underline{w},n}(\underline{z})$, *where* $\hat{x} = x_{s+1}j$. *We let* $\underline{w}^{min}$ *be the string* $\hat{x}\underline{w}$ *which minimises* $u_{x_n,\overline{x}\hat{x}\underline{w},n}(\underline{z})$.

In the case that ψ_1 is Markov, it is easy to express $u_{j,\underline{w},n+2}$ using terms $u_{j',\underline{w}',n}$, and from this one can show that $\Lambda_n(\underline{z})$ contracts. In the non-Markov case it is more difficult, but the following inequality is sufficient to show that $\lambda_n \to 1$.

Lemma 6.2.2.

$$u_{j,\underline{w},s+2}(\underline{z}) \le \sum_{x_n}\sum_{\overline{x}=x_{n+1}\cdots x_s}^{x_n} u_{x_n,\overline{x}\underline{w}^{max},n}(\underline{z})P^{(s+2,n)}(x_n,\overline{x},j,\underline{w}).$$

Proof. By definition, the *numerator* of $u_{j,\underline{w},s+2}(\underline{z})$, which is $\sum_{\underline{x}=x_0\cdots x_{s+1}j}\exp(\psi_1^{s+3}(\underline{x}\underline{w}))$, can be written

$$\sum_{x_n}\sum_{\underline{x}=x_0\cdots x_n}^{x_n}\sum_{\overline{x}=x_{n+1}\cdots x_s}^{x_n}\sum_{\hat{x}=x_{s+1}j}^{x_s}\exp(\psi_1^{n+1}(\underline{x}\overline{x}\hat{x}\underline{w}))\exp(\psi_1^{s+2-n}(\overline{x}\hat{x}\underline{w})).$$

We have used $\psi_1^{s+3}(\underline{x}\overline{x}\hat{x}\underline{w}) = \psi_1^{n+1}(\underline{x}\overline{x}\hat{x}\underline{w}) + \psi_1^{s+2-n}(\overline{x}\hat{x}\underline{w})$. We can further rewrite this as

$$\sum_{x_n} \sum_{\overline{x}=x_{n+1}\cdots x_s}^{x_n} \sum_{\hat{x}=x_{s+1}j}^{x_s} \underbrace{\left(\frac{\sum_{\underline{x}=x_0\cdots x_n}^{x_n} \exp(\psi_1^{n+1}(\underline{x}\overline{x}\hat{x}\underline{w}))}{\sum_{\underline{x}'=x_1\cdots x_n}^{x_n} \exp(\psi_1^{n}(\underline{x}'\overline{x}\hat{x}\underline{w}))} \right)}_{u_{x_n,\overline{x}\hat{x}\underline{w},n}(\underline{z})}$$

$$\times \underbrace{\left(\sum_{\underline{x}'=x_1\cdots x_n}^{x_n} \exp(\psi_1^{n}(\underline{x}'\overline{x}\hat{x}\underline{w})) \right) \exp(\psi_1^{s+2-n}(\overline{x}\hat{x}\underline{w})).}_{\sum_{\underline{x}'=x_1\cdots x_n} \exp(\psi_1^{s+2}(\underline{x}'\overline{x}\hat{x}\underline{w}))}$$

Now we wish to move the summation over $\hat{x}$ to the second bracket, but we note that the first bracket is not independent of $\hat{x}$. However using $\underline{w}^{\max}$ as defined above we can get an inequality.

$$\leq \sum_{x_n} \sum_{\overline{x}=x_{n+1}\cdots x_s}^{x_n} \underbrace{\left(\frac{\sum_{\underline{x}=x_0\cdots x_n}^{x_n} \exp(\psi_1^{n+1}(\underline{x}\overline{x}\underline{w}^{\max}))}{\sum_{\underline{x}'=x_1\cdots x_n}^{x_n} \exp(\psi_1^{n}(\underline{x}'\overline{x}\underline{w}^{\max}))} \right)}_{u_{x_n,\overline{x}\underline{w}^{\max},n}(\underline{z})} \underbrace{\left(\sum_{\hat{x}=x_{s+1}j}^{x_s} \sum_{\underline{x}'=x_1\cdots x_n}^{x_n} \exp(\psi_1^{s+2}(\underline{x}'\overline{x}\hat{x}\underline{w})) \right)}_{numerator(P^{s+2,n}(x_n,\overline{x},j,\underline{w}))}.$$

So by dividing by the denominator of $u_{j,\underline{w},s+2}(\underline{z})$, which equals the denominator of $P^{s+2,n}(x_n, \overline{x}, j, \underline{w})$, we see that

$$u_{j,\underline{w},s+2}(\underline{z}) \leq \sum_{x_n} \sum_{\overline{x}=x_{n+1}\cdots x_s}^{x_n} u_{x_n,\overline{x}\underline{w}^{\max},n}(\underline{z}).P^{s+2,n}(x_n, \overline{x}, j, \underline{w}).$$

We note that the only dependence on j in the above is in $P^{s+2,n}(x_n, \overline{x}, j, \underline{w})$, in particular, all the summations are over sets which are independent of j. □

Corollary 6.2.1.

$$\frac{u_{j,\underline{w},s+2}(\underline{z})}{u_{j',\underline{w}',s+2}(\underline{z})} \leq \frac{\sum_{x_n} \sum_{\overline{x}=x_{n+1}\cdots x_s}^{x_n} u_{x_n,\overline{x}\underline{w}^{max},n}(\underline{z}) P^{(s+2,n)}(x_n, \overline{x}, j, \underline{w})}{\sum_{x_n} \sum_{\overline{x}=x_{n+1}\cdots x_s}^{x_n} u_{x_n,\overline{x}\underline{w}'^{min},n}(\underline{z}) P^{(s+2,n)}(x_n, \overline{x}, j', \underline{w}'))}.$$

This follows from using $\underline{w}^{\min}$ in the previous lemma for the denominator.

We are finally able to state a lemma giving that λ_n tends to zero at a certain rate. This is crucial in showing that ψ_2 is well defined and for proving properties of the variation of ψ_2.

Lemma 6.2.3. *Suppose that for all $j, j' \in \Pi^{-1}(z_s), \underline{w}, \underline{v} \in \Sigma_1,\ s > n+1$:*

1. $\dfrac{u_{j,\underline{w},s+2}(\underline{z})}{u_{j',\underline{v},s+2}(\underline{z})} \leq \dfrac{\sum_{x_n}\sum_{\overline{x}=x_{n+1}\cdots x_s}^{x_n} u_{x_n,\overline{x}\underline{w}^{\max},n}(\underline{z}).P^{s+2,n}(x_n,\overline{x},j,\underline{w})}{\sum_{x_n}\sum_{\overline{x}=x_{n+1}\cdots x_s}^{x_n} u_{x_n,\overline{x}\underline{v}^{\min},n}(\underline{z}).P^{s+2,n}(x_n,\overline{x},j',\underline{v})}$;
2. *there exists* $c \in (0,1)$ *with* $c < \dfrac{P^{s+2,n}(x_n,\overline{x},j,\underline{w})}{P^{s+2,n}(x_n,\overline{x},j',\underline{v})}$ $\forall x_n, \overline{x}, j, k, \underline{w}, \underline{v}, s > n$; *and*
3. $\dfrac{u_{x_n,\overline{x}\underline{w}^{\max},n}(\underline{z})}{u_{x_n,\overline{x}\underline{v}^{\min},n}(\underline{z})} \leq \exp(2\sum_{k=s-n}^{s} var_k(\psi_1))$.

Then

$$\frac{u_{j,\underline{w},s+2}(\underline{z})}{u_{j',\underline{v},s+2}(\underline{z})} \leq c.\exp\left(2\sum_{k=s-n}^{s} var_k(\psi_1)\right) + (1-c).\max_{j,j',\underline{w},\underline{v}}\left(\frac{u_{j,\underline{w},n}(\underline{z})}{u_{j',\underline{v},n}(\underline{z})}\right).$$

We have already shown in corollary 6.2.1 that the first condition is satisfied, the proof that conditions 2 and 3 are satisfied is at the end of this section.

Proof. To simplify notation, we fix $\underline{z}$ and rewrite $\sum_{x_n}\sum_{\overline{x}=x_{n+1}\cdots x_s}^{x_n}$ as $\sum_{i\in I}$, letting i represent $x_n\overline{x}$ and I represent the finite set of possible choices of $x_n\overline{x}$. We are going to represent the above summations over $i \in I$ as the dot product of vectors with entries corresponding to symbols $i \in I$. Recall that we defined

$$P^{(s+2,n)}(x_n,\overline{x},j,\underline{w}) = \frac{\sum_{\underline{x}=x_1\cdots x_n}^{x_n}\sum_{\hat{x}=x_{s+1}j}^{x_s}\exp(\psi_1^{s+2}(\underline{x}\overline{x}\hat{x}\underline{w}))}{\sum_{\underline{x}'=x_1\cdots x_{s+1}j}\exp(\psi_1^{s+2}(\underline{x}'\underline{w}))},$$

which can be seen as the probability of picking the particular choice of $x_n\overline{x}$, or with our new notation, the probability of picking $i \in I$, given j and $\underline{w}$. Writing

$P^{(s+2,n)}(x_n, \overline{x}, j, \underline{w}) = P^{(s+2,n)}(i, j, \underline{w})$ for i corresponding to the correct choice of $x_n\overline{x}$, we construct the probability vector P_1 indexed by $i \in I$, by

$$P_1(i) := P^{(s+2,n)}(i, j, \underline{w}).$$

We let A be defined by

$$A(i) := a_i = (u_{x_n,\overline{x}\underline{w}^{max},n})$$

for $x_n, \overline{x}$ corresponding to $i \in I$.

Then we can rewrite the summation $u_{j,\underline{w},s+2}(\underline{z}) = P_1 \cdot A$.

We define P_2 and B by replacing j with j' and $\underline{w}$ with $\underline{w}'$ in the definitions of P_1 and A. The technical conditions of Hypothesis 6.0.1 ensure that replacing j with j' does not affect the possible choices of $x_n\overline{x}$, and so P_1, P_2, A and B are probability vectors indexed by the same set I. Hypothesis 1 of lemma 6.2.3 now becomes

$$\frac{u_{j,\underline{w},s+2}(\underline{z})}{u_{j',\underline{w}',s+2}(\underline{z})} \leq \frac{P_1 \cdot A}{P_2 \cdot B},$$

where, under Hypotheses 2 and 3 of lemma 6.2.3, there is a universal constant c such that $c < \frac{P_1(i)}{P_2(i)}$, and $\frac{A(i)}{B(i)} \leq \exp(2\sum_{k=s-n}^{s} var_k(\psi_1))$

We assume that $\max_{j,j',\underline{w},\underline{v}} \left(\frac{u_{j,\underline{w},n}(\underline{z})}{u_{j',\underline{v},n}(\underline{z})}\right) > \exp\left(2\sum_{k=s-n}^{s} var_k(\psi_1)\right)$, otherwise

$$\begin{aligned}
\frac{u_{j,\underline{w},s+2}(\underline{z})}{u_{j',\underline{v},s+2}(\underline{z})} &\leq (c + (1-c)) \max_{j,j',\underline{w},\underline{v}} \left(\frac{u_{j,\underline{w},n}(\underline{z})}{u_{j',\underline{v},n}(\underline{z})}\right) \\
&\leq c \exp\left(2\sum_{k=s-n}^{s} var_k(\psi_1)\right) + (1-c) \max_{j,j',\underline{w},\underline{v}} \left(\frac{u_{j,\underline{w},n}(\underline{z})}{u_{j',\underline{v},n}(\underline{z})}\right)
\end{aligned}$$

as required.

Now we use c to write

$$\frac{u_{j,\underline{w},s+2}(\underline{z})}{u_{j',\underline{v},s+2}(\underline{z})} \leq \frac{(c.P_1 \cdot A) + ((1-c).P_1 \cdot A)}{(c.P_1 \cdot B) + ((P_2 - cP_1) \cdot B)}$$

noting that $P_2 - cP_1 \geq 0$. We will use $\mathbb{1}$ to represent a vector of all 1s of length $|I|$. Now $A \leq \exp\left(2\sum_{k=s-n}^{s} var_k(\psi_1)\right) B$, so

$$\begin{aligned}
\frac{u_{j,\underline{w},s+2}(\underline{z})}{u_{j',\underline{v},s+2}(\underline{z})} &\leq \frac{\left(c.\exp\left(2\sum_{k=s-n}^{s} var_k(\psi_1)\right) P_1 \cdot B\right) + ((1-c).P_1 \cdot A)}{(c.P_1 \cdot B) + ((P_2 - cP_1) \cdot B)} \\
&\leq \frac{\left(c.\exp\left(2\sum_{k=s-n}^{s} var_k(\psi_1)\right) P_1 \cdot B\right) + ((1-c).P_1 \cdot \mathbb{1}.\max_i(a_i))}{(c.P_1 \cdot B) + ((P_2 - cP_1) \cdot \mathbb{1}\min_i(b_i))} \\
&\leq \frac{\left(c.\exp\left(2\sum_{k=s-n}^{s} var_k(\psi_1)\right) P_1 \cdot \mathbb{1}\min_i(b_i)\right) + ((1-c).P_1 \cdot \mathbb{1}.\max_i(a_i))}{(c.P_1 \cdot \mathbb{1}\min_i(b_i)) + ((P_2 - cP_1) \cdot \mathbb{1}\min_i(b_i))}
\end{aligned}$$

The justification for the last step is that we assumed $\dfrac{\max_i(a_i)}{\min_i(b_i)} > \exp\left(2\sum_{k=s-n}^{s} var_k(\psi_1)\right)$, and so shrinking terms on the top and bottom which are similar leads to more weight being given to those which are dissimilar. This is just the statement that if $\frac{a_1}{b_1} < \frac{a_2}{b_2}$ then

$$\frac{\alpha a_1 + a_2}{\alpha b_1 + b_2} > \frac{a_1 + a_2}{b_1 + b_2}$$

for any $\alpha \in (0,1)$.

Of course, $P_1 \cdot \mathbb{1} = P_2 \cdot \mathbb{1} = 1$, since P_1 and P_2 are probability vectors, so we can divide by $\min_i(b_i)$ to get

$$\begin{aligned}
\frac{u_{j,\underline{w},s+2}(\underline{z})}{u_{j',\underline{v},s+2}(\underline{z})} &\leq \frac{c.\exp\left(2\sum_{k=s-n}^{s} var_k(\psi_1)\right) + (1-c).\frac{\max_i(a_i)}{\min_i(b_i)}}{c + (1-c)} \\
&= c.\exp\left(2\sum_{k=s-n}^{s} var_k(\psi_1)\right) + (1-c)\max_{j,j',w,v}\left(\frac{u_{j,w,n}(\underline{z})}{u_{j',v,n}(\underline{z})}\right).
\end{aligned}$$

□

The previous lemma was useful to us as it allows us to prove the following corollary.

Corollary 6.2.2. *Given $s, n \in \mathbb{N}$ with $s > n$ we have*

$$\lambda_{s+2} \leq c. \exp\left(2 \sum_{k=s-n}^{s} var_k(\psi_1)\right) + (1-c)\lambda_n.$$

Proof. Using the conclusion of lemma 6.2.3 and taking the supremum over all choices of $j, \underline{w}, j'$ and $\underline{w}'$, we get

$$\lambda_{s+2}(\underline{z}) \leq c. \exp\left(2 \sum_{k=s-n}^{s} var_k(\psi_1)\right) + (1-c)\lambda_n(\underline{z}).$$

Then since each λ_n is finite we can take suprema over $\underline{z}$ and the corollary is proved.

□

In particular, for any n and any $\epsilon > 0$ we can choose s sufficiently large so that $\exp\left(2 \sum_{k=s-n}^{s} var_k(\psi_1)\right) < \frac{\lambda_n(\underline{z})}{2}$. Then

$$\lambda_{s+2}(\underline{z}) \leq (1 - \frac{c}{2})\lambda_n(\underline{z}),$$

and iterating we see that λ_n tends to one and so $\psi_2 := \log u$ is well defined and continuous.

This proves the first part of theorem 6.0.1, that if μ is a Gibbs measure and Π is a map satisfying Hypothesis 6.0.1, then the image measure ν is a Gibbs measure, under the assumption that the three conditions of lemma 6.2.3 are satisfied. Condition 1 was proved in corollary 6.2.1 and we prove conditions 2 and 3 now.

Lemma 6.2.4. *There is a constant $c > 0$ such that for all $s, n \in \mathbb{N}$ with $s > n$, $\underline{z} \in \Sigma_2$, and for all choices of $x_n, \overline{x}, j, j', \underline{w}$ and $\underline{w}'$,*

$$c \leq \frac{P^{(s+2,n)}(x_n, \overline{x}, j, \underline{w})}{P^{(s+2,n)}(x_n, \overline{x}, j', \underline{w}')}.$$

Proof. We can write

$$P^{(s+2,n)}(x_n, \overline{x}, j, \underline{w}) = \frac{\sum_{\underline{x}=x_1\cdots x_n}^{x_n} \sum_{\hat{x}=x_{s+1}j}^{x_s} \exp(\psi_1^{s+2}(\underline{x}\overline{x}\hat{x}\underline{w}))}{\sum_{\underline{x}'=x_1'\cdots x_s'} \sum_{\hat{x}=x_{s+1}j}^{x_s'} \exp(\psi_1^{s+2}(\underline{x}'\hat{x}\underline{w}))}.$$

We consider first the numerator. Given a choice of $\overline{x}$, changing j can only affect possible choices of x_{s+1}. There will always be at least one choice of x_{s+1} linking x_s to j by Hypothesis 6.0.1, and there can be at most $|A|$. So the number of terms in the summation for different choices of j can differ by a factor of at most $|A|$. Furthermore, given $\underline{x} = x_0 \cdots x_n, \overline{x} = x_{n+1} \cdots x_s, \hat{x}, \hat{x}', j, j', \underline{w}, \underline{w}'$, we have by lemma 6.1.1 that

$$\frac{\exp(\psi_1^{s+2}(\underline{x}\overline{x}\hat{x}\underline{w}))}{\exp(\psi_1^{s+2}(\underline{x}\overline{x}\hat{x}'\underline{w}'))} = \frac{\exp(\psi_1^{s}(\underline{x}\overline{x}\hat{x}\underline{w}))}{\exp(\psi_1^{s}(\underline{x}\overline{x}\hat{x}'\underline{w}'))} \frac{\exp(\psi_1^{2}(\hat{x}\underline{w}))}{\exp(\psi_1^{2}(\hat{x}'\underline{w}'))} \leq C.\exp(2var_0(\psi_1)).$$

Making identical calculations for the denominator we see that the lemma is proved with

$$c = \frac{1}{|A|^2 C^2 \exp(4var_0(\psi_1))}.$$

□

We now need only to confirm that the third condition of lemma 6.2.3 is satisfied.

Lemma 6.2.5. $\dfrac{u_{x_n,\overline{x}\underline{w}^{max},n}}{u_{x_n,\overline{x}\underline{w}^{min},n}} \leq \exp(2\sum_{k=s-n}^{s} var_k(\psi_1))$.

Proof. We recall that $\overline{x}$ was some choice of $x_{n+1} \cdots x_s$. Considering first the numerators, we see that

$$\frac{\text{numerator}(u_{x_n,\overline{x}\underline{w}^{max},n})}{\text{numerator}(u_{x_n,\overline{x}\underline{w}^{min},n})} = \frac{\sum_{\underline{x}=x_0\cdots x_n} \exp(\psi_1^{n+1}(\underline{x}\overline{x}\underline{w}^{max}))}{\sum_{\underline{x}=x_0\cdots x_n} \exp(\psi_1^{n+1}(\underline{x}\overline{x}\underline{w}^{min}))}.$$

Comparing termwise we see that $\sigma^k(\underline{x}\overline{x}\underline{w}^{min})$ and $\sigma^k(\underline{x}\overline{x}\underline{w}^{max})$ agree to $s-n+(n-k)$

places, and thus for any choice of $\underline{x}$,

$$\frac{\exp(\psi_1^{n+1}(\underline{x}\overline{x}\underline{w}^{max}))}{\exp\left(\psi_1^{n+1}(\underline{x}\overline{x}\underline{w}^{min})\right)} \leq \exp\left(\sum_{k=s-n}^{s} var_k(\psi_1)\right).$$

Summing over all choices of $\underline{x}$ and making the identical calculations for the denominator, the lemma is proved. □

Certain properties of Gibbs measures are dependent on the regularity of the potential. Some loss of regularity of the potential may be expected when a Gibbs measure is mapped under Π, and in the next section we use the inequalities above to prove the second part of theorem 6.0.1 giving relations between the regularity of ψ_1 and the regularity of ψ_2.

6.3 Regularity of the Potential ψ_2

This section, in which we consider the regularity properties of $\psi_2 := \log u$, is joint work with my supervisor Mark Pollicott. The following is our main result.

Theorem 6.3.1. *Let* $\kappa \geq 0$. *If* $\sum_{n=0}^{\infty} n^{\kappa+1} var_n(\psi_1) < \infty$ *then* $\sum_{n=0}^{\infty} n^{\kappa} var_n(\psi_2) < \infty$.

Proof. Let $0 < c < 1$ be as in lemma 6.2.4. Choose $1 < \beta < 1/(1-c)$ and an integer $M > 1$ sufficiently large that $\alpha := \beta(1-c)\left(1+\frac{1}{M}\right)\left(1-\frac{1}{M}\right)^{-\kappa} < 1$. Let us denote $a_n = \log \lambda_n$ and recall the trivial inequality $1 + x \leq \exp(x) \leq 1 + \beta x$, for $x > 0$ sufficiently small. Thus providing N_0 is sufficiently large we can deduce from corollary 6.2.2 that for any $n > N_0$

$$\begin{aligned}
1 + a_n \le \exp(a_n) &\le c.\exp\left(\sum_{m=[n/M]}^{n} var_m(\psi_1)\right) + (1-c)\exp(a_{n-[n/M]}) \\
&\le c + (1-c) + \beta c \sum_{m=[n/M]}^{n} var_m(\psi_1) + \beta(1-c)a_{n-[n/M]}
\end{aligned}$$

and hence that for any $N > N_0$,

$$\sum_{n=N_0}^{N} n^\kappa a_n \le \beta c \sum_{n=N_0}^{N} n^\kappa \sum_{m=[n/M]}^{n} var_m(\psi_1) + \beta(1-c)\sum_{n=N_0}^{N} n^\kappa a_{n-[n/M]}$$

(where $[\cdot]$ denotes the integer part).

We can bound

$$\begin{aligned}
\sum_{n=N_0}^{N} n^\kappa \sum_{m=[n/M]}^{n} var_m(\psi_1) &\le M^\kappa \sum_{n=N_0}^{N} \sum_{m=[n/M]}^{n} m^\kappa var_m(\psi_1) \\
&\le M^{\kappa+1} \sum_{n=N_0}^{N} n^{\kappa+1} var_n(\psi_1)
\end{aligned}$$

and

$$\begin{aligned}
&\sum_{n=N_0}^{N} n^\kappa a_{n-[n/M]} \\
&\le \frac{1}{\left(1-\frac{1}{M}\right)^\kappa} \sum_{n=N_0}^{N} \left(n - [n/M]\right)^\kappa a_{n-[n/M]} \\
&\le \frac{1}{\left(1-\frac{1}{M}\right)^\kappa} \left(\sum_{m=\left[N_0 - \frac{N_0}{M}\right]}^{\left[N-\frac{N}{M}\right]+1} m^\kappa a_m + \sum_{\substack{N_0 \le n \le N \\ M|n+1}} \left(n - [n/M]\right)^\kappa a_{n-[n/M]} \right) \\
&\le \frac{\left(1+\frac{1}{M}\right)}{\left(1-\frac{1}{M}\right)^\kappa} \sum_{m=N_0}^{N} m^\kappa a_m + O(1)
\end{aligned}$$

where we have used that

$$\sum_{\substack{N_0\leq n\leq N\\ M|n+1}} (n-[n/M])^{\kappa}\, a_{n-[n/M]} \leq \frac{1}{M} \sum_{m=N_0-[N_0/M]}^{N} m^k a_m$$

$$\text{and} \quad \sum_{m=N_0-[N_0/M]}^{N_0-1} m^k a_m = O(1).$$

Comparing the above inequalities we can bound

$$\underbrace{\left(1-\beta(1-c)\frac{(1+\frac{1}{M})}{(1-\frac{1}{M})^{\kappa}}\right)}_{>0} \sum_{n=N_0}^{N} n^{\kappa} a_n \leq \beta c M^{\kappa+1} \sum_{n=N_0}^{N} n^{\kappa+1} var_n(\psi_1) + O(1).$$

Letting $N \to +\infty$ we see that $\sum_{n=N_0}^{\infty} n^{\kappa} a_n < \infty$, which completes the proof. □

When $\kappa = 0$ we have the following corollary.

Corollary 6.3.1. *If $\sum_{n=0}^{\infty} n var_n(\psi_1) < \infty$ then $\sum_{n=0}^{\infty} var_n(\psi_2) < \infty$.*

Another application of corollary 6.2.2 is the following.

Theorem 6.3.2. *Suppose that there exists $c_1 > 0$ and $0 < \theta_1 < 1$ such that $var_n(\psi_1) \leq c_1 \theta_1^{\sqrt{n}}$ for all $n \geq 0$. Then there exists $c_2 > 0$ and $0 < \theta_2 < 1$ such that $var_n(\psi_2) \leq c_2 \theta_2^{\sqrt{n}}$ for all $n \geq 0$.*

Proof. By corollary 6.2.2 we can write

$$\begin{aligned}\lambda_n &\leq c \exp\left(c_1 \sum_{k=n-[\sqrt{n}]}^{n} \theta_1^{\sqrt{k}}\right) + (1-c)\lambda_{n-[\sqrt{n}]}\\ &\leq c \exp\left(C\theta^{[\sqrt{n}]}\right) + (1-c)\lambda_{n-[\sqrt{n}]}\end{aligned}$$

for any $\theta_1 < \theta < 1$ and some $C > 0$. Using this inequality inductively $[\sqrt{n}]$ times, we can write

$$\begin{aligned}
\lambda_n &\leq c\exp\left(C\theta^{[\sqrt{n}]}\right) + (1-c)\left(c\exp\left(C\theta^{[\sqrt{n}]}\right) + (1-c)\lambda_{n-2[\sqrt{n}]}\right) \\
&\quad \cdots \\
&\leq c\exp\left(C\theta^{[\sqrt{n}]}\right)\sum_{k=0}^{[\sqrt{n}]}(1-c)^k + (1-c)^{[\sqrt{n}]}\lambda_{n-[\sqrt{n}]^2} \\
&\leq \exp\left(C\theta^{[\sqrt{n}]}\right) + (1-c)^{[\sqrt{n}]}\lambda_0.
\end{aligned}$$

This generalises the results of Chazottes and Ugalde in [CU03], [CU11], and Verbitskiy in [Ver11]. In particular, we see that $|\lambda_n - 1| = O\left(\theta_2^{\sqrt{n}}\right)$, where $\theta_2 = \max\{\theta, (1-c)\}$ from which the result follows. □

The following is an easy consequence of the theorem and its proof.

Corollary 6.3.2. *Assume there exists $c_1 > 0$ and $0 < \theta < 1$ such that $var_n(\psi_1) \leq c_1\theta^n$ for all $n \geq 0$ (i.e. ψ_1 is Hölder continuous) then there exists $c_2 > 0$ such that $var_n(\psi_2) \leq c_2\theta^{\sqrt{n}}$ for all $n \geq 0$.*

Unfortunately we are unable to improve upon this estimate, and the question of whether Hölder continuity of the potential is preserved under mapping by Π remains open, with the exception of a special case given in the next section.

6.4 Examples and Comments

First we give an example which shows that some condition such as Hypothesis 6.0.1 on the map $\Pi : \Sigma_1 \to \Sigma_2$ is necessary. It was conjectured in [CU11] that for more general factor maps Π the image measure ν might still satisfy the inequality in definition 2.2.3 for almost every z. We show that this need not be the case.

Example 6.4.1. *Consider the shift (Σ_1, σ) associated to the transition matrix*

$$M = \begin{pmatrix} 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}$$

and consider a factor map $\Pi : \Sigma_1 \to \Sigma_2$ with $\Pi(1) = 1$, $\Pi(2) = 2$ and $\Pi(3) = \Pi(4) = 3$. Let $\psi_1 : \Sigma_1 \to \mathbb{R}$ be a Hölder continuous function (such that $P(\psi_1) = 0$) with Gibbs measure μ. We suppose that $\mu \circ \Pi^{-1} = \nu$ satisfies the inequality in Definition 2.2.3 for almost every $\underline{z}$. Then, for almost all $\underline{w}$,

$$C_1 \exp(\psi_2^{n+1}(1\underbrace{3\cdots3}_{n}\underline{w})) \leq \nu[1\underbrace{3\cdots3}_{n}] = \mu[1\underbrace{3\cdots3}_{n}]$$

and

$$C_2 \exp(\psi_2^{n+1}(2\underbrace{3\cdots3}_{n}\underline{w})) \geq \nu[2\underbrace{3\cdots3}_{n}] = \mu[2\underbrace{4\cdots4}_{n}]$$

If we further suppose that μ is a Bernoulli measure with $\mu[3] < \mu[4]$, we need only take n large enough such that

$$\frac{\mu[1]\mu[3]^n}{C_1 \exp(\inf_{\underline{z}} \psi_2(\underline{z}))} < \frac{\mu[2]\mu[4]^n}{C_2 \exp(\sup_{\underline{z}} \psi_2(\underline{z}))}$$

and we see

$$\psi_2^n(\underbrace{3\cdots3}_{n}\underline{w}) < \psi_2^n(\underbrace{3\cdots3}_{n}\underline{w})$$

for any $\underline{w}$, thus ψ_2^n is undefined on $[\underbrace{3\cdots3}_{n}]$ which is a set of positive measure. So there is a set of positive measure on which ν does not satisfy the inequality in Definition 2.2.3, and so ν is not a Gibbs measure.

We now assume that Π satisfies the conditions of hypothesis 6.0.1, and consider a class of potentials for which Hölder continuity is preserved under Π.

Example 6.4.2. *Suppose that ψ_1 can be expressed as*

$$\psi_1(\underline{x}) = f_0(x_0, x_1) + f_1(x_1, x_2) + \cdots$$

Then Hölder continuity of ψ_1 implies Hölder continuity of ψ_2.

Proof. Given some choice of w_0, the dependence of $u_{j,\underline{w},n}(\underline{z})$ on the later terms in $\underline{w}$ is less than $var_n(\psi_1)$. This is because, given $x_0 \cdots x_n$ and $\underline{w}$, $\underline{w}'$ with $w_0 = w_0'$,

$$\psi_1(\sigma^i(x_0 \cdots x_n\underline{w})) - \psi_1(\sigma^i(x_0 \cdots x_n\underline{w}')) = \sum_{k=0}^{\infty} f_{n-i+k}(w_k, w_{k+1}) - f_{n-i+k}(w_k', w_{k+1}')$$

and hence

$$\frac{\sum_{\underline{x}=x_0\cdots x_{n-1}j} \exp(\psi_1^{n+1}(\underline{xw}))}{\sum_{\underline{x}=x_0\cdots x_{n-1}j} \exp(\psi_1^{n+1}(\underline{xw}'))} \le \exp\left(\sum_{i=0}^{n}\sum_{k=0}^{\infty} f_{n-i+k}(w_k, w_{k+1}) - f_{n-i+k}(w_k', w_{k+1}')\right)$$

This is independent of the choice of $\underline{x}$. Thus we have

$$\begin{aligned}
\frac{u_{j,\underline{w},n}(\underline{z})}{u_{j,\underline{w}',n}(\underline{z})} &= \frac{\sum_{\underline{x}=x_0\cdots x_{n-1}j} \exp(\psi_1^{n+1}(\underline{xw}))}{\sum_{\underline{x}'=x_1\cdots x_{n-1}j} \exp(\psi_1^{n}(\underline{x}'\underline{w}))} \cdot \frac{\sum_{\underline{x}'=x_1\cdots x_{n-1}j} \exp(\psi_1^{n}(\underline{x}'\underline{w}'))}{\sum_{\underline{x}=x_0\cdots x_{n-1}j} \exp(\psi_1^{n+1}(\underline{xw}'))} \\
&= \frac{\sum_{\underline{x}=x_0\cdots x_{n-1}j} \exp(\psi_1^{n+1}(\underline{xw}))}{\sum_{\underline{x}=x_0\cdots x_{n-1}j} \exp(\psi_1^{n+1}(\underline{xw}'))} \frac{\sum_{\underline{x}'=x_1\cdots x_{n-1}j} \exp(\psi_1^{n}(\underline{x}'\underline{w}'))}{\sum_{\underline{x}'=x_1\cdots x_{n-1}j} \exp(\psi_1^{n}(\underline{x}'\underline{w}))} \\
&= \frac{\exp(\sum_{i=0}^{n}\sum_{k=0}^{\infty} f_{n-i+k}(w_k, w_{k+1}) - f_{n-i+k}(w_k', w_{k+1}'))}{\exp(\sum_{i=1}^{n}\sum_{k=0}^{\infty} f_{n-i+k}(w_k, w_{k+1}) - f_{n-i+k}(w_k', w_{k+1}'))} \\
&= \exp(\sum_{k=0}^{\infty} f_{n+k}(w_k, w_{k+1}) - f_{n+k}(w_k', w_{k+1}')) \\
&\le var_n(\psi_1).
\end{aligned}$$

The appearance of $\exp(2\sum_{k=s-n}^{s} var_k(\psi_1))$ in the statement of lemma 6.2.3 appears as a maximal value of $\frac{u_{j,\overline{x}\underline{w},n}(\underline{z})}{u_{j,\overline{x}\underline{w}',n}(\underline{z})}$. Choosing $s = n + 1$ and putting $\overline{x} = w_0$, the statement of lemma 6.2.3 now becomes

$$\lambda_{n+3} \leq c(\exp(var_n(\psi_1))) + (1 - c)\lambda_n$$

which in particular gives that Hölder potentials project to Hölder potentials. This generalizes the result in [CU03], where it was shown that Gibbs measures with locally constant potentials (Markov measures) project to Gibbs measures with Hölder potentials. □

6.5 Comments on the Technical Hypothesis

We recall that, given a set $B \subset \Sigma_1$, the set $\mathcal{A}_n(B)$ was defined to be the set of values of x_n for sequences $\underline{x}$ in B. Our technical hypothesis on Π was as follows.

Hypothesis. *We assume that for* $\Pi : \Sigma_1 \to \Sigma_2$ *there exists a natural number* N *such that for any* $\underline{z} \in \Sigma_2$,

1. *If* $\mathcal{A}_n\{\underline{x} : x_{n+m} = j, \Pi(\underline{x}) = \underline{z}\}$ *is non-empty for some* $m > N$*, then* $\mathcal{A}_n\{\underline{x} : x_{n+m} = j, \Pi(\underline{x}) = \underline{z}\} = \mathcal{A}_n\{\Pi(\underline{x}) = \underline{z}\}$.
2. $\mathcal{A}_n\{\underline{x} : \Pi(x_{n-N} \cdots x_{n+N}) = z_{n-N} \cdots z_{n+N}\} = \mathcal{A}_n\{\underline{x} : \Pi(\underline{x}) = \underline{z}\}$.

Some understanding of these conditions can be gained by considering the non-homogeneous symbolic spaces of Fan and Pollicott in [FP00]. Let M be the incidence matrix of Σ_1. If for each $\underline{z}$ we consider the submatrix M_n of M given by rows corresponding to symbols in $\mathcal{A}_n\{\underline{x} : \Pi(\underline{x}) = \underline{z}\}$ and columns corresponding to symbols in $\mathcal{A}_{n+1}\{\underline{x} : \Pi(\underline{x}) = \underline{z}\}$, then the matrices M_n give rise to a non-homogeneous symbolic space. Sequences $\underline{x}$ projecting to $\underline{z}$ correspond to sequences

$\{\underline{x} : M_n(x_n, x_{n+1}) = 1 \forall n \in \mathbb{N}\}$. Part (i) of Hypothesis 6.0.1 corresponds to equation (1) of [FP00], that there exists an N such that for all $j > 0$ the product $\Pi_{n=j}^{j+N} M_n$ is a strictly positive matrix. Part (ii) requires that we can determine the matrices M_n by looking only at $z_{n-N} \cdots z_{n+N}$ rather than considering all of $\underline{z}$.

The topological conditions of [CU03] can also be understood with reference to non-homogeneous symbolic spaces. In that article, matrices M'_n were defined to be the submatrices of M with rows corresponding to elements of $\Pi^{-1}(z_n)$ and columns corresponding to $\Pi^{-1}(z_{n+1})$. The matrices M'_n are larger than our matrices M_n. The first topological condition was that, for a word $z_n \cdots z_{n+k}$ with $z_n = z_{n+k}$, the product of matrices $M'_n \cdots M'_{n+k}$ should be a positive matrix. The second topological condition was that any word $x_1 \cdots x_n$ projecting to $z_1 \cdots z_n$ should be extendable to a sequence $\underline{x}$ projecting to $\underline{z}$, or that no row in any matrix M'_n should be completely empty.

Since $A = \{1, \cdots, k_1\}$ is finite, we have by the pigeonhole principle that any word $x_m \cdots x_{m+k_1+1}$ must have at least one repeated digit, and then the two topological conditions of [CU03] give that $M_n \cdots M_{n+k_1+1}$ must be a strictly positive matrix, and hence imply Part (i) of our Hypothesis 6.0.1. The second topological condition of [CU03] gives that any word $x_1 \cdots x_n$ projecting to $z_1 \cdots z_n$ can be extended to a sequence $\underline{x} \in \Pi^{-1}(\underline{z})$, which implies part (ii) of our Hypothesis 6.0.1. Thus our topological conditions are weaker than those in [CU03].

6.6 Renormalization of Gibbs Measures in Statistical Mechanics

This problem fits into the broader framework of the study of renormalizations of Gibbs measures in statistical mechanics, where one considers a Gibbs measure μ on a space X and a map $\Pi : X \to Y$, and asks about the image measure $\nu = \mu \circ \Pi^{-1}$ on Y. The map Π is called a renormalization map, this term is common among the statistical mechanics community because of the connections with renormalization group theory in physics. Certain technical problems in renormalization group theory were found to be the result of maps Π under which Gibbs measures map to non Gibbs measures and so concerted efforts have been made to understand the conditions under which this can happen.

In the case that Π maps a Gibbs measure μ to a non Gibbs measure ν, Dobrushin asked whether any of the properties of Gibbs measures still hold for ν. This became known as Dobrushin's restoration programme, and has been the focus of much work within statistical mechanics. Example 6.4.1 gives a negative answer to part of the programme, but other aspects remain open or have positive answers. For example, Verbitskiy has shown in a recent article [Ver10] that renormalized Gibbs measures still satisfy a variational principle even if they are not Gibbs measures.

Bibliography

[Abr59] L. M. Abramov. On the entropy of a flow. *Dokl. Akad. Nauk SSSR*, 128:873–875, 1959.

[AL11] D. Ahmadi Dastjerdi and S. Lamei. Generating functions for special flows over the 1-step countable topological markov chains. *ArXiv e-prints*, January 2011.

[BG08] A. I. Bufetov and B. M. Gurevich. On a measure with maximum entropy for the Teichmüller flow on the moduli space of abelian differentials. *Funktsional. Anal. i Prilozhen.*, 42(3):75–77, 2008.

[BI06] L. Barreira and G. Iommi. Suspension flows over countable Markov shifts. *J. Stat. Phys.*, 124(1):207–230, 2006.

[Bil99] P. Billingsley. *Convergence of probability measures.* Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons Inc., New York, second edition, 1999. A Wiley-Interscience Publication.

[BLL10] A. T. Baraviera, R. Leplaideur, and A. O. Lopes. Selection of measures for a potential with two maxima at the zero temperature limit. *ArXiv e-prints*, April 2010.

[Bow70] R. Bowen. Topological entropy and axiom A. In *Global Analysis (Proc. Sympos. Pure Math., Vol. XIV, Berkeley, Calif., 1968)*, pages 23–41. Amer. Math. Soc., Providence, R.I., 1970.

[Bow72] R. Bowen. Periodic orbits for hyperbolic flows. *Amer. J. Math.*, 94:1–30, 1972.

[Bow73] R. Bowen. Topological entropy for noncompact sets. *Trans. Amer. Math. Soc.*, 184:125–136, 1973.

[BP11] M. Boyle and K. Petersen. Hidden Markov processes in the context of symbolic dynamics. In *Entropy of Hidden Markov Processes and Connections to Dynamical Systems*. 2011.

[BR75] R. Bowen and D. Ruelle. The ergodic theory of Axiom A flows. *Invent. Math.*, 29(3):181–202, 1975.

[Bré03] J. Brémont. Gibbs measures at temperature zero. *Nonlinearity*, 16(2):419–426, 2003.

[BS03] J. Buzzi and O. Sarig. Uniqueness of equilibrium measures for countable Markov shifts and multidimensional piecewise expanding maps. *Ergodic Theory Dynam. Systems*, 23(5):1383–1400, 2003.

[CGU09] J. R. Chazottes, J. M. Gambaudo, and E. Ugalde. Zero-temperature limits of one-dimensional Gibbs states via renormalization: the case of locally constant potentials, 2009.

[CH10] J. R. Chazottes and M. Hochman. On the zero-temperature limit of Gibbs states. *Comm. Math. Phys.*, 297(1):265–281, 2010.

[Coe90] Z. Coelho. *Entropy and Ergodicity of Skew-Products over Subshifts of Finite Type and Central Limit Asymptotics.* PhD thesis, University of Warwick, 1990.

[CU03] J. R. Chazottes and E. Ugalde. Projection of Markov measures may be Gibbsian. *J. Statist. Phys.*, 111(5-6):1245–1272, 2003.

[CU11] J. R. Chazottes and E. Ugalde. Preservation of Gibbsianness under amalgamation of symbols. In *Entropy of Hidden Markov Processes and Connections to Dynamical Systems.* 2011.

[DdV07] K. Dajani and M. de Vries. Invariant densities for random β-expansions. *J. Eur. Math. Soc. (JEMS)*, 9(1):157–176, 2007.

[DK02] K. Dajani and C. Kraaikamp. *Ergodic theory of numbers*, volume 29 of *Carus Mathematical Monographs.* Mathematical Association of America, Washington, DC, 2002.

[DK03] K. Dajani and C. Kraaikamp. Random β-expansions. *Ergodic Theory Dynam. Systems*, 23(2):461–479, 2003.

[FP00] A. H. Fan and M. Pollicott. Non-homogeneous equilibrium states and convergence speeds of averaging operators. *Math. Proc. Cambridge Philos. Soc.*, 129(1):99–115, 2000.

[GK01] B. Gurevich and S. Katok. Arithmetic coding and entropy for the positive geodesic flow on the modular surface. *Mosc. Math. J.*, 1(4):569–582, 645, 2001. Dedicated to the memory of I. G. Petrovskii on the occasion of his 100th anniversary.

[Ham10] U. Hamenstädt. Bowen's construction for the Teichmüller flow. *ArXiv e-prints*, July 2010.

[Iom07] G. Iommi. Ergodic optimization for renewal type shifts. *Monatsh. Math.*, 150(2):91–95, 2007.

[Iom10] G. Iommi. Thermodynamic formalism for the positive geodesic flow on the modular surface. *ArXiv e-prints*, September 2010.

[Jen06] O. Jenkinson. Ergodic optimization. *Discrete Contin. Dyn. Syst.*, 15(1):197–224, 2006.

[JMU05] O. Jenkinson, R. D. Mauldin, and M. Urbański. Zero temperature limits of Gibbs-equilibrium states for countable alphabet subshifts of finite type. *J. Stat. Phys.*, 119(3-4):765–776, 2005.

[JMU06] O. Jenkinson, R. D. Mauldin, and M. Urbański. Ergodic optimization for countable alphabet subshifts of finite type. *Ergodic Theory Dynam. Systems*, 26(6):1791–1803, 2006.

[Kat96] S. Katok. Coding of closed geodesics after Gauss and Morse. *Geom. Dedicata*, 63(2):123–145, 1996.

[Kem11] T. Kempton. Factors of Gibbs measures for subshifts of finite type. *Bulletin of the London Mathematical Society*, 2011.

[KP11] T. Kempton and M. Pollicott. Factors of Gibbs measures for full shifts. In *Entropy of Hidden Markov Processes and Connections to Dynamical Systems*. 2011.

[Kre67] U. Krengel. Entropy of conservative transformations. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 7:161–181, 1967.

[Lep05] R. Leplaideur. A dynamical proof for the convergence of Gibbs measures at temperature zero. *Nonlinearity*, 18(6):2847–2880, 2005.

[Mor07] I. D. Morris. Entropy for zero-temperature limits of Gibbs-equilibrium states for countable-alphabet subshifts of finite type. *J. Stat. Phys.*, 126(2):315–324, 2007.

[MU01] R. D. Mauldin and M. Urbański. Gibbs states on the symbolic space over an infinite alphabet. *Israel J. Math.*, 125:93–130, 2001.

[MU03] R. D. Mauldin and M. Urbański. *Graph directed Markov systems*, volume 148 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 2003. Geometry and dynamics of limit sets.

[Pol01] A. B. Polyakov. On a measure with maximal entropy for a special flow over a local perturbation of a countable topological Bernoulli scheme. *Mat. Sb.*, 192(7):73–96, 2001.

[PP84] Y. B. Pesin and B. S. Pitskel. Topological pressure and the variational principle for noncompact sets. *Funktsional. Anal. i Prilozhen.*, 18(4):50–63, 96, 1984.

[Sal88] I. A. Salama. Topological entropy and recurrence of countable chains. *Pacific J. Math.*, 134(2):325–341, 1988.

[Sar] O. Sarig. Lecture notes on thermodynamic formalism for countable Markov shifts. http://www.math.psu.edu/sarig/.

[Sar99] O. Sarig. Thermodynamic formalism for countable Markov shifts. *Ergodic Theory Dynam. Systems*, 19(6):1565–1593, 1999.

[Sar03] O. Sarig. Existence of Gibbs measures for countable Markov shifts. *Proc. Amer. Math. Soc.*, 131(6):1751–1758 (electronic), 2003.

[Sav98] S. V. Savchenko. Special flows constructed from countable topological Markov chains. *Funktsional. Anal. i Prilozhen.*, 32(1):40–53, 96, 1998.

[Ser85] C. Series. The modular surface and continued fractions. *J. London Math. Soc. (2)*, 31(1):69–80, 1985.

[SF] N. Sidorov and D. J. Feng. Growth rate for beta-expansions. *To appear in Monash. Math.*

[Sid03a] N. Sidorov. Almost every number has a continuum of β-expansions. *Amer. Math. Monthly*, 110(9):838–842, 2003.

[Sid03b] N. Sidorov. Ergodic-theoretic properties of certain Bernoulli convolutions. *Acta Math. Hungar.*, 101(4):345–355, 2003.

[Sin72] J. G. Sinaĭ. Gibbs measures in ergodic theory (English translation). *Russian Math. Surveys*, 27(4):21–64, 1972.

[Tho11] D. Thompson. A thermodynamic definition of topological pressure for non-compact sets. *Ergodic Theory Dynam. Systems*, 2011.

[Ver10] E. Verbitskiy. A variational principle for fuzzy Gibbs measures. *Moscow Math. Journal*, 10(4):811–829, 2010.

[Ver11] E. Verbitskiy. Thermodynamics of hidden Markov processes. In *Entropy of Hidden Markov Processes and Connections to Dynamical Systems*. 2011.

[Wal82] P. Walters. *An introduction to ergodic theory*, volume 79 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1982.