

Matrix Gibbs states, factor maps and transfer operators

by

Mark Piraino

M.S., DePaul University, 2015

B.S., DePaul University, 2014

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

DOCTOR OF PHILOSOPHY

in the Department of Mathematics and Statistics

© Mark Piraino, 2019

University of Victoria

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

Matrix Gibbs states, factor maps and transfer operators

by

Mark Piraino

M.S., DePaul University, 2015

B.S., DePaul University, 2014

Supervisory Committee

Dr. Christopher Bose, Co-supervisor
(Department of Mathematics and Statistics)

Dr. Anthony Quas, Co-supervisor
(Department of Mathematics and Statistics)

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

Supervisory Committee

Dr. Christopher Bose, Co-supervisor
(Department of Mathematics and Statistics)

Dr. Anthony Quas, Co-supervisor
(Department of Mathematics and Statistics)

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

ABSTRACT

We study two problems. The first concerning ergodic properties of measures on $\Sigma^{\mathbb{Z}}$ such that $\mu_{\mathcal{A},t}[x_0 \cdots x_{n-1}] \approx e^{-nP} \left\| A_{x_0} \cdots A_{x_{n-1}} \right\|^t$ where $\mathcal{A} = (A_0, \dots, A_{M-1})$ is a collection of matrices, such measures are known as matrix Gibbs states. In particular we give a sufficient condition for $\mu_{\mathcal{A},t}$ to be isomorphic to a Bernoulli shift and mix at an exponential rate. The second problem concerns factors of Gibbs states. In particular we show that all of classical uniqueness regimes for Gibbs states are closed under factor maps which satisfy a mixing in fibers condition. The unifying approach to both of these problems is to realize the measure of cylinder sets in terms of positive operators.

Table of Contents

Supervisory Committee	ii
Abstract	iii
Table of Contents	iv
List of Figures	vi
Acknowledgements	vi
Dedication	viii
1 Introduction	1
1.1 Overview	1
1.2 Shifts of finite type and factor maps	4
1.3 (Scalar) Thermodynamic formalism	8
1.3.1 g-measures	9
1.3.2 Equilibrium and Gibbs states	10
1.3.3 Gibbs states for Hölder potentials on a full shift	12
2 Matrix Equilibrium States	15
2.1 Introduction	15
2.2 Matrices which preserve a common cone	19
2.3 Transfer operators and exponential mixing	26
2.4 The Weak Bernoulli Property	35
3 Factors of Gibbs States	38
3.1 Introduction	38
3.2 Examples: Hidden Markov measures	40
3.3 Factors of Hölder Gibbs states on a full shift	43

3.4	The classical uniqueness regimes	46
3.5	Final remarks	63
Appendix A Cone Theory		64
A.1	Cones in $C(\Sigma_A^+)$	64
A.2	Ruelle-Perron-Frobenius theorems	73
A.3	Finite dimensional Perron-Frobenius Theory	79
Bibliography		88

List of Figures

- Figure 1: A graph.
- Figure 2: A graph in which some of the vertexes have been colored.

Acknowledgements

I would like to thank:

- **My supervisors:** Christopher Bose and Anthony Quas for their guidance, advice and support. Without them this work would not have been possible.
- **The examining committee:** Jairo Bochi for a careful reading of this thesis and many useful comments. Pavel Kovtun for a careful reading of this thesis and useful suggestions. In addition I would like to thank the anonymous referee of the article [36] for pointing me to [21] which greatly improved the contents of Chapter 2.
- **My teachers:** Many individuals have invested time and effort teaching me mathematics, without them I would never have been in a position to attempt this.
- **My friends:** All of the people who have made the last four years an enjoyable experience.

Dedication

To my parents.

“Go west, young man, go west and grow up with the country.” -Horace Greeley

Chapter 1

Introduction

1.1 Overview

This thesis concerns two problems, the first about ergodic properties of equilibrium states in matrix thermodynamic formalism and the second, regularity properties of factors of Gibbs states. To motivate these problems let's consider an example which sits at the intersection of these two areas. Consider a particle which moves according to a biased random walk on the following graph

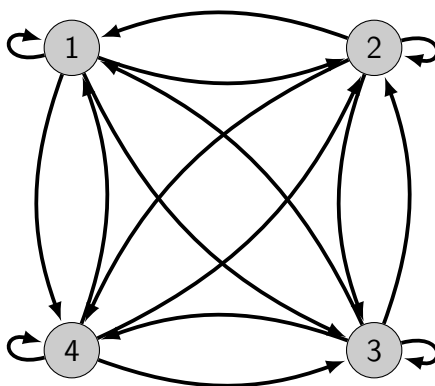


Figure 1: A graph.

so that the probability of the particle moving to state i given that it is at state j is given by S_{ij} where S is the matrix

$$S = \begin{bmatrix} 1/3 & 1/3 & 1/6 & 1/6 \\ 1/4 & 1/8 & 1/8 & 1/2 \\ 1/3 & 1/3 & 1/6 & 1/6 \\ 1/4 & 1/8 & 1/8 & 1/2 \end{bmatrix}.$$

This process has 1 step of memory in the sense that the probability of the particle being in state i given the entire past only depends on the particle's previous state j . These kind of random processes which have only 1 step of memory are called *Markov*. Denote by ν the left eigenvector for S normalized so that $\sum_i \nu_i = 1$ then the probability of seeing the sequence 1123 is given by

$$\nu_1 S_{11} S_{12} S_{23}.$$

Now suppose that we color the vertexes

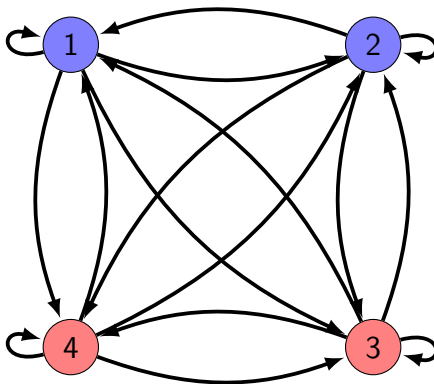


Figure 2: A graph in which some of the vertexes have been colored.

and we can't observe what state the particle is in but we can observe what color the state is. This gives us a new random process which takes the values "red" and "blue"(called a hidden Markov process), interestingly this random process is

not Markov. In fact the process has infinite memory and it is thus natural to ask if the memory of the process depends weakly on the past or decays at some rate (for example exponentially fast). This is the basic question we address in chapter 3 where we consider a more general situation known as factors of Gibbs states. Next notice that if we define the matrices

$$\begin{aligned} \mathcal{S}_{rr} &= \begin{bmatrix} 1/3 & 1/3 \\ 1/4 & 1/8 \end{bmatrix}, \mathcal{S}_{rb} = \begin{bmatrix} 1/6 & 1/6 \\ 1/8 & 1/2 \end{bmatrix} \\ \mathcal{S}_{br} &= \begin{bmatrix} 1/3 & 1/3 \\ 1/4 & 1/8 \end{bmatrix}, \mathcal{S}_{bb} = \begin{bmatrix} 1/6 & 1/6 \\ 1/8 & 1/2 \end{bmatrix} \end{aligned}$$

and the vectors

$$\nu_r = \begin{bmatrix} \nu_1 & \nu_2 \end{bmatrix}, \nu_b = \begin{bmatrix} \nu_3 & \nu_4 \end{bmatrix},$$

then a simple computation shows that the probability of observing the sequence $rrbr$ can be written

$$\nu_r \mathcal{S}_{rr} \mathcal{S}_{rb} \mathcal{S}_{br} [1] \tag{1.1}$$

where $[1]$ is the vector of all 1s. As $\nu_i > 0$ we have that this is approximately $\|\mathcal{S}_{rr} \mathcal{S}_{rb} \mathcal{S}_{br}\|$. Probability measures with this property are known as matrix Gibbs states. The focus of chapter 2 will be the study of ergodic and statistical properties of these matrix Gibbs states.

This work provides a unified approach to the study of both of these objects. The key insight is to realize the measure of a cylinder in terms of positive operators and use techniques from functional analysis. In particular we are able to show that ergodic and statistical properties of matrix Gibbs states can be deduced from the spectrum of a suitable transfer operator. This is similar to a well known method originally

due to Ruelle [39] who first used transfer operators to study Gibbs states in scalar thermodynamic formalism (see section 1.3). In the same way that we have realized the measure of a cylinder set for a hidden Markov measure in terms of products of non-negative matrices (equation (1.1)) we can realize the measure of a cylinder set for the factor of a Gibbs state in terms of positive operators acting on an infinite dimensional space. Using cone techniques we are able to study the regularity properties of the conditional probabilities for these measures.

1.2 Shifts of finite type and factor maps

In this section we collect some background on shifts of finite type and 1-block factor maps which we will use in future chapters.

Proposition 1.2.1. Suppose that Σ is a finite set.

1. Let $\{\alpha_n\}$ be a positive decreasing sequence converging to 0. The function

$$d(x, y) = \inf \{ \alpha_n : x_i = y_i \text{ for all } -n \leq i \leq n \}$$

is an ultra-metric on the set $\Sigma^{\mathbb{Z}}$. Moreover the topology induced by d is the same as the product topology on $\Sigma^{\mathbb{Z}}$ (where Σ is given the discrete topology).

2. Let $\{\alpha_n\}$ be a positive decreasing sequence converging to 0. The function

$$d(x, y) = \inf \{ \alpha_n : x_i = y_i \text{ for all } 0 \leq i \leq n \}$$

is an ultra-metric on the set $\Sigma^{\mathbb{N}}$. Moreover the topology induced by d is the same as the product topology on $\Sigma^{\mathbb{N}}$ (where Σ is given the discrete topology).

It is standard to take the sequence $\alpha_n = 2^{-n}$ and functions on $\Sigma^{\mathbb{Z}}$ are said to

be Hölder if they are Hölder continuous in the metric induced by this sequence. For $f \in C(\Sigma^{\mathbb{N}})$ or $C(\Sigma^{\mathbb{Z}})$ define

$$\text{var}_n f = \sup \{|f(x) - f(y)| : x_i = y_i \text{ for all } |i| \leq n - 1\}.$$

Definition 1.2.2. Suppose that Σ is a finite set (we refer to Σ as the set of *symbols*). Define the *two-sided shift* $\sigma : \Sigma^{\mathbb{Z}} \rightarrow \Sigma^{\mathbb{Z}}$ by $\sigma(x_i)_{i=-\infty}^{\infty} = (x_{i+1})_{i=-\infty}^{\infty}$ and similarly the *one-sided shift* $\sigma : \Sigma^{\mathbb{N}} \rightarrow \Sigma^{\mathbb{N}}$ by $\sigma(x_i)_{i=0}^{\infty} = (x_{i+1})_{i=0}^{\infty}$.

Often we think of the two-sided shift as acting by moving a distinguished zero position by one place. Consider a point $(x_i)_{i=-\infty}^{\infty}$ the shift action does the following

$$\cdots x_{-2}x_{-1}\widehat{x}_0x_1x_2 \cdots \xrightarrow{\sigma} \cdots x_{-1}x_0\widehat{x}_1x_2x_3 \cdots$$

where $\widehat{}$ marks the distinguished zero position. The one-sided shift we often think of as deleting the zeroth entry of a string, that is the shift does the following

$$x_0x_1x_2 \cdots \xrightarrow{\sigma} x_1x_2x_3 \cdots .$$

It is not difficult to see that σ is continuous and in the case of $\Sigma^{\mathbb{Z}}$ it is a bijection.

Definition 1.2.3. A subset X of $\Sigma^{\mathbb{Z}}$ or $\Sigma^{\mathbb{N}}$ is called a *subshift* if X is closed and $\sigma(X) \subseteq X$. The pair $(X, \sigma|_X)$ is called a *symbolic dynamical system*.

To avoid ambiguity and to emphasize the space X we will sometimes write the shift action on X as σ_X , otherwise we simply write σ when the space is understood. We refer to a finite string of symbols $x_0 \cdots x_n$ as a *word*. A word w is *admissible* or *allowed* if there exists a point $x \in X$ with $w = x_0 \cdots x_n$ the collection of all admissible words is called the *language* of the subshift X .

One important class of subshifts are shifts of finite type. On the one hand these subshifts appear relatively simple but they play an important role in many areas of dynamical systems. Perhaps the most well known example is the connection between shifts of finite type and axiom A diffeomorphisms and flows via Markov partitions [3], [4].

Definition 1.2.4. Suppose that $\Sigma = \{1, \dots, m\}$ and A is an $m \times m$ matrix with $A_{ij} \in \{0, 1\}$. Define the *shift of finite type* determined by the matrix A to be

$$\Sigma_A = \left\{ (x_i)_{i=-\infty}^{\infty} \in \Sigma^{\mathbb{Z}} : A_{x_i x_{i+1}} = 1 \text{ for all } i \in \mathbb{Z} \right\}$$

and similarly on the one-sided shift

$$\Sigma_A^+ = \left\{ (x_i)_{i=0}^{\infty} \in \Sigma^{\mathbb{N}} : A_{x_i x_{i+1}} = 1 \text{ for all } i \geq 0 \right\}.$$

It can be verified that Σ_A and Σ_A^+ are subshifts. A shift of finite type Σ_A is called *topologically mixing* if there exists an M such that $(A^M)_{ij} > 0$ for all i, j .

Definition 1.2.5. Let X be a subshift. A Borel probability measure μ on X is called *shift invariant* if $\mu(\sigma^{-1}B) = \mu(B)$ for all Borel sets $B \subseteq X$.

Recall the definition of the *cylinder set*

$$[x_0 \cdots x_n] = \{z \in X : z_i = x_i \text{ for all } 0 \leq i \leq n\}$$

we will also write

$$[{}_t x_0 \cdots x_n] = \{z \in X : z_i = x_i \text{ for all } t \leq i \leq t+n\}.$$

Borel measures are determined by their values on cylinder sets and ergodic properties

of a shift invariant measure can be deduced using only cylinder sets. One of the most common ways of constructing measures on $\Sigma^{\mathbb{Z}}$ is using the Kolmogorov extension theorem we summarize the method in the following proposition.

Proposition 1.2.6. Let Σ be a finite set. Suppose that for each $n \geq 0$ and $a_0, \dots, a_n \in \Sigma$ we have a non-negative number $p_n(a_0 \cdots a_n)$ such that

$$\sum_{i \in \Sigma} p_0(i) = 1$$

and

$$\sum_{i \in \Sigma} p_{n+1}(a_0 \cdots a_n i) = p_n(a_0 \cdots a_n) = \sum_{i \in \Sigma} p_{n+1}(i a_0 \cdots a_n). \quad (1.2)$$

Then there exists a unique shift invariant Borel measure μ on $\Sigma^{\mathbb{Z}}$ such that

$$\mu[t x_0 \cdots x_n] = p_n(x_0 \cdots x_n)$$

for all $t \in \mathbb{Z}$ and $n \geq 0$.

Proof. This follows from the Kolmogorov extension theorem as (1.2) implies the required consistency conditions. \square

Note also that same result holds where $\Sigma^{\mathbb{Z}}$ is replaced by $\Sigma^{\mathbb{N}}$. Thus studying measures on $\Sigma^{\mathbb{Z}}$ or $\Sigma^{\mathbb{N}}$ can be reduced to finding and studying the values of the measure on cylinder sets. The following proposition is often useful.

Proposition 1.2.7. Suppose that μ is a shift invariant measure.

1. μ is ergodic if and only if for all cylinder sets $[I], [J]$

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} \mu(\sigma^{-i}[I] \cap [J]) = \mu([I])\mu([J]).$$

2. μ is mixing if and only if for all cylinder sets $[I], [J]$

$$\lim_{n \rightarrow \infty} \mu(\sigma^{-n}[I] \cap [J]) = \mu([I])\mu([J]).$$

Proof. This is [42, Theorem 1.17]. □

Proposition 1.2.8. Suppose that Σ and $\bar{\Sigma}$ are two sets of symbols with $|\bar{\Sigma}| \leq |\Sigma|$ and $\pi : \Sigma \rightarrow \bar{\Sigma}$. Define $\pi : \Sigma^{\mathbb{N}} \rightarrow \bar{\Sigma}^{\mathbb{N}}$ by

$$\pi[(x_i)_{i=0}^{\infty}] = (\pi(x_i))_{i=0}^{\infty}.$$

Then π is continuous and $\sigma_{\bar{\Sigma}^{\mathbb{N}}} \circ \pi = \pi \circ \sigma_{\Sigma^{\mathbb{N}}}$.

Maps of the type described in Proposition 1.2.8 are called *1-block factor maps*.

Proposition 1.2.9. If $X \subseteq \Sigma^{\mathbb{N}}$ is a subshift then $\pi(X) \subseteq \bar{\Sigma}^{\mathbb{N}}$ is a subshift.

Set $Y = \pi(X)$. Notice that the map π induces a map $C(Y) \rightarrow C(X)$, $f \mapsto f \circ \pi$, by duality π also induces a map between Borel measures on X and Borel measures on Y $\mu \mapsto \pi_*\mu$. Moreover because $\sigma_{\bar{\Sigma}^{\mathbb{N}}} \circ \pi = \pi \circ \sigma_{\Sigma^{\mathbb{N}}}$ we have that if μ is shift invariant then $\pi_*\mu$ is shift invariant. It can be shown that for any Borel set $B \subseteq Y$ we have that

$$\pi_*\mu(B) = \mu(\pi^{-1}B).$$

1.3 (Scalar) Thermodynamic formalism

Thermodynamic formalism is a field which sits at the intersection of probability, statistical physics and dynamical systems. Broadly the purpose of thermodynamic

formalism is to select invariant measures with specific properties. For example measures of maximal entropy or measures with certain conditional probabilities. These measures are typically constructed using the eigendata of a Ruelle operator. Recall that given a continuous function $\varphi : X \rightarrow \mathbb{R}$ we define the *Ruelle operator* (sometimes referred to as the *transfer operator*) for φ , $L_\varphi : C(\Sigma_A^+) \rightarrow C(\Sigma_A^+)$, by

$$L_\varphi f(x) = \sum_{\sigma y=x} e^{\varphi(y)} f(y).$$

Moreover ergodic and statistical properties are deduced from the convergence properties of $\rho(L_\varphi)^{-n} L_\varphi^n$ (where $\rho(L_\varphi)$ is the spectral radius) using a suitable Ruelle-Perron-Frobenius (RPF) theorem for example those in theorems A.2.2 and A.2.1.

1.3.1 g-measures

Let Σ_A^+ be a topologically mixing shift of finite type and $g : \Sigma_A^+ \rightarrow \mathbb{R}$ such that $g > 0$ and

$$\sum_{\sigma y=x} g(y) = 1 \text{ for all } x \in \Sigma_A^+.$$

It is natural to think of the value $g(ix_1x_2\cdots)$ as being the conditional probability that one observes state i given one has seen $x_1x_2\cdots$. However this immediately begs the question is there a probability measure \mathbb{P} such that $\mathbb{P}(i|x_1x_2\cdots) = g(ix_1\cdots)$? Such a probability is called a g -measure. These g -measures are connected to the eigendata for the transfer operator associated to the function $\log g$ by the following theorem.

Theorem 1.3.1. (Ledrappier [28]) Suppose that $g : \Sigma_A^+ \rightarrow \mathbb{R}$ is continuous, $g > 0$, and $\sum_{\sigma y=x} g(y) = 1$ for all $x \in \Sigma_A^+$. Then $L_{\log g}^* \mu = \mu$ if and only if μ is shift invariant and

$$\mathbb{E}_\mu \left(\chi_{[i]} \middle| \sigma^{-1} \mathcal{B} \right) (x) = g(i(\sigma x))$$

for all $i \in \Sigma$ and for μ almost every $x \in \Sigma_A^+$, where \mathcal{B} is the Borel σ -algebra.

Given a shift invariant measure μ we define the g function for μ to be

$$g(x) = \lim_{n \rightarrow \infty} \frac{\mu[x_0 x_1 \cdots x_n]}{\mu[x_1 \cdots x_n]}$$

where the limit exists for almost every x by the increasing martingale theorem.

1.3.2 Equilibrium and Gibbs states

Next, let's recall the related notations of equilibrium states and Gibbs states. For more details the standard reference is Bowen's book [6]. The notion of an equilibrium state is based on the principle that nature minimizes free energy. Given a continuous function $\varphi : \Sigma_A^+ \rightarrow \mathbb{R}$ (which we refer to as a *potential*), an equilibrium state is a shift invariant measure which maximizes the quantity

$$h_\mu(\sigma) + \int_{\Sigma_A^+} \varphi d\mu \tag{1.3}$$

where $h_\mu(\sigma)$ is the Kolmogorov-Sinai entropy. The maximum value of equation (1.3) is called the pressure of φ denoted $P(\varphi)$. Recall the following definitions.

Definition 1.3.2. We say that a function is *Hölder* if there exists a constant $|\varphi|_\theta$

and $0 < \theta < 1$ such that

$$\text{var}_n \varphi \leq |\varphi|_\theta \theta^n$$

for all $n \geq 0$ (that is φ is Hölder in the 2^{-n} metric). We say that a function is *Walters* if

$$\sup_{n \geq 1} \text{var}_{n+k} S_n \varphi \xrightarrow{k \rightarrow \infty} 0$$

where $S_n \varphi(x) = \sum_{i=0}^{n-1} \varphi(\sigma^i x)$. We say that a function is *Bowen* if φ is continuous and there exists a constant K such that

$$\sup_{n \geq 1} \text{var}_n S_n \varphi \leq K.$$

We will refer to these as the classical uniqueness regimes. It can be shown that

$$\text{Hölder} \subset \text{Walters} \subset \text{Bowen}.$$

Remark 1. Another common class of potentials are those potentials which have *summable variations*. That is, potentials for which

$$\sum_{n=0}^{\infty} \text{var}_n \varphi < \infty.$$

It can be shown that

$$\text{summable variations} \subset \text{Walters}$$

and in some sense summable variations is simply a verifiable condition which implies

the Walters property.

For all of these classes of potentials equilibrium states exist and are unique. Moreover they satisfy the Gibbs inequality (and are thus known as *Gibbs states*). That is, if Σ_A^+ is a topologically mixing shift of finite type and $\varphi : \Sigma_A^+ \rightarrow \mathbb{R}$ is Bowen then there exists a unique equilibrium state μ_φ and there exist constants $C > 0$ and P such that

$$C^{-1} \leq \frac{\mu_\varphi([x_0 \cdots x_{n-1}])}{e^{-nP + S_n \varphi(x)}} \leq C \quad (1.4)$$

for all $x \in \Sigma_A^+$ and $n > 0$ (it can be shown that $P = P(\varphi)$). These Gibbs states can be constructed in the following way. Take h and ν from the Ruelle-Perron-Frobenius Theorem A.2.2. Then the Gibbs state for φ , μ_φ , is the measure defined by $d\mu_\varphi = h d\nu$ and $P(\varphi) = \log \rho(L_\varphi)$. We will show how one can deduce ergodic properties from convergence properties of the Ruelle operator in Proposition 1.3.4.

1.3.3 Gibbs states for Hölder potentials on a full shift

Let $\varphi : \Sigma^{\mathbb{N}} \rightarrow \mathbb{R}$ be Hölder and assume without loss of generality that the pressure of φ is 0 (otherwise replace φ with $\varphi - P(\varphi)$). As the pressure of φ is 0 we have that $\rho(L_\varphi) = 1$. Take $h > 0$ and ν from the RPF Theorem A.2.1. The standard method for analyzing Gibbs states for Hölder continuous potentials on topologically mixing shifts of finite type can be found in Bowen's book [6]. Let's briefly sketch an alternative way of working with these Gibbs states which we will adapt to matrix equilibrium states in chapter 2. We work on a full shift for simplicity of exposition. This method is easily adapted to shifts of finite type.

Define for each $i \in \Sigma$ an operator $L_i : C(\Sigma^{\mathbb{N}}) \rightarrow C(\Sigma^{\mathbb{N}})$ by

$$L_i f(x) = e^{\varphi(ix)} f(ix) \quad (1.5)$$

where ix is the point $(ix)_0 = 1$ and $(ix)_k = x_{k-1}$ for $k \neq 0$.

Lemma 1.3.3. Let $d\mu_\varphi = hd\nu$ be the unique Gibbs state for φ . Then

$$\mu_\varphi[x_0x_1 \cdots x_{n-1}] = \left\langle L_{x_{n-1}} \cdots L_{x_1} L_{x_0} h, \nu \right\rangle.$$

Proof. Notice that for any word $x_0x_1 \cdots x_{n-1} \in \Sigma^n$

$$\begin{aligned} \mu_\varphi[x_0x_1 \cdots x_{n-1}] &= \int_{\Sigma^\mathbb{N}} \chi_{[x_0x_1 \cdots x_{n-1}]} h d\nu \\ &= \int_{\Sigma^\mathbb{N}} \chi_{[x_0x_1 \cdots x_{n-1}]} h d(L_\varphi^*)^n \nu \\ &= \int_{\Sigma^\mathbb{N}} L_\varphi^n (\chi_{[x_0x_1 \cdots x_{n-1}]} h) d\nu \\ &= \int_{\Sigma^\mathbb{N}} e^{S_n \varphi(x_0x_1 \cdots x_{n-1}z)} h(x_0x_1 \cdots x_{n-1}z) d\nu(z) \\ &= \int_{\Sigma^\mathbb{N}} L_{x_{n-1}} \cdots L_{x_1} L_{x_0} h d\nu. \end{aligned} \quad \square$$

Given a word $I = i_0i_1 \cdots i_{n-1}$ we write

$$L_I = L_{i_n} \cdots L_{i_1} L_{i_0}.$$

Notice that

$$\sum_{|I|=n} L_I = L_\varphi^n.$$

Proposition 1.3.4. The measure μ_φ is mixing. In particular there exist constants $C > 0$ and $0 < \gamma < 1$ such that

$$\left| \mu_\varphi([J] \cap \sigma^{-n-|J|}[I]) - \mu_\varphi([I])\mu_\varphi([J]) \right| \leq C\mu_\varphi([I])\mu_\varphi([J])\gamma^n. \quad (1.6)$$

for all cylinder sets $[I], [J]$.

Proof. Notice

$$\begin{aligned}
& \left| \mu_\varphi([J] \cap \sigma^{-n-|J|}[I]) - \mu_\varphi([I])\mu_\varphi([J]) \right| \\
&= \left| \sum_{|K|=n} \mu_\varphi([JKI]) - \mu_\varphi([I])\mu_\varphi([J]) \right| \\
&= \left| \sum_{|K|=n} \langle L_I L_K L_J h, \nu \rangle - \langle L_I, \nu \rangle \langle L_J h, \nu \rangle \right| \\
&= \left| \left\langle L_I \left(\sum_{|K|=n} L_K \right) L_J h, \nu \right\rangle - \langle L_I h, \nu \rangle \langle L_J h, \nu \rangle \right| \\
&= \left| \langle L_I L_\varphi^n L_J h, \nu \rangle - \langle L_I h, \nu \rangle \langle L_J h, \nu \rangle \right| \\
&= \left| \langle L_I (L_\varphi^n L_J h - \langle L_J h, \nu \rangle h), \nu \rangle \right| \\
&\leq \|L_I\|_{\text{op}} \left\| L_\varphi^n L_J h - \langle L_J h, \nu \rangle h \right\|_\infty \\
&\leq C \langle L_I h, \nu \rangle \left\| L_\varphi^n L_J h - \langle L_J h, \nu \rangle h \right\|_\infty \text{ in a similar way as equation (A.1)} \\
&\leq C \langle L_I h, \nu \rangle \langle L_J h, \nu \rangle \gamma^n \text{ by Theorem A.2.1} \\
&= C \mu_\varphi([I]) \mu_\varphi([J]) \gamma^n \text{ by Lemma 1.3.3} \quad \square
\end{aligned}$$

The inequality (1.6) implies in addition that the measure is weak Bernoulli and thus the natural extension of μ_φ is isomorphic to a Bernoulli shift. This approach is the one we will adapt to matrix thermodynamic formalism. In this case the natural transfer operator no longer acts on $C(\Sigma_A^+)$ but on $C(\mathbb{RP}^d)$, continuous functions on the projective space of \mathbb{R}^d . One can then take the equality in Lemma 1.3.3 as a definition of the Gibbs state and deduce ergodic properties via the same argument as Proposition 1.3.4. We will also use Lemma 1.3.3 to realize the measure of cylinder sets for factors of Gibbs states (that is $\pi_*\mu$ for some 1-block factor map π) in terms of positive operators. This will play an important role in chapter 3.

Chapter 2

Matrix Equilibrium States

This chapter is to appear in Ergodic Theory and Dynamical Systems [36].

2.1 Introduction

By analogy with the scalar thermodynamic formalism if $\mathcal{A} = (A_0, \dots, A_{M-1}) \in M_d(\mathbb{R})^M$ and $t > 0$ we say that a shift invariant measure $\mu_{\mathcal{A},t}$ is a *matrix Gibbs state* for (\mathcal{A}, t) , or a *t-Gibbs state* when \mathcal{A} is understood, provided there exists a constant $C > 0$ and P such that

$$C^{-1}\mu_{\mathcal{A},t}([x_0 \cdots x_{n-1}]) \leq e^{-nP} \left\| A_{x_0} \cdots A_{x_{n-1}} \right\|^t \leq C\mu_{\mathcal{A},t}([x_0 \cdots x_{n-1}]) \quad (2.1)$$

for all $x \in \Sigma^{\mathbb{Z}}$ ($\Sigma = \{0, \dots, M-1\}$) and $n > 0$. As all finite dimensional norms are equivalent one can take any choice of norm in (2.1). Notice we are working with the two-sided shift and not, as has been done in previous literature, the one-sided shift. Thus in a strict sense one may consider that we are working with the invertible extension of matrix Gibbs states. This is important when working on the isomorphism problem and it is also necessary so that we can apply the results in [9]. When $t = 1$

we refer to the measure simply as the Gibbs state for \mathcal{A} . A computation shows that

$$P = \lim_{n \rightarrow \infty} \frac{1}{n} \log \left(\sum_{x_0 \cdots x_{n-1}} \|A_{x_0} \cdots A_{x_{n-1}}\|^t \right).$$

In particular P is uniquely determined by (2.1) and is called the *pressure* denoted $P(\mathcal{A}, t)$. For the remainder of this chapter a Gibbs state will always refer to a matrix Gibbs state. Matrix Gibbs states are also equilibrium states for a sub-additive variational principle [10]

$$P(\mathcal{A}, t) = \sup_{\mu \in \mathcal{M}(\sigma)} [h(\mu) + t\Lambda(\mathcal{A}, \mu)]. \quad (2.2)$$

where $\Lambda(\mathcal{A}, \mu)$ is the *maximal Lyapunov exponent*

$$\Lambda(\mathcal{A}, \mu) = \lim_{n \rightarrow \infty} \frac{1}{n} \int \log \|A_{x_0} \cdots A_{x_{n-1}}\| d\mu(x).$$

Measures which achieve the supremum are called *matrix equilibrium states*. Such measures always exist by weak* compactness and upper semi-continuity of $h(\mu) + t\Lambda(\mathcal{A}, \mu)$. The connection between Gibbs states and equilibrium states for the variation principle (2.2) was studied in [18]. The study of these measures was originally motivated by their applications to dimension theory [19]. However recently interest has been shown in determining their ergodic properties [30] [31]. In the classical case for Hölder continuous functions, scalar Gibbs states are well known to have many nice statistical properties. It is natural to ask to what extent matrix Gibbs states share these properties.

One of the strongest of these properties is that the dynamical system defined by the shift map and a scalar Gibbs state for a Hölder potential is isomorphic to a Bernoulli shift and this is the problem we will focus on in this chapter. This is a par-

ticularly appealing property because Bernoulli shifts are classified up to isomorphism by their entropy [33]. In general it is very difficult to explicitly construct isomorphisms between measure preserving systems. One of the most common methods for demonstrating a measure preserving system is isomorphic to a Bernoulli shift is to show that it is weak Bernoulli and appeal to [20]. This is the strategy we will take in this paper. The same method has been used by Bowen [5] for scalar Gibbs states. Recall what it means for a dynamical system to be weak Bernoulli.

Definition 2.1.1. We say that partitions \mathcal{Q} and \mathcal{R} are ε -independent (written $\mathcal{Q} \perp^\varepsilon \mathcal{R}$) if

$$\sum_{q \in \mathcal{Q}, r \in \mathcal{R}} |\mu(q \cap r) - \mu(q)\mu(r)| < \varepsilon.$$

We say that a partition \mathcal{P} is *weak Bernoulli* if for every $\varepsilon > 0$ there exists N such that $\bigvee_{i=0}^{s-1} \sigma^{-i} \mathcal{P} \perp^\varepsilon \bigvee_{i=t}^{t+r-1} \sigma^{-i} \mathcal{P}$ for all $r, s \geq 0$ and $t \geq s + N$. We say that $\mu_{\mathcal{A}, t}$ is *weak Bernoulli* if the standard partition $\mathcal{P} = \{[i] : 0 \leq i \leq M - 1\}$ is weak Bernoulli.

For a word $I = i_0 i_1 \cdots i_{n-1}$ we write

$$A_I = A_{i_0} A_{i_1} \cdots A_{i_{n-1}}$$

and we denote the length of the word I by $|I|$. We say that $\mathcal{A} = (A_0, \dots, A_{M-1}) \in M_d(\mathbb{R})^M$ is *irreducible* if the matrices have no common proper and non-trivial invariant subspace. This implies that there exists a constant $\delta > 0$ such that

$$\sum_{|K| \leq d} \|A_I A_K A_J\| \geq \delta \|A_I\| \|A_J\| \tag{2.3}$$

for all I, J , see for instance [30, lemma 12]. With this in mind we make the following definition

Definition 2.1.2. We say that $\mathcal{A} = (A_0, \dots, A_{M-1})$ is *primitive* if there exists an N and a $\delta > 0$ such that

$$\sum_{|K|=N} \|A_I A_K A_J\| \geq \delta \|A_I\| \|A_J\| \quad (2.4)$$

for all I, J .

For irreducible (and for primitive) collections of matrices, matrix Gibbs states are known to exist and be unique [17, theorem 5.5] for all $t > 0$. The terms irreducible and primitive are familiar from Perron-Frobenius theory and indeed the notions are connected. Let $L_{\mathcal{A}} : M_d(\mathbb{R}) \rightarrow M_d(\mathbb{R})$ be defined by $L_{\mathcal{A}}B = \sum_i A_i^* B A_i$, then $L_{\mathcal{A}}$ preserves the cone of positive semi-definite matrices. The operator $L_{\mathcal{A}}$ appears in connection with a class of measures related to fractal geometry called Kusuoka measures [27] (see example 2.2.3). One can check that if $L_{\mathcal{A}}$ is irreducible (respectively primitive) in the sense of Perron-Frobenius theory then \mathcal{A} satisfies equation (2.3) (respectively equation (2.4)). For the details see proposition A.3.6. Our main theorem is the following.

Theorem 2.1.3. Suppose that $\mathcal{A} = (A_0, \dots, A_{M-1})$ is primitive. Then for any $t > 0$ the unique t -Gibbs state for \mathcal{A} is weak Bernoulli.

The proof of theorem 2.1.3 can be found in section 2.4. The proof relies on a general result of Bradley [9], which is somewhat opaque. With this in mind we also present a method for understanding matrix Gibbs states through transfer operators which is interesting in its own right. Understanding the ergodic/statistical properties of Gibbs states in sub-additive thermodynamic formalism has long been a challenge, with most results being achieved using fairly ad-hoc methods. This is in contrast to the case for scalar Gibbs states which has a well developed methodology for deducing ergodic/statistical properties relying on the transfer operator. In this chapter we

adapt the classical doctrine of transfer operators for scalar Gibbs states to matrix Gibbs states.

In section 2.2 we show that in the case when t is an even integer the ergodic properties of $\mu_{A,t}$ can be readily understood by studying the convergence properties of iterates of a matrix. As a consequence we can obtain an exponential mixing result which includes an explicit rate determined by the spectral gap of a finite dimensional matrix. This naturally leads to the problem of generalizing this approach to $t > 0$. In section 2.3 we generalize section 2.2 using operators on a suitable infinite dimensional vector space. A major advantage of the approach in sections 2.2 and 2.3 is that we can give an explicit construction of certain Gibbs states, including a formula for the measure of a cylinder set. Previous methods have relied on abstract compactness arguments, realizing the Gibbs state as a weak* limit point of a sequence of measures. As many properties are not preserved under weak* limits this makes an analysis of the Gibbs state difficult. Our transfer operator approach allows us to give direct proofs of ergodic properties. It also provides a strong intuition for understanding how properties of the collection \mathcal{A} are reflected in the ergodic properties of $\mu_{A,t}$.

2.2 Matrices which preserve a common cone

One particular class of matrix Gibbs states has appeared extensively in applications. Consider the following examples.

Example 2.2.1. Bernoulli measures, take $d = 1$.

Example 2.2.2. Factors of Markov measures. The 1-Gibbs states for collections of non-negative matrices are precisely factors of Markov measures; for details see [7] or [11], [45]. In fact, allowing the operators in \mathcal{A} to act on an infinite dimensional space, factors of Gibbs states for Hölder potentials can be viewed as Gibbs states for

a suitable collection of operators, see [35].

Example 2.2.3. The Kusuoka measure [27] was originally studied because of its connections to fractal geometry. We briefly recall the construction. Let $L_i B = A_i^* B A_i$ and $L_{\mathcal{A}} = \sum_i L_i$. When \mathcal{A} is irreducible there exist U, V positive definite matrices such that $L_{\mathcal{A}} U = \rho(L_{\mathcal{A}}) U$, $L_{\mathcal{A}}^* V = \rho(L_{\mathcal{A}}) V$ (notice that $L_{\mathcal{A}}^* B = \sum_i A_i B A_i^*$) and $\langle U, V \rangle_{\text{HS}} = 1$ (where $\langle A, B \rangle_{\text{HS}} = \text{tr}(A^* B)$ is the Hilbert-Schmidt inner product). The Kusuoka measure is then obtained by extending

$$\mu[x_0 \cdots x_{n-1}] = \rho(L_{\mathcal{A}})^{-n} \langle L_{x_0} L_{x_1} \cdots L_{x_{n-1}} U, V \rangle_{\text{HS}}$$

to a measure using Carathéodory's extension theorem. It was shown in [30] that the Kusuoka measure is a 2-Gibbs state. We will generalize this result to k -Gibbs states for k even in example 2.2.7. Observe that thinking of the linear maps L_i as matrices we have that the Kusuoka measure is the 1-Gibbs state for the collection $\hat{\mathcal{A}} = (L_0, \dots, L_{M-1})$ each of which preserves the cone of positive semi-definite matrices.

The property shared by all of these matrix equilibrium states is that all of the matrices preserve a common cone. Our goal for this section is then to treat these measures in an abstract manner. As one of the applications of this section is the Kusuoka measure, we work with matrices preserving an abstract cone K . For the most part, the reader will lose no intuition by simply thinking of K as being the non-negative orthant of \mathbb{R}^d . For the reader's convenience we have collected some definitions and facts about abstract cones in finite dimensional vector spaces in the appendix. For $\theta \in (0, 1)$ define

$$\mathcal{H}_{\theta} = \left\{ f \in C(\Sigma^{\mathbb{Z}}) : \text{There exists a constant } K > 0 \text{ for which } \text{var}_n f \leq K\theta^n \right\}.$$

We denote the least such constant by $|f|_{\theta}$ and \mathcal{H}_{θ} becomes a Banach space with norm

$\|f\|_\theta = \|f\|_\infty + |f|_\theta$. The goal of this section is to prove the following theorem.

Theorem 2.2.4. Let $\mathcal{A} = (A_0, \dots, A_{M-1}) \in M_d(\mathbb{R})^M$. Suppose that each A_i is non-negative with respect to a cone K and $A = \sum_i A_i$ is such that $\sum_{k=0}^{d-1} A^k$ maps $K \setminus \{0\}$ into the interior of K (that is, A is K -irreducible). Then there exists a 1-Gibbs state for \mathcal{A} denoted $\mu_{\mathcal{A}}$. Moreover:

1. $\mu_{\mathcal{A}}$ is ergodic and thus unique, and $P(\mathcal{A}, 1) = \log \rho(A)$.
2. If there exists an N such that A^N maps $K \setminus \{0\}$ into the interior of K (that is, A is K -primitive) then:
 - (a) $\mu_{\mathcal{A}}$ is weak Bernoulli.
 - (b) $\mu_{\mathcal{A}}$ has exponential decay of correlations for Hölder continuous functions.

That is for a fixed $\theta \in (0, 1)$ there are constants D and $\gamma \in (0, 1)$ such that

$$\left| \int f \cdot g \circ \sigma^n d\mu_{\mathcal{A}} - \int f d\mu_{\mathcal{A}} \int g d\mu_{\mathcal{A}} \right| \leq D \|f\|_\theta \|g\|_\theta \gamma^n$$

for all $f, g \in \mathcal{H}_\theta$, $n \geq 0$. In addition, the rate γ is determined by θ and the eigenvalues of A .

For the Kusuoka measure, part 2(b) is known [23]. However our proof is fundamentally different and significantly more elementary. In particular the method in [23] uses the g -function for the Kusuoka measure and transfer operator techniques. This is technically challenging largely due to the fact that the g -function can fail to be continuous.

We can explicitly construct the measure $\mu_{\mathcal{A}}$. As A is irreducible we may take u, v to be right and left eigenvectors respectively corresponding to the spectral radius

$\rho(A)$ with $\langle u, v \rangle = 1$. On cylinder sets we define

$$\mu_{\mathcal{A}}[x_0 x_1 \cdots x_{n-1}] = \rho(A)^{-n} \langle A_{x_0} A_{x_1} \cdots A_{x_{n-1}} u, v \rangle. \quad (2.5)$$

Using the fact that u, v are eigenvectors for A it is readily checked that

$$\sum_i \mu_{\mathcal{A}}[i x_0 \cdots x_{n-1}] = \mu_{\mathcal{A}}[x_0 \cdots x_{n-1}] = \sum_i \mu_{\mathcal{A}}[x_0 \cdots x_{n-1} i].$$

Proposition 1.2.6 implies that this extends to a shift invariant measure on $\Sigma^{\mathbb{Z}}$. Next our goal is to show that this is a 1-Gibbs state for \mathcal{A} and that it is unique. To do so, we prove the following proposition.

Proposition 2.2.5. Suppose that $\mathcal{A} = (A_0, \dots, A_{M-1}) \in M_d(\mathbb{R})^M$ is such that each A_i is non-negative with respect to a cone K and $A = \sum_i A_i$ is K -irreducible. Then

1. $\mu_{\mathcal{A}}$ is ergodic.
2. $\mu_{\mathcal{A}}$ satisfies the Gibbs inequality (2.1) with $P = \log \rho(A)$.

Proof. 1. Observe that

$$A^n = \left(\sum_i A_i \right)^n = \sum_{|K|=n} A_K. \quad (2.6)$$

Let I, J be words.

$$\begin{aligned} & \left| \frac{1}{n} \sum_{k=1}^n \mu_{\mathcal{A}}([I] \cap \sigma^{-k}[J]) - \mu_{\mathcal{A}}([I]) \mu_{\mathcal{A}}([J]) \right| \\ & \leq \left| \frac{1}{n} \sum_{k=1}^{|I|} \mu_{\mathcal{A}}([I] \cap \sigma^{-k}[J]) \right| \\ & \quad + \rho(A)^{-|I|-|J|} \left| \left\langle A_I \left(\frac{1}{n} \sum_{k=|I|+1}^n \rho(A)^{|I|-k} A^{k-|I|} \right) A_J u, v \right\rangle - \langle A_I u, v \rangle \langle A_J u, v \rangle \right| \\ & \xrightarrow{n \rightarrow \infty} 0 + \rho(A)^{-|I|-|J|} |\langle A_I \langle A_J u, v \rangle u, v \rangle - \langle A_I u, v \rangle \langle A_J u, v \rangle| = 0 \end{aligned}$$

by the Perron-Frobenius theorem A.3.4 2(b). As cylinder sets are a generating semi-algebra this implies by Proposition 1.2.7 that $\mu_{\mathcal{A}}$ is ergodic.

2. From the Perron-Frobenius theorem we have that $u \in \text{int}(K)$, $v \in \text{int}(K^*)$.

Thus the Gibbs inequality follows directly from an application of lemma A.3.5.

□

As ergodic measures are mutually singular this implies that $\mu_{\mathcal{A}}$ is the unique 1-Gibbs state for \mathcal{A} . The proof of the previous lemma shows that mixing properties of $\mu_{\mathcal{A}}$ are related to the convergence of A^n . It is this fact that we will exploit to prove the remaining assertions in theorem 2.2.4.

Proposition 2.2.6. Suppose that $\mathcal{A} = (A_0, \dots, A_{M-1}) \in M_d(\mathbb{R})^M$ is such that each A_i is non-negative with respect to a cone K and $A = \sum_i A_i$ if A is K -primitive then the measure $\mu_{\mathcal{A}}$ is weak Bernoulli.

Proof. Let $r, s \geq 1$, $t \geq s$ and take $[I] \in \mathcal{V}_{i=0}^{s-1} \sigma^{-i} \mathcal{P}$ and $[{}_t J] \in \mathcal{V}_{i=t}^{t+r-1} \sigma^{-i} \mathcal{P}$. Notice

$$\begin{aligned} & |\mu_{\mathcal{A}}([I] \cap [{}_t J]) - \mu_{\mathcal{A}}([I])\mu_{\mathcal{A}}([{}_t J])| \\ &= \left| \sum_{|K|=t-s} \mu_{\mathcal{A}}([IKJ]) - \mu_{\mathcal{A}}([I])\mu_{\mathcal{A}}([{}_t J]) \right| \\ &= \left| \sum_{|K|=t-s} \rho(A)^{-(s+r+(t-s))} \langle A_I A_K A_J u, v \rangle - \rho(A)^{-(s+r)} \langle A_I u, v \rangle \langle A_J u, v \rangle \right| \\ &= \rho(A)^{-(s+r)} \left| \left\langle A_I \left(\rho(A)^{-(t-s)} \sum_{|K|=t-s} A_K \right) A_J u, v \right\rangle - \langle A_I u, v \rangle \langle A_J u, v \rangle \right| \end{aligned}$$

Notice that

$$\rho(A)^{-(t-s)} \sum_{|K|=t-s} A_K = \rho(A)^{-(t-s)} A^{t-s} = uv^T + (\rho(A)^{-(t-s)} A^{t-s} - uv^T).$$

Thus

$$\begin{aligned}
& |\mu_{\mathcal{A}}([I] \cap [{}_t J]) - \mu_{\mathcal{A}}([I])\mu_{\mathcal{A}}([J])| \\
&= \rho(A)^{-(s+r)} \left| \left\langle A_I(\rho(A)^{-(t-s)}A^{t-s} - uv^T)A_J u, v \right\rangle \right| \\
&\leq \rho(A)^{-(s+r)} \|A_I^* v\| \|A_J u\| \left\| \rho(A)^{-(t-s)}A^{t-s} - uv^T \right\| \\
&\leq C\beta^{t-s} \rho(A)^{-s} \|A_I\| \rho(A)^{-r} \|A_J\| \\
&\leq C'\beta^{t-s} \mu_{\mathcal{A}}(I)\mu_{\mathcal{A}}(J) \text{ by Proposition 2.2.5}
\end{aligned}$$

where $\beta = \frac{|\lambda_2| + \varepsilon}{\rho(A)} < 1$ for a small $\varepsilon > 0$ as in Perron-Frobenius theorem A.3.4. Then we have

$$\sum_{I,J} |\mu_{\mathcal{A}}([I] \cap [{}_t J]) - \mu_{\mathcal{A}}([I])\mu_{\mathcal{A}}([J])| \leq K\beta^{t-s} \sum_{I,J} \mu_{\mathcal{A}}([I])\mu_{\mathcal{A}}([J]) = K\beta^{t-s}.$$

Hence $\mu_{\mathcal{A}}$ is weak Bernoulli. □

Thus we have proven theorem 2.2.4 2(a); part 2(b) follows by an approximation argument, see Bowen's book [6, theorem 1.26]. Finally we end this section with an example which shows that k -Gibbs states can be understood in terms of matrices preserving a common cone, for k an even integer.

Example 2.2.7. The following example generalizes the Kusuoka measure (the Kusuoka measure is the case of $k = 2$). Let k be an even integer and define

$$S = \text{span} \{v^{\otimes k} : v \in \mathbb{R}^d\}$$

We consider the following cone in S^*

$$K = \left\{ w \in S^* : \left\langle v^{\otimes k}, w \right\rangle_{(\mathbb{R}^d)^{\otimes k}} \geq 0 \text{ for all } v \in \mathbb{R}^d \right\}$$

Note that when k is odd this set is $\{0\}$. When k is even, K is a cone with non-void interior (see proposition A.3.7). The cone K is sometimes referred to as the positive semi-definite tensor cone: in the case of $k = 2$ this cone can be identified with positive semi-definite matrices. Suppose that $\mathcal{A} = (A_0, \dots, A_{M-1})$ is a collection of matrices with no common proper, non-trivial invariant subspace. Consider the collection $\mathcal{A}' = ((A_0^{\otimes k})^*, \dots, (A_{M-1}^{\otimes k})^*)$. The collection \mathcal{A}' preserves the cone K . We claim that in fact $A = \sum_i (A_i^{\otimes k})^*$ is irreducible with respect to K . To prove this it is enough to show that no eigenvector of A lies on the boundary of K [40, theorem 4.1]. Suppose that $w \in K$, $w \neq 0$ and that $Aw = \lambda w$ and define

$$W = \text{span} \left\{ u : \langle u^{\otimes k}, w \rangle_{(\mathbb{R}^d)^{\otimes k}} = 0 \right\}$$

We claim that W is invariant under \mathcal{A} . If $\langle u^{\otimes k}, w \rangle_{(\mathbb{R}^d)^{\otimes k}} = 0$ then

$$0 = \langle u^{\otimes k}, Aw \rangle_{(\mathbb{R}^d)^{\otimes k}} = \sum_i \langle (A_i u)^{\otimes k}, w \rangle_{(\mathbb{R}^d)^{\otimes k}}$$

as $w \in K$ this implies that $\langle (A_i u)^{\otimes k}, w \rangle_{(\mathbb{R}^d)^{\otimes k}} = 0$ for each i . Thus W is \mathcal{A} invariant, so it is either \mathbb{R}^d or $\{0\}$. As $w \neq 0$ we must have that $W = \{0\}$. Therefore $w \in \text{int}(K)$ by lemma A.3.2 and A is irreducible. Constructing the 1-Gibbs state for \mathcal{A}' , we see that it satisfies the Gibbs inequality: there exist constants $C > 0$ and P such that

$$C^{-1} \mu_{\mathcal{A}'}([x_0 \cdots x_{n-1}]) \leq e^{-nP} \left\| (A_{x_0}^{\otimes k})^* (A_{x_1}^{\otimes k})^* \cdots (A_{x_{n-1}}^{\otimes k})^* \right\| \leq C \mu_{\mathcal{A}'}([x_0 \cdots x_{n-1}]).$$

As $A_{x_{n-1}}^{\otimes k} A_{x_{n-2}}^{\otimes k} \cdots A_{x_0}^{\otimes k} = (A_{x_{n-1}} A_{x_{n-2}} \cdots A_{x_0})^{\otimes k}$ we have that

$$C^{-1} \mu_{\mathcal{A}'}([x_0 \cdots x_{n-1}]) \leq e^{-nP} \left\| A_{x_{n-1}} A_{x_{n-2}} \cdots A_{x_0} \right\|^k \leq C \mu_{\mathcal{A}'}([x_0 \cdots x_{n-1}]).$$

Strictly speaking the order of the product of matrices is backwards from the Gibbs inequality in equation (2.1). By taking $\mathcal{A} = (A_0^*, \dots, A_{M-1}^*)$ this can be changed (see proposition A.3.8). Thus we have found an elementary way of constructing k -Gibbs states for all even integers.

2.3 Transfer operators and exponential mixing

The goal of this section is to explore a method for constructing matrix Gibbs states and proving ergodic and statistical properties using transfer operators. This approach is interesting for number of reasons. In particular it is an application of transfer operator methods to a problem in sub-additive ergodic theory. It is also a reasonable generalization of Example 2.2.7 using operators on infinite dimensional spaces. We will need the following definitions.

Definition 2.3.1. We say that a collection of invertible $d \times d$ matrices (A_0, \dots, A_{M-1}) is *strongly irreducible* if they do not preserve a finite union of proper and nontrivial subspaces.

Definition 2.3.2. An element $B \in M_d(\mathbb{R})$ is called *proximal* if B has a simple eigenvalue of modulus $\rho(B)$ and any other eigenvalue has modulus strictly smaller than $\rho(B)$. The collection (A_0, \dots, A_{M-1}) is called *proximal* if there exists a product $B = A_{x_0} \cdots A_{x_n}$ that is proximal.

We have the following theorem.

Theorem 2.3.3. Suppose that $\mathcal{A} = (A_0, \dots, A_{M-1})$ is a collection of real invertible $d \times d$ matrices which is proximal and strongly irreducible. Then for any $t \geq 0$ there exists a unique Gibbs state for (\mathcal{A}, t) , denoted $\mu_{\mathcal{A}, t}$. Moreover

1. $\mu_{\mathcal{A}, t}$ is weak Bernoulli.

2. $\mu_{A,t}$ has exponential decay of correlations for Hölder continuous functions. That is, for a fixed $\theta \in (0, 1)$ there are constants D and $\gamma \in (0, 1)$ such that

$$\left| \int f \cdot g \circ \sigma^n d\mu_{A,t} - \int f d\mu_{A,t} \int g d\mu_{A,t} \right| \leq D \|f\|_\theta \|g\|_\theta \gamma^n$$

for all $f, g \in \mathcal{H}_\theta$, $n \geq 0$.

In the previous section we have seen that the role of the transfer operator for $t = 2k$ was played by $A = \sum_i A_i^{\otimes 2k}$ we need to find a suitable replacement. By identifying 2-tensors with bilinear forms which are in turn a subspace of the 2-homogeneous functions one is naturally led to consider the action of the matrices on t -homogeneous functions. This is then equivalent to the action of the matrices on the projective space \mathbb{RP}^{d-1} weighted by the functions $\left\| A_i \frac{u}{\|u\|} \right\|^t$. That is, define a transfer operator by

$$L_t f(\bar{u}) = \sum_{i=0}^{M-1} \left\| A_i \frac{u}{\|u\|} \right\|^t f(\overline{A_i u}) \quad (2.7)$$

which acts on $C(\mathbb{RP}^{d-1})$. The connection between matrix Gibbs states and this operator is made clear in proposition 2.3.4. First we fix some notation. For a function h and a measure ν we write

$$\langle h, \nu \rangle = \int h d\nu.$$

Recall that \mathbb{RP}^{d-1} is obtained by taking the quotient of $\mathbb{R}^d \setminus \{0\}$ by the equivalence relation $x \sim y$ if and only if $x = \lambda y$ for some $\lambda \neq 0$. We denote the equivalence class of a vector v by \bar{v} . Define a metric on \mathbb{RP}^{d-1} by

$$d(\bar{u}, \bar{w}) = \inf \left\{ \|u' - w'\| : \|u'\| = \|w'\| = 1 \text{ and } \bar{u'} = \bar{u}, \bar{w'} = \bar{w} \right\}.$$

Proposition 2.3.4. Let $t \geq 0$ and $\mathcal{A} = (A_0, \dots, A_{M-1})$ be a collection of invertible matrices. Suppose that there exists ν_t a Borel probability measure not supported on a projective proper subspace and h_t a strictly positive continuous function such that $L_t h_t = \rho(L_t) h_t$, $L_t^* \nu_t = \rho(L_t) \nu_t$ and $\langle h_t, \nu_t \rangle = 1$. Define L_i by

$$L_i f(\bar{u}) = \left\| A_i \frac{u}{\|u\|} \right\|^t f(\overline{A_i u}).$$

Then the formula

$$\mu_{\mathcal{A}, t}[x_0 x_1 \cdots x_{n-1}] = \rho(L_t)^{-n} \int_{\mathbb{R}P^{d-1}} L_{x_{n-1}} \cdots L_{x_1} L_{x_0} h_t(\bar{u}) d\nu_t(\bar{u}) \quad (2.8)$$

extends to a shift invariant measure on $\Sigma^{\mathbb{Z}}$. Moreover $\mu_{\mathcal{A}, t}$ is a Gibbs state for (\mathcal{A}, t) .

Proof. The assumption that h_t, ν_t are eigenvectors corresponding to $\rho(L_t)$ implies that the formula in (2.8) extends to a shift invariant measure by proposition 1.2.6. All that remains to be shown is that $\mu_{\mathcal{A}, t}$ satisfies the Gibbs inequality. To see why the Gibbs inequality holds notice that the function

$$A \mapsto \int_{\mathbb{R}P^{d-1}} \left\| A \frac{u}{\|u\|} \right\|^t d\nu_t(\bar{u})$$

from the set of norm one $d \times d$ matrices to \mathbb{R} is continuous and strictly positive (by the assumption that ν_t is not supported on a projective proper subspace). Take $C > 0$ such that

$$\int_{\mathbb{R}P^{d-1}} \left\| A \frac{u}{\|u\|} \right\|^t d\nu_t(\bar{u}) \geq C \|A\|^t$$

for all $A \in M_d(\mathbb{R})$. Thus

$$\rho(L)^{-n} \langle L_{x_{n-1}} \cdots L_{x_1} L_{x_0} h_t, \nu_t \rangle \geq (\inf h_t) C \rho(L)^{-n} \|A_{x_0} A_{x_1} \cdots A_{x_{n-1}}\|^t$$

and

$$\rho(L)^{-n} \langle L_{x_{n-1}} \cdots L_{x_1} L_{x_0} h_t, \nu_t \rangle \leq (\sup h_t) \rho(L)^{-n} \|A_{x_0} A_{x_1} \cdots A_{x_{n-1}}\|^t.$$

Which shows that the measure $\mu_{A,t}$ satisfies the Gibbs inequality. \square

If $I = i_0 i_1 \cdots i_{n-1}$ we will use the notation that

$$L_I = L_{i_{n-1}} \cdots L_{i_1} L_{i_0}.$$

Notice that this is backward from the definition of A_I . To see why consider

$$\begin{aligned} L_{x_1} L_{x_0} f(\bar{u}) &= \left\| A_{x_1} \frac{u}{\|u\|} \right\|^t L_{x_0} f(\overline{A_{x_1} u}) \\ &= \left\| A_{x_1} \frac{u}{\|u\|} \right\|^t \left\| A_{x_0} \frac{A_{x_1} u}{\|A_{x_1} u\|} \right\|^t f(\overline{A_{x_0} A_{x_1} u}) \\ &= \left\| A_{x_0} A_{x_1} \frac{u}{\|u\|} \right\|^t f(\overline{A_{x_0} A_{x_1} u}). \end{aligned}$$

As we can see pre-composition reverses the order of the products.

Operators like L_t have appeared frequently in the study of random matrix products. This is however the first time they have been used to construct a measure on $\Sigma^{\mathbb{Z}}$ and deduce ergodic and statistical properties. To prove theorem 2.3.3 all we require is a suitable Perron-Frobenius theorem. For each $\varepsilon > 0$ denote by $C^\varepsilon(\mathbb{RP}^{d-1})$ the space of ε -Hölder continuous functions in the d metric on \mathbb{RP}^{d-1} . This becomes a Banach space in the usual way with norm $\|\cdot\|_\varepsilon = \|\cdot\|_\infty + |\cdot|_\varepsilon$ (where $|f|_\varepsilon$ is the least ε -Hölder constant for f). Set $\bar{t} = \min\{1, t\}$. The following theorem is a result of Guivarc'h and Le Page [21].

Theorem 2.3.5 (Guivarc'h and Le Page [21]). Let $t > 0$. Suppose that (A_0, \dots, A_{M-1}) are real, invertible, strongly irreducible and proximal. Then there exists an ε with $0 < \varepsilon \leq \bar{t}$ such that the following hold

1. $L_t : C^\varepsilon(\mathbb{RP}^{d-1}) \rightarrow C^\varepsilon(\mathbb{RP}^{d-1})$, that is L_t preserves the space of ε -Hölder functions.
2. The spectral radius of $L_t : C^\varepsilon(\mathbb{RP}^{d-1}) \rightarrow C^\varepsilon(\mathbb{RP}^{d-1})$ is equal to $e^{P(\mathcal{A}, t)}$. That is

$$\log \rho(L_t) = \lim_{n \rightarrow \infty} \frac{1}{n} \log \left(\sum_{|I|=n} \|A_I\|^t \right) = P(\mathcal{A}, t).$$

3. There exists a unique Borel probability measure ν_t on \mathbb{RP}^{d-1} , not supported on a projective subspace, such that $L_t^* \nu_t = \rho(L_t) \nu_t$.
4. There exists a unique \bar{t} -Hölder function $h_t : \mathbb{RP}^{d-1} \rightarrow (0, \infty)$ such that $L_t h_t = \rho(L_t) h_t$ and $\langle h_t, \nu_t \rangle = 1$.
5. The operator L_t has a spectral gap on $C^\varepsilon(\mathbb{RP}^{d-1})$. That is to say there exists a decomposition of L_t as $L_t = \rho(L_t)(P_t + R_t)$ where $\rho(R_t) < 1$, $P_t R_t = R_t P_t = 0$ and

$$P_t f = \langle f, \nu_t \rangle h_t \text{ for all } f \in C^\varepsilon(\mathbb{RP}^{d-1}).$$

Proof. If we take the measure on $GL_d(\mathbb{R})$ to be $\mu = \frac{1}{M} \sum_{i=0}^{M-1} \delta_{A_i}$ then the operator called P^t in [21] is a scalar multiple of L_t and the result follows from [21, Theorem 8.8]. That h_t is \bar{t} -Hölder is [21, lemma 4.8]. \square

Corollary 2.3.6. Under the assumptions of Theorem 2.3.5 there exist constants $C > 0$ and β with $0 < \beta < 1$ such that for any $f \in C^\varepsilon(\mathbb{RP}^{d-1})$ we have

$$\left\| \rho(L_t)^{-n} L_t^n f - \langle f, \nu_t \rangle h_t \right\|_\varepsilon \leq C \|f\|_\varepsilon \beta^n$$

for all $n \geq 0$.

Proof. Since $P_t^2 = P_t$ and $P_t R_t = R_t P_t = 0$ we have that $\rho(L_t)^{-n} L_t^{-n} = P_t + R_t^n$. Thus

$$\left\| \rho(L_t)^{-n} L_t^n f - \langle f, \nu_t \rangle h_t \right\|_\varepsilon = \|R_t^n f\|_\varepsilon \leq \|R_t^n\|_{\varepsilon, \text{op}} \|f\|_\varepsilon.$$

Taking $\beta = \rho(R_t) + \eta < 1$ for a small $\eta > 0$ we have the result. \square

In order to obtain decay of correlation results we are thus forced into controlling the regularity of $L_J h_t$. This is the content of the next lemma.

Lemma 2.3.7. 1. For any $A \in GL_d(\mathbb{R})$ we have that

$$d(\overline{Au}, \overline{Aw}) \leq \frac{2 \|A\|}{\left\| A \frac{u}{\|u\|} \right\|} d(\overline{u}, \overline{w}).$$

for all $u, w \in \mathbb{R}^d$.

2. For any $A \in GL_d(\mathbb{R})$ and $t \geq 0$ we have that

$$\left| \left\| A \frac{u}{\|u\|} \right\|^t - \left\| A \frac{w}{\|w\|} \right\|^t \right| \leq (t+1) \|A\|^t d(\overline{u}, \overline{w})^{\bar{t}}$$

for all $u, w \in \mathbb{R}^d \setminus \{0\}$.

3. For any $0 < \varepsilon \leq \bar{t}$ there exists a constant K such that $\|L_J h_t\|_\varepsilon \leq K \|A_J\|^t$ for all J .

Proof. 1. This is essentially [21, Lemma 4.6]. We provide the details for the sake of completeness. Notice for any u, w

$$\begin{aligned} \|Au\| \|Aw\| \left(\frac{Au}{\|Au\|} - \frac{Aw}{\|Aw\|} \right) &= \|Aw\| Au - \|Au\| Aw \\ &= \|Aw\| Au - \|Aw\| Aw + \|Aw\| Aw - \|Au\| Aw \\ &= \|Aw\| (Au - Aw) + (\|Aw\| - \|Au\|)Aw. \end{aligned}$$

By taking the norm of both sides we have that

$$\|Au\| \|Aw\| \left\| \frac{Au}{\|Au\|} - \frac{Aw}{\|Aw\|} \right\| \leq 2 \|Aw\| \|A(u - w)\|.$$

Thus

$$\begin{aligned} d(\overline{Au}, \overline{Aw}) &\leq \left\| \frac{A \frac{u}{\|u\|}}{\|A \frac{u}{\|u\|}\|} - \frac{A \frac{w}{\|w\|}}{\|A \frac{w}{\|w\|}\|} \right\| \\ &\leq \frac{2}{\|A \frac{u}{\|u\|}\|} \left\| A \left(\frac{u}{\|u\|} - \frac{w}{\|w\|} \right) \right\| \\ &\leq \frac{2 \|A\|}{\|A \frac{u}{\|u\|}\|} \left\| \frac{u}{\|u\|} - \frac{w}{\|w\|} \right\|. \end{aligned}$$

The same argument holds for $\left\| \frac{-u}{\|u\|} - \frac{w}{\|w\|} \right\|$. Hence the result.

2. This is [21, lemma 4.6].

3. Notice

$$\begin{aligned} &|L_J h(\bar{u}) - L_J h(\bar{w})| \\ &= \left| \left\| A_J \frac{u}{\|u\|} \right\|^t h_t(\overline{A_J u}) - \left\| A_J \frac{w}{\|w\|} \right\|^t h_t(\overline{A_J w}) \right| \end{aligned}$$

$$\begin{aligned}
&\leq \left\| A_J \frac{u}{\|u\|} \right\|^t |h_t(\overline{A_J u}) - h_t(\overline{A_J w})| + \|h_t\|_\infty \left| \left\| A_J \frac{u}{\|u\|} \right\|^t - \left\| A_J \frac{w}{\|w\|} \right\|^t \right| \\
&\leq \left\| A_J \frac{u}{\|u\|} \right\|^t |h_t|_{\bar{t}} d(\overline{A_J u}, \overline{A_J w})^{\bar{t}} + \|h_t\|_\infty (t+1) \|A_J\|^t d(\bar{u}, \bar{w})^{\bar{t}} \\
&\leq \left\| A_J \frac{u}{\|u\|} \right\|^t |h_t|_{\bar{t}} \left(\frac{2 \|A_J\|}{\|A_J \frac{u}{\|u\|}\|} \right)^{\bar{t}} d(\bar{u}, \bar{w})^{\bar{t}} + \|h_t\|_\infty (t+1) \|A_J\|^t d(\bar{u}, \bar{w})^{\bar{t}} \\
&= \left\| A_J \frac{u}{\|u\|} \right\|^{t-\bar{t}} \|A_J\|^{\bar{t}} |h_t|_{\bar{t}} 2^{\bar{t}} d(\bar{u}, \bar{w})^{\bar{t}} + \|h_t\|_\infty (t+1) \|A_J\|^t d(\bar{u}, \bar{w})^{\bar{t}} \\
&\leq \|A_J\|^t |h_t|_{\bar{t}} 2^{\bar{t}} d(\bar{u}, \bar{w})^{\bar{t}} + \|h_t\|_\infty (t+1) \|A_J\|^t d(\bar{u}, \bar{w})^{\bar{t}} \\
&= \left[|h_t|_{\bar{t}} 2^{\bar{t}} + \|h_t\|_\infty (t+1) \right] \|A_J\|^t d(\bar{u}, \bar{w})^{\bar{t}}.
\end{aligned}$$

Thus for $0 < \varepsilon \leq \bar{t}$ we have

$$|L_J h_t|_\varepsilon \leq 2^{\bar{t}-\varepsilon} |L_J h_t|_{\bar{t}} \leq \|A_J\|^t 2^{\bar{t}-\varepsilon} \left[|h_t|_{\bar{t}} 2^{\bar{t}} + \|h_t\|_\infty (t+1) \right].$$

Therefore

$$\|L_J h_t\|_\varepsilon = \|L_J h_t\|_\infty + |L_J h_t|_\varepsilon \leq \|A_J\|^t \left(\|h_t\|_\infty + 2^{\bar{t}-\varepsilon} (|h_t|_{\bar{t}} 2^{\bar{t}} + \|h_t\|_\infty (t+1)) \right).$$

□

Proof of theorem 2.3.3. The proof now follows in exactly the same way as proposition 1.3.4 and theorem 2.2.4. Notice

$$\begin{aligned}
&\left| \mu_{\mathcal{A},t}([J] \cap \sigma^{-n-|J|}[I]) - \mu_{\mathcal{A},t}([I])\mu_{\mathcal{A},t}([J]) \right| \\
&= \left| \sum_{|K|=n} \mu_{\mathcal{A},t}([JKI]) - \mu_{\mathcal{A},t}([I])\mu_{\mathcal{A},t}([J]) \right| \\
&= \left| \sum_{|K|=n} \rho(L)^{-(n+|I|+|J|)} \langle L_I L_K L_J h_t, \nu_t \rangle - \rho(L)^{-(|I|+|J|)} \langle L_I h_t, \nu_t \rangle \langle L_J h_t, \nu_t \rangle \right|
\end{aligned}$$

$$\begin{aligned}
&= \rho(L)^{-(|I|+|J|)} \left| \left\langle L_I \left(\rho(L)^{-n} \sum_{|K|=n} L_K \right) L_J h_t, \nu_t \right\rangle - \langle L_I h_t, \nu_t \rangle \langle L_J h_t, \nu_t \rangle \right| \\
&= \rho(L)^{-(|I|+|J|)} \left| \left\langle L_I \rho(L)^{-n} L_t^n L_J h_t, \nu_t \right\rangle - \langle L_I h_t, \nu_t \rangle \langle L_J h_t, \nu_t \rangle \right| \text{ by (2.6)} \\
&= \rho(L)^{-(|I|+|J|)} \left| \left\langle L_I (\rho(L)^{-n} L_t^n L_J h_t - \langle L_J h_t, \nu_t \rangle h_t), \nu_t \right\rangle \right| \\
&\leq \rho(L)^{-(|I|+|J|)} \|L_I\|_{\infty, \text{op}} \left\| \rho(L)^{-n} L_t^n L_J h_t - \langle L_J h_t, \nu_t \rangle h_t \right\|_{\infty} \\
&\leq \rho(L)^{-(|I|+|J|)} \|L_I\|_{\infty, \text{op}} \|L_J h_t\|_{\varepsilon} \beta^n \text{ by Corollary 2.3.6} \\
&\leq K \rho(L)^{-(|I|+|J|)} \|A_I\|^t \|A_J\|^t \beta^n \text{ by Lemma 2.3.7} \\
&\leq C^2 K \mu_{\mathcal{A}, t}([I]) \mu_{\mathcal{A}, t}([J]) \beta^n \text{ by Proposition 2.3.4}
\end{aligned}$$

This proves part (1) of Theorem 2.3.3, (2) follows by an approximation argument as in Bowen's book [6, Theorem 1.26]. \square

Recently in addition to the interest in Gibbs states associated with the norms of matrices there has also been significant interest in the so called singular value potential [2], [15]. In this case norms of matrices are replaced by a product of singular values, one can associate a suitable transfer operator to this potential see [21]. It seems likely that the method presented in this chapter could be extended to give decay of correlations results for Gibbs states of the singular value potential (in particular taking advantage of [21, theorem 8.10]). In addition it seems likely this method could be particularly well suited to studying Gibbs states when $t < 0$. From the perspective of thermodynamic formalism it is likely that these measures for $t < 0$ are significantly more interesting; for example it is known that the pressure function can fail to be analytic [16] and thus one expects that these systems can exhibit phase transitions. We leave this for future work.

2.4 The Weak Bernoulli Property

The purpose of this section is to prove theorem 2.1.3. The proof is similar to [44] where scalar potentials satisfying the Bowen property are considered. The key tool is a result of Bradley on ψ -mixing sequences of random variables [9] which implies lemma 2.4.1. Let $\mathcal{P} = \{[i] : i \in \Sigma\}$ be the standard partition for an invariant measure μ define

$$\psi_n^* = \sup \left\{ \frac{\mu(A \cap B)}{\mu(A)\mu(B)} : A \in \bigvee_{i=n}^{\infty} \sigma^{-i}\mathcal{P}, B \in \bigvee_{i=-\infty}^{-1} \sigma^{-i}\mathcal{P}, \mu(A)\mu(B) > 0 \right\}$$

$$\psi_n' = \inf \left\{ \frac{\mu(A \cap B)}{\mu(A)\mu(B)} : A \in \bigvee_{i=n}^{\infty} \sigma^{-i}\mathcal{P}, B \in \bigvee_{i=-\infty}^{-1} \sigma^{-i}\mathcal{P}, \mu(A)\mu(B) > 0 \right\}$$

Recall that an invariant measure μ is ψ -mixing if

$$\lim_{n \rightarrow \infty} \psi_n^* = \lim_{n \rightarrow \infty} \psi_n' = 1.$$

The following lemma is essentially a rephrasing of [8, theorem 4.1(2)].

Lemma 2.4.1. Let μ be a shift invariant measure on $\Sigma^{\mathbb{Z}}$. Suppose that for some $N > 0$ there exists a constant $C > 0$ such that

$$C^{-1}\mu([I])\mu([J]) \leq \mu([I] \cap \sigma^{-N-|J|}[J]) \leq C\mu([I])\mu([J]) \quad (2.9)$$

for all words I, J . Then μ is weak Bernoulli.

Proof sketch. Notice that for $n \geq N$ we have that

$$\mu([I] \cap \sigma^{-n-|J|}[J]) = \sum_{|K|=n-N} \mu([I] \cap \sigma^{-N-|K|-|J|}[KJ])$$

$$\begin{aligned}
&\geq C^{-1} \sum_{|K|=n-N} \mu([I])\mu([KJ]) \\
&= C^{-1}\mu([I]) \sum_{|K|=n-N} \mu([KJ]) \\
&= C^{-1}\mu([I])\mu([J]).
\end{aligned}$$

A similar argument for the other inequality shows that in fact (2.9) holds with the same constant C for all $n \geq N$. Thus we have by an approximation argument that

$$\limsup_{n \rightarrow \infty} \mu(X \cap \sigma^{-n}Y) \leq C\mu(X)\mu(Y)$$

and

$$\liminf_{n \rightarrow \infty} \mu(X \cap \sigma^{-n}Y) \geq C^{-1}\mu(X)\mu(Y)$$

for all X, Y Borel measurable. The second inequality gives that μ is totally ergodic and the first then implies that μ is mixing by a theorem of Ornstein [34, Theorem 2.1]. By an approximation argument we have that

$$\begin{aligned}
\psi_n^* &= \sup \left\{ \frac{\mu(A \cap B)}{\mu(A)\mu(B)} : A \in \bigvee_{i=n}^{\infty} \sigma^{-i}\mathcal{P}, B \in \bigvee_{i=-\infty}^{-1} \sigma^{-i}\mathcal{P}, \mu(A)\mu(B) > 0 \right\} \leq C \\
\psi_n' &= \inf \left\{ \frac{\mu(A \cap B)}{\mu(A)\mu(B)} : A \in \bigvee_{i=n}^{\infty} \sigma^{-i}\mathcal{P}, B \in \bigvee_{i=-\infty}^{-1} \sigma^{-i}\mathcal{P}, \mu(A)\mu(B) > 0 \right\} \geq C^{-1}
\end{aligned}$$

for all $n \geq N$. A result of Bradley [9, Theorem 1] implies that μ is ψ -mixing; that ψ -mixing implies weak Bernoulli is easily verified. \square

With this lemma in hand the proof of theorem 2.1.3 is merely an application of the Gibbs inequality.

Proof of theorem 2.1.3. Let N be as in the definition of primitive. Let $t > 1$ and take q such that $1/t + 1/q = 1$. Then for any I, J

$$\begin{aligned}
\mu_{\mathcal{A},t}([I] \cap \sigma^{-N-|J|}[J]) &= \sum_{|K|=N} \mu_{\mathcal{A},t}([IKJ]) \\
&\geq C^{-1} e^{-(|I|+N+|J|)P(\mathcal{A},t)} \sum_{|K|=N} \|A_I A_K A_J\|^t \\
&\geq C^{-1} e^{-(|I|+N+|J|)P(\mathcal{A},t)} M^{-Nt/q} \left(\sum_{|K|=N} \|A_I A_K A_J\| \right)^t \\
&\geq C^{-1} e^{-(|I|+N+|J|)P(\mathcal{A},t)} M^{-Nt/q} \delta^t \|A_I\|^t \|A_J\|^t \\
&\geq C^{-3} e^{-NP(\mathcal{A},t)} M^{-Nt/q} \delta^t \mu_{\mathcal{A},t}([I]) \mu_{\mathcal{A},t}([J])
\end{aligned}$$

where $M = |\Sigma|$, note that we have used that the collection is primitive in the second to the last step. For $0 < t \leq 1$ we have that

$$\begin{aligned}
\mu_{\mathcal{A},t}([I] \cap \sigma^{-N-|J|}[J]) &= \sum_{|K|=N} \mu_{\mathcal{A},t}([IKJ]) \\
&\geq C^{-1} e^{-(|I|+N+|J|)P(\mathcal{A},t)} \sum_{|K|=N} \|A_I A_K A_J\|^t \\
&\geq C^{-1} e^{-(|I|+N+|J|)P(\mathcal{A},t)} \left(\sum_{|K|=N} \|A_I A_K A_J\| \right)^t \\
&\geq C^{-1} e^{-(|I|+N+|J|)P(\mathcal{A},t)} \delta^t \|A_I\|^t \|A_J\|^t \\
&\geq C^{-3} e^{-NP(\mathcal{A},t)} \delta^t \mu_{\mathcal{A},t}([I]) \mu_{\mathcal{A},t}([J]).
\end{aligned}$$

For matrix Gibbs states the right hand inequality in equation (2.9) always holds. This is a simple consequence of the Gibbs inequality and the fact that the norm is sub-multiplicative, see [30, theorem 5]. The result then follows from lemma 2.4.1. \square

Chapter 3

Factors of Gibbs States

3.1 Introduction

Hidden Markov measures are of great interest in many areas of science, both pure and applied. It is well known that a hidden Markov measure can fail to be Markov. In fact hidden Markov measures can fail to have conditional probabilities which are continuous [25, example 4.2]. Our goal here is to study a generalization of hidden Markov measures, single site factors of g measures. These measures have attracted a significant amount attention ([41], [45], [11], [12], [24], [25]). Broadly speaking there are two main questions: when do these measures have continuous conditional probabilities (and what is their modulus of continuity) and what classes of measures are preserved by single site factors. We will focus on the second question, for results on continuity rates see [37] and [25].

Let us recall the definition of a single site factor map. Suppose that Σ and $\bar{\Sigma}$ are two alphabets and $\pi : \Sigma \rightarrow \bar{\Sigma}$ and Σ_A^+ is a shift of finite type over Σ . We define a map from $\Sigma_A^+ \rightarrow \bar{\Sigma}^{\mathbb{N}}$ which we again call π by $\pi[(x_i)_{i=0}^{\infty}] = (\pi(x_i))_{i=0}^{\infty}$, such maps are called *single site factor maps*. This map is continuous and intertwines the shift maps. Let Y be the image of π given a shift invariant measure μ on Σ_A we can define $\pi_*\mu$ on Y

as the pushforward under π .

When the shift of finite type Σ_A^+ is a full shift it is known that single site factors of Markov measures have Hölder continuous g functions [45]. However when the shift Σ_A^+ has excluded words this is no longer true. See, for instance [35, example 4] or [25, example 4.2]. However by imposing conditions on the factor map π to ensure that the fibers of π are “topologically mixing” in a certain sense, the result can be recovered. See, for instance [11], [45]. The goal of this paper is to prove the analogous results for equilibrium states associated to more general potentials. The following definition appears in [45].

Definition 3.1.1. We say that a single site factor map π is *fiber-wise sub-positive mixing* if there exists an N such that for any word $b_0 \cdots b_N$ admissible in Y and words $u_0 \cdots u_N, w_0 \cdots w_N$ such that $\pi(u_0 \cdots u_N) = \pi(w_0 \cdots w_N) = b_0 \cdots b_N$ there exists a word $a_0 \cdots a_N$ admissible in Σ_A^+ projecting to $b_0 \cdots b_N$ with $a_0 = u_0$ and $a_N = w_N$.

The goal of this chapter is then to prove the following theorem.

Theorem 3.1.2. Suppose that Σ is a finite alphabet, $\Sigma_A^+ \subseteq \Sigma^{\mathbb{N}}$ is a topologically mixing shift of finite type, μ_φ a Gibbs measure for a potential φ and π a fiber-wise sub-positive mixing factor map. If φ is Bowen (respectively Walters, Hölder) then $\pi_*\mu_\varphi$ is the Gibbs state for a potential which is Bowen (respectively Walters, Hölder).

The proof also yields the following:

Corollary 3.1.3. Suppose that Σ is a finite alphabet, $\Sigma_A^+ \subseteq \Sigma^{\mathbb{N}}$ is a topologically mixing shift of finite type, μ_g a g -measure for g and π a fiber-wise sub-positive mixing factor map. If $\log g$ is Bowen (respectively Walters, Hölder) then the logarithm of the g -function for $\pi_*\mu_g$ is Bowen (respectively Walters, Hölder).

3.2 Examples: Hidden Markov measures

Example 3.2.1. (a Markov measure which projects to a Markov measure) Let μ_S be the Markov measure on $\{0, 1, 2\}^{\mathbb{N}}$ defined by the stochastic matrix

$$S = \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \\ 1/6 & 1/6 & 2/3 \end{bmatrix}$$

and let ν be its stationary distribution. In this case

$$\nu = \begin{bmatrix} 1/4 \\ 1/4 \\ 1/2 \end{bmatrix}.$$

Suppose that the states 0 and 1 are labeled red and 2 is labeled blue. That is $\pi : \{0, 1, 2\}^{\mathbb{N}} \rightarrow \{r, b\}^{\mathbb{N}}$ is the map induced by the function $\pi(0) = \pi(1) = r$ and $\pi(2) = b$. We would like to study $\pi_*\mu_S$ let's start by making a simple observation. Consider

$$\begin{aligned} \pi_*\mu_S[rrbr] &= \mu_S[0020] + \mu_S[0021] \\ &\quad + \mu_S[0120] + \mu_S[0121] \\ &\quad + \mu_S[1020] + \mu_S[1021] \\ &\quad + \mu_S[1120] + \mu_S[1121] \end{aligned}$$

$$\begin{aligned}
&= \nu_0 S_{00} S_{02} S_{20} + \nu_0 S_{00} S_{02} S_{21} \\
&\quad + \nu_0 S_{01} S_{12} S_{20} + \nu_0 S_{01} S_{12} S_{21} \\
&\quad + \nu_1 S_{10} S_{02} S_{20} + \nu_1 S_{10} S_{02} S_{21} \\
&\quad + \nu_1 S_{11} S_{12} S_{20} + \nu_1 S_{11} S_{12} S_{21} \\
&= \begin{bmatrix} \nu_0 & \nu_1 \end{bmatrix} \begin{bmatrix} S_{00} & S_{01} \\ S_{10} & S_{11} \end{bmatrix} \begin{bmatrix} S_{02} \\ S_{12} \end{bmatrix} \begin{bmatrix} S_{20} & S_{21} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix}.
\end{aligned}$$

That is, we can write the $\pi_*\mu_S$ measure of cylinder sets in terms of a product of non-negative matrices. Of course there is nothing special about the cylinder set $rrbr$ nor about this particular example. This naturally leads us to define the matrices

$$\begin{aligned}
\mathcal{S}_{rr} &= \begin{bmatrix} 1/3 & 1/3 \\ 1/3 & 1/3 \end{bmatrix}, \quad \mathcal{S}_{rb} = \begin{bmatrix} 1/3 \\ 1/3 \end{bmatrix} \\
\mathcal{S}_{br} &= \begin{bmatrix} 1/6 & 1/6 \\ 1/6 & 1/6 \end{bmatrix}, \quad \mathcal{S}_{bb} = \begin{bmatrix} 2/3 \\ 2/3 \end{bmatrix} \\
\nu_r &= \begin{bmatrix} 1/4 & 1/4 \end{bmatrix} \quad \text{and} \quad \nu_b = \begin{bmatrix} 1/2 \end{bmatrix}.
\end{aligned}$$

The previous computation then reads nicely as

$$\pi_*\mu_S[rrbr] = \nu_r \mathcal{S}_{rr} \mathcal{S}_{rb} \mathcal{S}_{br} [1]$$

where $[1]$ is the vector of all 1's in the correct dimension. In general we have that

$$\pi_*\mu_S[y_0 y_1 \cdots y_n] = \nu_{y_0} \mathcal{S}_{y_0 y_1} \mathcal{S}_{y_1 y_2} \cdots \mathcal{S}_{y_{n-1} y_n} [1].$$

Of course for this particular choice of S we can observe that all of the matrices $\{\mathcal{S}_{rr}, \mathcal{S}_{rb}, \mathcal{S}_{br}, \mathcal{S}_{bb}\}$ are rank 1 and as a consequence the g function for $\pi_*\mu_S$ depends on 2 coordinates; in particular, $\pi_*\mu_S$ is a Markov measure.

Of course in general it is not the case that the projection of a Markov measure is again Markov.

Example 3.2.2. (a Markov measure which projects to a measure which is not Markov) Let μ_S be the Markov measure on $\{0, 1, 2\}^{\mathbb{N}}$ defined by the stochastic matrix

$$S = \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 1/3 & 0 & 2/3 \\ 1/6 & 1/6 & 2/3 \end{bmatrix}.$$

Again suppose that the states 0 and 1 are labeled red and 2 is labeled blue. That is $\pi : \{0, 1, 2\}^{\mathbb{N}} \rightarrow \{r, b\}^{\mathbb{N}}$ is the map induced by the function $\pi(0) = \pi(1) = r$ and $\pi(2) = b$. Then the same computation shows that the measure of a cylinder is determined by the matrices

$$\begin{aligned} \mathcal{S}_{rr} &= \begin{bmatrix} 1/3 & 1/3 \\ 1/3 & 0 \end{bmatrix}, \quad \mathcal{S}_{rb} = \begin{bmatrix} 1/3 \\ 2/3 \end{bmatrix} \\ \mathcal{S}_{br} &= \begin{bmatrix} 1/6 & 1/6 \end{bmatrix}, \quad \mathcal{S}_{bb} = \begin{bmatrix} 2/3 \end{bmatrix}. \end{aligned}$$

In this case \mathcal{S}_{rr} is not rank 1 and it can be shown that $\pi_*\mu_S$ is not (1-step) Markov by computing the g function at certain points, in particular $g(r^\infty)$ and $g(rrb^\infty)$. In fact it be shown that g has infinite range by computing $g(rr \cdots rb^\infty)$. However observe that all products of length 2 are positive. As positive matrices are strict contractions of the Hilbert projective metric this implies $\pi_*\mu_S$ is the Gibbs state for a Hölder continuous potential. See [45] for details.

These examples demonstrate two things. First, the presence of excluded words (or more accurately the structure of fibers $\pi^{-1}(y)$) has a significant impact on the regularity properties of the g function for hidden Markov measures. Second, they

suggest a method for extending results beyond hidden Markov measures where the matrices \mathcal{S}_{rr} , \mathcal{S}_{rb} , \mathcal{S}_{br} and \mathcal{S}_{rr} are replaced by suitable operators acting on subspaces of $C(\Sigma_A^+)$.

3.3 Factors of Hölder Gibbs states on a full shift

This section serves as the prototype for the argument in the next section while avoiding the technicalities of excluded words and potentials less regular than Hölder continuous. Suppose that $\Sigma^{\mathbb{N}}$ is a full shift and $\varphi : \Sigma^{\mathbb{N}} \rightarrow \mathbb{R}$ is Hölder continuous. That is, there exists $0 < \theta < 1$ and a constant $|\varphi|_{\theta}$ such that

$$\text{var}_n \varphi \leq |\varphi|_{\theta} \theta^n$$

for all $n \geq 0$. Assume without loss of generality that the pressure of φ is 0. Recall that the transfer operator $L_{\varphi} : C(\Sigma^{\mathbb{N}}) \rightarrow C(\Sigma^{\mathbb{N}})$ defined by

$$L_{\varphi} f(x) = \sum_i e^{\varphi(ix)} f(ix)$$

has a unique positive eigenfunction $h > 0$ and eigenmeasure ν corresponding to the eigenvalue 1 normalized so that $\langle h, \nu \rangle = 1$ (as the pressure is 0 we have that $\rho(L_{\varphi}) = 1$). Moreover the unique Gibbs state for φ is the measure $d\mu_{\varphi} = h d\nu$. Let $\pi : \Sigma^{\mathbb{N}} \rightarrow \bar{\Sigma}^{\mathbb{N}}$ be a 1-block factor map and let L_i be as in equation (1.5). For each $\bar{i} \in \bar{\Sigma}$ an operator $\mathcal{L}_{\bar{i}} : C(\Sigma^{\mathbb{N}}) \rightarrow C(\Sigma^{\mathbb{N}})$ by

$$\mathcal{L}_{\bar{i}} = \sum_{\pi(i)=\bar{i}} L_i. \tag{3.1}$$

It is simple to check that these operators are linear and bounded. We will see that these operators are strict contractions with respect to the Hilbert metric on the cone

$$\mathcal{C} = \left\{ f \in C(\Sigma^{\mathbb{N}}) : f(x) > 0, f(x) \leq f(y)e^{K\theta^{k(x,y)}} \right\}$$

where $k(x, y) = \min \{i \geq 0 : x_i \neq y_i\}$ and $K = \frac{|\varphi|_{\theta}}{\sigma - \theta}$, where σ is an arbitrary number such that $\theta < \sigma < 1$. For background on the Hilbert metric and this type of cone see the appendix.

Lemma 3.3.1. 1. Let π and μ_{φ} be as above. Then

$$\pi_*\mu_{\varphi}[\bar{z}_0\bar{z}_1 \cdots \bar{z}_{n-1}] = \left\langle \mathcal{L}_{\bar{z}_{n-1}} \cdots \mathcal{L}_{\bar{z}_1} \mathcal{L}_{\bar{z}_0} h, \nu \right\rangle.$$

2. For any $\bar{i} \in \bar{\Sigma}$ we have that $\text{diam}_{\mathcal{C}}(\mathcal{L}_{\bar{i}}\mathcal{C}) < \infty$.

Proof. 1. By Lemma 1.3.3, for any word $z_0z_1 \cdots z_{n-1} \in \Sigma^n$ we have

$$\mu_{\varphi}[z_0z_1 \cdots z_{n-1}] = \int_{\Sigma^{\mathbb{N}}} L_{z_{n-1}} \cdots L_{z_1} L_{z_0} h d\nu.$$

Therefore

$$\begin{aligned} \pi_*\mu_{\varphi}[\bar{z}_0\bar{z}_1 \cdots \bar{z}_{n-1}] &= \sum_{\pi(z_0z_1 \cdots z_{n-1}) = \bar{z}_0\bar{z}_1 \cdots \bar{z}_{n-1}} \mu_{\varphi}[z_0z_1 \cdots z_{n-1}] \\ &= \sum_{\pi(z_0z_1 \cdots z_{n-1}) = \bar{z}_0\bar{z}_1 \cdots \bar{z}_{n-1}} \left\langle L_{z_{n-1}} \cdots L_{z_1} L_{z_0} h, \nu \right\rangle \\ &= \left\langle \mathcal{L}_{\bar{z}_{n-1}} \cdots \mathcal{L}_{\bar{z}_1} \mathcal{L}_{\bar{z}_0} h, \nu \right\rangle. \end{aligned}$$

2. Suppose that $f \in \mathcal{C}$. Then

$$\mathcal{L}_{\bar{i}}f(x) = \sum_{\pi(i) = \bar{i}} e^{\varphi(ix)} f(ix)$$

$$\begin{aligned}
&= \sum_{\pi(i)=\bar{i}} e^{\varphi(ix)-\varphi(iy)} \frac{f(ix)}{f(iy)} f(iy) \\
&\leq \exp \left[|\varphi|_{\theta} \theta^{k(x,y)+1} + \frac{|\varphi|_{\theta} \theta}{\sigma - \theta} \theta^{k(x,y)+1} \right] \mathcal{L}_{\bar{i}} f(y) \\
&= \exp \left[\frac{\sigma |\varphi|_{\theta} \theta^{k(x,y)+1} - |\varphi|_{\theta} \theta^{k(x,y)+2} + |\varphi|_{\theta} \theta^{k(x,y)+2}}{\sigma - \theta} \right] \mathcal{L}_{\bar{i}} f(y) \\
&= \exp \left[\frac{\sigma |\varphi|_{\theta} \theta^{k(x,y)+1}}{\sigma - \theta} \right] \mathcal{L}_{\bar{i}} f(y).
\end{aligned}$$

Thus by corollary A.1.9 we have that for any $f, g \in \mathcal{C}$,

$$\Theta_{\mathcal{C}}(\mathcal{L}_{\bar{i}} f, \mathcal{L}_{\bar{i}} g) \leq 2 \log \frac{1 + \sigma}{1 - \sigma} + 2\sigma \frac{|\varphi|_{\theta} \theta}{\sigma - \theta}.$$

□

Define

$$D = 2 \log \frac{1 + \sigma}{1 - \sigma} + 2\sigma \frac{|\varphi|_{\theta} \theta}{\sigma - \theta}$$

and

$$\gamma = \tanh \left(\frac{D}{4} \right).$$

Theorem 3.3.2. (Redig and Wang [38]) Let $\pi : \Sigma^{\mathbb{N}} \rightarrow \bar{\Sigma}^{\mathbb{N}}$ be a 1-block factor map, φ Hölder continuous and μ_{φ} the unique Gibbs state for φ . Then the g function for $\pi_* \mu_{\varphi}$

$$g(\bar{z}) = \lim_{n \rightarrow \infty} \frac{\pi_* \mu_{\varphi}[\bar{z}_0 \bar{z}_1 \cdots \bar{z}_n]}{\pi_* \mu_{\varphi}[\bar{z}_1 \cdots \bar{z}_n]}$$

is Hölder continuous.

Proof. Notice that for $x, y \in \bar{\Sigma}^{\mathbb{N}}$ such that $x_i = y_i$ for all $0 \leq i \leq N$ and any $n, m \geq 1$ we have

$$\begin{aligned} & \left| \log \frac{\pi_* \mu_\varphi[x_0 x_1 \cdots x_n]}{\pi_* \mu_\varphi[x_1 \cdots x_n]} - \log \frac{\pi_* \mu_\varphi[y_0 y_1 \cdots y_m]}{\pi_* \mu_\varphi[y_1 \cdots y_m]} \right| \\ &= \left| \log \frac{\langle \mathcal{L}_{x_n} \cdots \mathcal{L}_{x_1} \mathcal{L}_{x_0} h, \nu \rangle}{\langle \mathcal{L}_{x_n} \cdots \mathcal{L}_{x_1} h, \nu \rangle} - \log \frac{\langle \mathcal{L}_{y_m} \cdots \mathcal{L}_{y_1} \mathcal{L}_{y_0} h, \nu \rangle}{\langle \mathcal{L}_{y_m} \cdots \mathcal{L}_{y_1} h, \nu \rangle} \right| \\ &= \left| \log \frac{\langle \mathcal{L}_{x_N} \cdots \mathcal{L}_{x_1} \mathcal{L}_{x_0} h, \nu_{x,n} \rangle \langle \mathcal{L}_{x_N} \cdots \mathcal{L}_{x_1} h, \nu_{y,m} \rangle}{\langle \mathcal{L}_{x_N} \cdots \mathcal{L}_{x_1} h, \nu_{x,n} \rangle \langle \mathcal{L}_{x_N} \cdots \mathcal{L}_{x_1} \mathcal{L}_{x_0} h, \nu_{y,m} \rangle} \right|. \end{aligned}$$

Where

$$\nu_{x,n} = \mathcal{L}_{x_{N+1}}^* \cdots \mathcal{L}_{x_{n-1}}^* \mathcal{L}_{x_n}^* \nu$$

and similarly for $\nu_{y,m}$. Therefore by proposition A.1.4 we have that

$$\begin{aligned} & \left| \log \frac{\pi_* \mu_\varphi[x_0 x_1 \cdots x_n]}{\pi_* \mu_\varphi[x_1 \cdots x_n]} - \log \frac{\pi_* \mu_\varphi[y_0 y_1 \cdots y_m]}{\pi_* \mu_\varphi[y_1 \cdots y_m]} \right| \\ & \leq \Theta_{\mathcal{C}}(L_{x_N} \cdots L_{x_1} L_{x_0} h, L_{x_N} \cdots L_{x_1} h) \\ & \leq \gamma^{N-1} \Theta_{\mathcal{C}}(L_{x_1} L_{x_0} h, L_{x_1} h) \leq \gamma^N \frac{D}{\gamma}. \end{aligned}$$

Taking the limit we have that g is Hölder. □

3.4 The classical uniqueness regimes

The purpose of this section is to consider factors of Gibbs states on subshifts of finite type. It is known from [45] that factors of Markov measures under a fiber-wise sub-positive mixing map have Hölder continuous g functions. We have seen that factors of Hölder Gibbs states on full shifts have Hölder continuous g functions. The goal of this section is to show that in fact all of the classical uniqueness regimes from definition

1.3.2 are closed under fiber-wise sub-positive mixing factor maps. In particular in this section we will prove theorem 3.1.2. Let L_φ be the Ruelle operator associated to φ and assume without loss that $\rho(L_\varphi) = 1$, in other words the pressure of φ is 0. Take ν such that $\mathcal{L}^*\nu = \nu$. Assume that φ is Bowen.

Let $\pi : \Sigma \rightarrow \bar{\Sigma}$ be a map inducing a 1-block factor map $\pi : \Sigma_A^+ \rightarrow Y \subseteq \bar{\Sigma}^{\mathbb{N}}$ and assume that π is fiber-wise sub-positive mixing. For each $i, j \in \Sigma$ with ji admissible in Σ_A^+ define the operator

$$L_{ij}f(x) = e^{\varphi(jx)} f(jx) \chi_{[i]}(x).$$

For example in the case of the Markov measure μ_S from example 3.2.2 L_{00} is simply the matrix

$$\begin{bmatrix} 1/3 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

As μ_S is the Gibbs state for the potential $\varphi(x) = \log S_{x_0x_1}$. In a similar manner as (3.1) for bb' admissible in Y define an operator $\mathcal{L}_{bb'} : C(\Sigma_A^+) \rightarrow C(\Sigma_A^+)$ by

$$\mathcal{L}_{bb'}f = \sum_{\pi(i)=b, \pi(j)=b'} L_{ij}f.$$

In the case of example 3.2.2 the operator $\mathcal{L}_{br} = L_{20} + L_{21}$ is the matrix

$$\mathcal{L}_{br} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1/3 & 2/3 & 0 \end{bmatrix}.$$

Recall that when $\varphi(x) = \log S_{x_0x_1}$ the matrix representation for the transfer operator

L_φ acting on $\text{span} \{ \chi_{[a]} : a \in \Sigma \}$ is given by the transpose of the stochastic matrix S . Given a word $w = w_0 \cdots w_n$ admissible in Σ_A^+ define

$$L_w = L_{w_n w_{n-1}} \cdots L_{w_2 w_1} L_{w_1 w_0}$$

and notice that

$$\begin{aligned} L_{w_2 w_1} L_{w_1 w_0} f(x) &= e^{\varphi(w_0 x)} L_{w_1 w_0} f(w_1 x) \chi_{[w_2]}(x) \\ &= e^{\varphi(w_0 x)} \left(e^{\varphi(w_0 w_1 x)} f(w_0 w_1 x) \chi_{[w_1]}(w_1 x) \right) \chi_{[w_2]}(x) \\ &= e^{S_2 \varphi(w_0 w_1 x)} f(w_0 w_1 x) \chi_{[w_2]}(x). \end{aligned}$$

Thus by iteration

$$L_w f(x) = e^{S_n \varphi(w_0 \cdots w_{n-1} x)} f(w_0 \cdots w_{n-1} x) \chi_{[w_n]}(x).$$

Similarly for $\bar{w} = \bar{w}_0 \cdots \bar{w}_n$ admissible in Y define

$$\mathcal{L}_{\bar{w}} = \mathcal{L}_{\bar{w}_n \bar{w}_{n-1}} \cdots \mathcal{L}_{\bar{w}_2 \bar{w}_1} \mathcal{L}_{\bar{w}_1 \bar{w}_0}.$$

We will make use of a strategy of using sequences of cones to obtain sub exponential rates of convergence that has been used previously in the decay of correlations literature [26]. Give a metric d and a symbol a define a cone

$$\mathcal{C}([a], d) = \left\{ f \in C([a]) : f \geq 0 \text{ and } f(x) \leq f(x') e^{d(x, x')} \text{ for all } x, x' \in [a] \right\}$$

and given a number $B > 0$ define a cone

$$\mathcal{C}([a], +B) = \left\{ f \in C([a]) : f \geq 0 \text{ and } f(x) \leq e^B f(x') \text{ for all } x, x' \in [a] \right\}.$$

We now produce a sequence of metrics for which long products of the operators L_{ij} map $\mathcal{C}([a], d_j)$ to $\mathcal{C}([b], \sigma d_{j+1})$ for some $0 < \sigma < 1$. This is crucial to bounding the projective diameter (see corollary A.1.9).

Procedure 3.4.1. This procedure has two inputs. A function φ with the Bowen property (which is fixed throughout this section) and a sequence α_k which is positive and decreasing to 0 (we will make different choices for α_k depending on our needs). What it produces is a sequence of natural numbers $\{n_j\}_{j=0}^{\infty}$ and a sequence of metrics $\{d_j\}_{j=0}^{\infty}$ with specific properties. Construct a sequence $\{n_j\}_{j=0}^{\infty}$ of numbers and metrics $\{d_j\}_{j=0}^{\infty}$ in following way. Set

$$\alpha(0, k) = \alpha_k \text{ and } \alpha(m, k) = \text{var}_{m+k} S_m \varphi.$$

Fix some $0 < \sigma < 1$. Inductively define a sequence $\{n_i\}_{i=1}^{\infty}$: take n_1 such that for all $n \geq n_1$ we have $\alpha(0, n) \leq \frac{\sigma}{2}$. As α_k converges to 0 and $S_{n_1} \varphi$ is continuous we may take n_2 such that for all $n \geq n_2$ we have that

$$\alpha(0, n_1 + n) \leq \frac{\sigma^2}{2^2} \text{ and } \alpha(n_1, n) \leq \frac{\sigma}{2}.$$

Continue in this way, that is given $\{n_i\}_{i=1}^{j-1}$ choose n_j such that for all $n \geq n_j$ we have

$$\begin{aligned} \alpha(0, n_1 + n_2 + \cdots + n_{j-1} + n) &\leq \frac{\sigma^j}{2^j}, \\ \alpha(n_1, n_2 + n_3 + \cdots + n_{j-1} + n) &\leq \frac{\sigma^{j-1}}{2^{j-1}}, \\ \alpha(n_2, n_3 + \cdots + n_{j-1} + n) &\leq \frac{\sigma^{j-2}}{2^{j-2}}, \\ &\vdots \\ \alpha(n_i, n_{i+1} + n_{i+2} + \cdots + n_{j-1} + n) &\leq \frac{\sigma^{j-i}}{2^{j-i}} \\ &\vdots \end{aligned}$$

$$\alpha(n_{j-1}, n) \leq \frac{\sigma}{2}$$

Define

$$d_{j,k} = \sum_{i=0}^j \frac{\alpha(n_i, n_{i+1} + n_{i+2} + \cdots + n_j + k)}{\sigma^{j-i}} \quad (3.2)$$

where $n_0 = 0$. Notice that for each j , $d_{j,k} \xrightarrow{k \rightarrow \infty} 0$ thus we can then define a sequence of metrics $\{d_j\}_{j=0}^{\infty}$ by

$$d_j(x, y) = d_{j,k(x,y)} \quad (3.3)$$

where $k(x, y) = \min \{i : x_i \neq y_i\}$. Notice that $d_{0,k} = \alpha_k$.

Remark 2. Unless φ is Hölder $n_j \xrightarrow{j \rightarrow \infty} \infty$. In the case of a Hölder potential (lemma 3.3.1) we can have that $n_j = 1$ and $d_{j,k} = \frac{|\varphi|_{\theta} \theta^{k+1}}{\sigma - \theta}$ for all j .

What we have gained from this construction of $\{n_i\}_{i=0}^{\infty}$ and $d_{j,k}$ is the following. First,

$$\begin{aligned} \alpha(n_{j+1}, k) + d_{j, n_{j+1} + k} &= \alpha(n_{j+1}, k) + \sum_{i=0}^j \frac{\alpha(n_i, n_{i+1} + n_{i+2} + \cdots + n_j + n_{j+1} + k)}{\sigma^{j-i}} \\ &= \sigma \left(\frac{\alpha(n_{j+1}, k)}{\sigma} + \sum_{i=0}^j \frac{\alpha(n_i, n_{i+1} + n_{i+2} + \cdots + n_j + n_{j+1} + k)}{\sigma^{j-i+1}} \right) \\ &= \sigma \left(\sum_{i=0}^{j+1} \frac{\alpha(n_i, n_{i+1} + n_{i+2} + \cdots + n_j + n_{j+1} + k)}{\sigma^{(j+1)-i}} \right) \\ &= \sigma d_{j+1, k}. \end{aligned}$$

This implies that n_{j+1} products of the operators L_{ij} map a cone of the type $\mathcal{C}([a], d_j)$ into a cone of the type $\mathcal{C}([b], \sigma d_{j+1})$ which is the content of lemma 3.4.2. Second, we have that for $k \geq 0$

$$d_{j,k} = \sum_{i=0}^j \frac{\alpha(n_i, n_{i+1} + n_{i+2} + \cdots + n_{j-1} + n_j + k)}{\sigma^{j-i}} \leq \sum_{i=0}^{j-1} \frac{1}{2^{j-i}} + \frac{\alpha(n_j, k)}{\sigma} \leq 2 + \frac{K}{\sigma}.$$

Where K is the constant from the Bowen property for φ , this implies that for all j and a we have that $\mathcal{C}([a], d_j) \subseteq \mathcal{C}([a], +B)$ where $B = 2 + \frac{K}{\sigma}$.

Lemma 3.4.2. For any j and word $w = w_0 \cdots w_{n_{j+1}}$ we have that

$$L_w : \mathcal{C}([w_0], d_j) \rightarrow \mathcal{C}([w_{n_{j+1}}], \sigma d_{j+1}).$$

Proof. Suppose that $f \in \mathcal{C}([w_0], d_j)$. For $x, x' \in [w_{n_{j+1}}]$ with $x_i = x'_i$ for all $0 \leq i \leq k-1$. We have by definition that

$$\left| S_{n_{j+1}} \varphi(w_0 \cdots w_{n_{j+1}-1} x) - S_{n_{j+1}} \varphi(w_0 \cdots w_{n_{j+1}-1} x') \right| \leq \alpha(n_{j+1}, k)$$

and

$$\begin{aligned} \frac{f(w_0 \cdots w_{n_{j+1}-1} x)}{f(w_0 \cdots w_{n_{j+1}-1} x')} &\leq \exp \left[d_j(w_0 \cdots w_{n_{j+1}-1} x, w_0 \cdots w_{n_{j+1}-1} x') \right] \\ &= \exp \left[d_{j, n_{j+1}+k} \right] \end{aligned}$$

because $f \in \mathcal{C}([w_{n_{j+1}}], d_j)$ and $w_0 \cdots w_{n_{j+1}-1} x, w_0 \cdots w_{n_{j+1}-1} x'$ agree for $n_{j+1} + k$ places. Therefore

$$\begin{aligned} L_w f(x) &= e^{S_{n_{j+1}} \varphi(w_0 \cdots w_{n_{j+1}-1} x)} f(w_0 \cdots w_{n_{j+1}-1} x) \\ &= e^{S_{n_{j+1}} \varphi(w_0 \cdots w_{n_{j+1}-1} x) - S_{n_{j+1}} \varphi(w_0 \cdots w_{n_{j+1}-1} x') + S_{n_{j+1}} \varphi(w_0 \cdots w_{n_{j+1}-1} x')} \\ &\quad \times \frac{f(w_0 \cdots w_{n_{j+1}-1} x)}{f(w_0 \cdots w_{n_{j+1}-1} x')} f(w_0 \cdots w_{n_{j+1}-1} x') \\ &\leq \exp \left[\alpha(n_{j+1}, k) + d_{j, n_{j+1}+k} \right] L_w f(x') \\ &= e^{\sigma d_{j+1}(x, x')} L_w f(x') \end{aligned}$$

Hence $L_w f \in \mathcal{C}([w_{n_{j+1}}], \sigma d_{j+1})$. \square

The strategy of the proof of Theorem 3.1.2 will be the following. Define $\psi : Y \rightarrow \mathbb{R}$ by

$$\psi(y) = \lim_{m \rightarrow \infty} \log \frac{\langle \mathcal{L}_{y_m y_{m-1}} \cdots \mathcal{L}_{y_2 y_1} \mathcal{L}_{y_1 y_0} 1, \nu \rangle}{\langle \mathcal{L}_{y_m y_{m-1}} \cdots \mathcal{L}_{y_2 y_1} 1, \nu \rangle}. \quad (3.4)$$

Recall that if ψ is Bowen then it has a unique Gibbs state μ_ψ which satisfies the Gibbs inequality

$$C^{-1} \leq \frac{\mu_\psi[y_0 \cdots y_{n-1}]}{e^{S_n \psi(y)}} \leq C$$

(that the pressure is 0 is a consequence of the assumption that $\rho(L_\varphi) = 1$). A computation (Proposition 3.4.6) shows that

$$e^{S_n \psi(y)} \approx \pi_* \nu[y_0 \cdots y_{n-1}] \approx \pi_* \mu_\varphi[y_0 \cdots y_{n-1}].$$

Thus μ_ψ and $\pi_* \mu_\varphi$ are mutually absolutely continuous and therefore equal as they are ergodic. $\pi_* \mu_\varphi$ is then the Gibbs state for ψ and theorem 3.1.2 will follow by proving that ψ is Hölder, Walters, or Bowen respectively.

This proof strategy is similar to [25] and [37] with the observation that the choice of the measure ν is arbitrary (although natural). The notable difference being that we appeal to cone-theoretic techniques while [25] and [37] are based on “hands on” bounds. That is, the function defined in [25] and [37] is essentially the same as (3.4) with ν replaced by a point mass. The real benefit to our method is that we can prove that classes are closed under factor maps. This is in contrast to [25] and [37] where the main application is continuity rates.

We must now construct a sequence of cones. To do so we identify $C(\Sigma_{\mathcal{A}})$ with $\bigoplus_{a \in \Sigma} C([a])$ and define cones as in the proposition A.1.6. Given a word $\bar{w} = \bar{w}_0 \cdots \bar{w}_n$

admissible in Y and a metric d , define cones

$$\mathcal{C}(\text{in}, \bar{w}, d) = \bigoplus_{a: \exists w, \pi(w) = \bar{w}, w_0 = a} \mathcal{C}([a], d)$$

and

$$\mathcal{C}(\text{out}, \bar{w}, d) = \bigoplus_{a: \exists w, \pi(w) = \bar{w}, w_n = a} \mathcal{C}([a], d).$$

We will give names to the sets

$$\bar{w}_{\text{in}} = \{a : \exists w, \pi(w) = \bar{w}, w_0 = a\} \subseteq \Sigma$$

and

$$\bar{w}_{\text{out}} = \{a : \exists w, \pi(w) = \bar{w}, w_n = a\} \subseteq \Sigma.$$

In addition we will call

$$\begin{aligned} \mathcal{C}(\text{in}, \bar{w}, +) &= \bigoplus_{a \in \bar{w}_{\text{in}}} \mathcal{C}([a], +) \\ &= \left\{ f \geq 0 : f(x) = 0 \text{ for all } x \notin \bigcup_{a \in \bar{w}_{\text{in}}} [a] \right\} \end{aligned}$$

and

$$\begin{aligned} \mathcal{C}(\text{in}, \bar{w}, +B) &= \bigoplus_{a \in \bar{w}_{\text{in}}} \mathcal{C}([a], +B) \\ &= \left\{ f \in \mathcal{C}(\text{in}, \bar{w}, +) : f(x) \leq e^B f(x') \text{ for all } x, x' \text{ with } x_0 = x'_0 \right\} \end{aligned}$$

and similarly for “out”. To get a sense of these definitions observe that in example

3.2.2 we have that

$$\mathcal{C}(\text{in}, br, +) = \left\{ \begin{bmatrix} a_0 \\ a_1 \\ 0 \end{bmatrix} : a_0, a_1 \geq 0 \right\}$$

and

$$\mathcal{C}(\text{out}, br, +) = \left\{ \begin{bmatrix} 0 \\ 0 \\ a_2 \end{bmatrix} : a_2 \geq 0 \right\}.$$

Recall that because the potential in example 3.2.2 depends only on 2 coordinates we can think of these cones as the non-negative quadrant of $\text{span} \{ \chi_{[a]} : a \in \Sigma \} \cong \mathbb{R}^{|\Sigma|}$. In general the cones $\mathcal{C}(\text{in}, \bar{w}, d)$, $\mathcal{C}(\text{out}, \bar{w}, d)$ are simply non-negative functions supported on the cylinder sets corresponding to \bar{w}_{in} and \bar{w}_{out} for which the restriction to any cylinder set $[a]$ is in the cone $\mathcal{C}([a], d)$.

Lemma 3.4.3. Suppose that $\bar{w} = \bar{w}_0 \cdots \bar{w}_{n_{j+1}}$ is a word admissible in Y . We have that

$$\mathcal{L}_{\bar{w}} : \mathcal{C}(\text{in}, \bar{w}, d_j) \rightarrow \mathcal{C}(\text{out}, \bar{w}, \sigma d_{j+1}).$$

Proof. This is a consequence of the fact that $\mathcal{L}_{\bar{w}} = \sum_{\pi(w)=\bar{w}} L_w$ and lemma 3.4.2. \square

Lemma 3.4.4. Let N be as in definition 3.1.1. There exists a constant C such that

$$\Theta_{\mathcal{C}(\text{out}, \bar{w}, +)}(\mathcal{L}_{\bar{w}} f, 1) \leq C$$

for all $f \in \mathcal{C}(\text{in}, \bar{w}, +B)$ and admissible words $\bar{w} = \bar{w}_0 \cdots \bar{w}_N$.

Proof. Let $z \in \cup_{\pi(a)=w_0}[a]$ be such that $f(z) = \|f\|_\infty$. By the assumption that π is fiberwise sub-positive mixing we have that for any $x \in \cup_{\pi(a)=\bar{w}_N}[a]$ there exists a word w with $\pi(w) = \bar{w}$ and $w_0 = z_0, w_N = x_0$. Thus

$$\mathcal{L}_{\bar{w}}f(x) \geq L_w f(x) = e^{S_N \varphi(w_0 \cdots w_{N-1} x_0 \cdots)} f(w_0 \cdots w_{N-1} x_0 \cdots) \geq e^{-N\|\varphi\|_\infty} \|f\|_\infty e^{-B}$$

on the other hand we have that

$$\mathcal{L}_{\bar{w}}f(x) \leq \|\mathcal{L}_{\bar{w}}\|_{\text{op}} \|f\|_\infty \leq \|L_\varphi\|_{\text{op}}^N \|f\|_\infty.$$

Therefore

$$\Theta_{\mathcal{C}(\text{out}, \bar{w}, +)}(\mathcal{L}_{\bar{w}}f, 1) \leq \log \left(\frac{\|L_\varphi\|_{\text{op}}^N}{e^{-B-N\|\varphi\|_\infty}} \right).$$

□

Remark 3. Note by potentially taking n_j slightly larger if necessary we can have that $n_j \geq N$ for all j .

Lemma 3.4.5. 1. There exists a constant D such that for any j and $\bar{w} = \bar{w}_0 \cdots \bar{w}_{n_{j+1}}$ admissible in Y we have that

$$\text{diam}_{\mathcal{C}(\text{out}, \bar{w}, d_{j+1})}(\mathcal{L}_{\bar{w}}\mathcal{C}(\text{in}, \bar{w}, d_j)) \leq D.$$

2. Let $j \geq 0, k \geq 1$ and set $m = \sum_{i=1}^k n_{j+i}$. For any word $y_0 \cdots y_m$ admissible in Y we have that

$$\Theta_{\mathcal{C}(\text{out}, y_{n_{j+1}} \cdots y_m, d_{j+2})}(\mathcal{L}_{y_m y_{m-1}} \cdots \mathcal{L}_{y_1 y_0} f, \mathcal{L}_{y_m y_{m-1}} \cdots \mathcal{L}_{y_1 y_0} g) \leq \gamma^k \Theta_{\mathcal{C}(\text{in}, y_0 \cdots y_{n_{j+1}}, d_j)}(f, g)$$

where $\gamma = \tanh(D/4)$.

Proof. 1. Suppose that $f, g \in \mathcal{C}(\text{in}, \bar{w}, d_j)$ as $\mathcal{L}_{\bar{w}}f, \mathcal{L}_{\bar{w}}g \in \mathcal{C}(\text{out}, \bar{w}, \sigma d_{j+1})$ we have by corollary A.1.9 that

$$\begin{aligned} \Theta_{\mathcal{C}(\text{out}, \bar{w}, d_{j+1})}(\mathcal{L}_{\bar{w}}f, \mathcal{L}_{\bar{w}}g) &\leq 2 \log \frac{1 + \sigma}{1 - \sigma} + \Theta_{\mathcal{C}(\text{out}, \bar{w}, +)}(\mathcal{L}_{\bar{w}}f, \mathcal{L}_{\bar{w}}g) \\ &\leq 2 \log \frac{1 + \sigma}{1 - \sigma} + 2C \end{aligned}$$

where C is as in lemma 3.4.4.

2. We will show the second assertion when $k = 2$ the general case for $k \geq 1$ is similar. Set $\bar{w}^1 = y_0 \cdots y_{n_{j+1}}$ and $\bar{w}^2 = y_{n_{j+1}} \cdots y_m$. Notice that

$$\mathcal{L}_{y_m y_{m-1}} \cdots \mathcal{L}_{y_1 y_0} = \mathcal{L}_{\bar{w}^2} \mathcal{L}_{\bar{w}^1}.$$

Define a map $R_{\bar{w}_{\text{in}}^2} : \mathcal{C}(\text{out}, \bar{w}^1, d_{j+1}) \rightarrow \mathcal{C}(\text{in}, \bar{w}^2, d_{j+1})$ by

$$R_{\bar{w}_{\text{in}}^2} f(x) = \sum_{a \in \bar{w}_{\text{in}}^2} \chi_{[a]}(x) f(x).$$

Notice that for any $f \in C(\Sigma_A)$, $f - R_{\bar{w}_{\text{in}}^2} f \in \ker \mathcal{L}_{\bar{w}^2}$ because is it supported on cylinder sets for which no word projecting to \bar{w}^2 begins ($R_{\bar{w}_{\text{in}}^2}$ is nothing but a restriction map). Thus $\mathcal{L}_{\bar{w}^2} f = \mathcal{L}_{\bar{w}^2} R_{\bar{w}_{\text{in}}^2} f$. To get a sense of what is happening consider the following diagram

$$\mathcal{C}(\text{in}, \bar{w}^1, d_j) \xrightarrow{\mathcal{L}_{\bar{w}^1}} \mathcal{C}(\text{out}, \bar{w}^1, d_{j+1}) \xrightarrow{R_{\bar{w}_{\text{in}}^2}} \mathcal{C}(\text{in}, \bar{w}^2, d_{j+1}) \xrightarrow{\mathcal{L}_{\bar{w}^2}} \mathcal{C}(\text{out}, \bar{w}^2, d_{j+2}) \quad (3.5)$$

Hence

$$\begin{aligned}
\Theta_{\mathcal{C}(\text{out}, \bar{w}^2, d_{j+2})}(\mathcal{L}_{\bar{w}^2} \mathcal{L}_{\bar{w}^1} f, \mathcal{L}_{\bar{w}^2} \mathcal{L}_{\bar{w}^1} g) &= \Theta_{\mathcal{C}(\text{out}, \bar{w}^2, d_{j+2})}(\mathcal{L}_{\bar{w}^2} R_{\bar{w}_{\text{in}}^2} \mathcal{L}_{\bar{w}^1} f, \mathcal{L}_{\bar{w}^2} R_{\bar{w}_{\text{in}}^2} \mathcal{L}_{\bar{w}^1} g) \\
&\leq \gamma \Theta_{\mathcal{C}(\text{in}, \bar{w}^2, d_{j+1})}(R_{\bar{w}_{\text{in}}^2} \mathcal{L}_{\bar{w}^1} f, R_{\bar{w}_{\text{in}}^2} \mathcal{L}_{\bar{w}^1} g) \\
&\leq \gamma \Theta_{\mathcal{C}(\text{out}, \bar{w}^1, d_{j+1})}(\mathcal{L}_{\bar{w}^1} f, \mathcal{L}_{\bar{w}^1} g) \\
&\leq \gamma^2 \Theta_{\mathcal{C}(\text{in}, \bar{w}^1, d_j)}(f, g)
\end{aligned}$$

by theorem A.1.3. □

Proposition 3.4.6. Suppose that φ is Bowen.

1. The sequence of functions $\{f_m\} \subseteq C(Y)$ defined by

$$f_m(y) = \log \frac{\langle \mathcal{L}_{y_m y_{m-1}} \cdots \mathcal{L}_{y_2 y_1} \mathcal{L}_{y_1 y_0} 1, \nu \rangle}{\langle \mathcal{L}_{y_m y_{m-1}} \cdots \mathcal{L}_{y_2 y_1} 1, \nu \rangle} = \log \frac{\pi_* \nu[y_0 y_1 \cdots y_m]}{\pi_* \nu[y_1 \cdots y_m]}$$

converges uniformly to a function $\psi : Y \rightarrow \mathbb{R}$. Moreover ψ is Bowen.

2. If μ_ψ is the unique Gibbs states for ψ then there exist constants $C_1, C_2 > 0$ such that

$$C_1 \leq \frac{\mu_\psi[y_0 \cdots y_{n-1}]}{\pi_* \mu_\varphi[y_0 \cdots y_{n-1}]} \leq C_2.$$

for all y .

3. $\mu_\psi = \pi_* \mu_\varphi$.

Proof. 1. Let $\varepsilon > 0$. Let $\alpha_k = \text{var}_k \varphi$, set d_j to be the metrics as in equation (3.3) and notice that this implies that $L_{ij} 1 \in \mathcal{C}([i], d_0)$. Let D and γ be as in lemma 3.4.5. Take $N = 1 + \sum_{i=1}^k n_i$ where k is chosen such that $\gamma^{k-1} D < \varepsilon$. Suppose

that $y \in Y$ and that $n, m \geq N$. Then

$$\begin{aligned}
& |f_n(y) - f_m(y)| \\
&= \left| \log \frac{\langle \mathcal{L}_{y_n y_{n-1}} \cdots \mathcal{L}_{y_2 y_1} \mathcal{L}_{y_1 y_0} 1, \nu \rangle}{\langle \mathcal{L}_{y_n y_{n-1}} \cdots \mathcal{L}_{y_2 y_1} 1, \nu \rangle} - \log \frac{\langle \mathcal{L}_{y_m y_{m-1}} \cdots \mathcal{L}_{y_2 y_1} \mathcal{L}_{y_1 y_0} 1, \nu \rangle}{\langle \mathcal{L}_{y_m y_{m-1}} \cdots \mathcal{L}_{y_2 y_1} 1, \nu \rangle} \right| \\
&= \left| \log \frac{\langle \mathcal{L}_{y_N y_{N-1}} \cdots \mathcal{L}_{y_2 y_1} \mathcal{L}_{y_1 y_0} 1, \nu_{y,n} \rangle \langle \mathcal{L}_{y_N y_{N-1}} \cdots \mathcal{L}_{y_2 y_1} 1, \nu_{y,m} \rangle}{\langle \mathcal{L}_{y_N y_{N-1}} \cdots \mathcal{L}_{y_2 y_1} 1, \nu_{y,n} \rangle \langle \mathcal{L}_{y_N y_{N-1}} \cdots \mathcal{L}_{y_2 y_1} \mathcal{L}_{y_1 y_0} 1, \nu_{y,m} \rangle} \right|
\end{aligned}$$

where

$$\nu_{y,n} = \mathcal{L}_{y_{N+1} y_N}^* \cdots \mathcal{L}_{y_n y_{n-1}}^* \nu$$

and similarly for $\nu_{y,m}$. For the sake of notation divide $y_1 y_2 \cdots y_N$ into words $\bar{w}^1 = y_1 \cdots y_{n_1+1}$, $\bar{w}^2 = y_{n_1+1} \cdots y_{n_1+n_2+1}$ and so on. Thus by lemma A.1.4

$$|f_n(y) - f_m(y)| \leq \Theta_+(\mathcal{L}_{y_n y_{n-1}} \cdots \mathcal{L}_{y_2 y_1} (\mathcal{L}_{y_1 y_0} 1), \mathcal{L}_{y_n y_{n-1}} \cdots \mathcal{L}_{y_2 y_1} 1),$$

where Θ_+ is the Hilbert metric on the cone of non-negative functions. In addition, by our choice of α_k we have

$$R_{\bar{w}_{\text{in}}^1} 1, R_{\bar{w}_{\text{in}}^1} \mathcal{L}_{y_0 y_1} 1 \in \mathcal{C}(\text{in}, \bar{w}^1, d_0).$$

Thus by theorem A.1.3 and lemma 3.4.5 we have that

$$\begin{aligned}
|f_n(y) - f_m(y)| &\leq \Theta_+(\mathcal{L}_{y_n y_{n-1}} \cdots \mathcal{L}_{y_2 y_1} (\mathcal{L}_{y_1 y_0} 1), \mathcal{L}_{y_n y_{n-1}} \cdots \mathcal{L}_{y_2 y_1} 1) \\
&\leq \Theta_{\mathcal{C}(\text{out}, \bar{w}^k, d_k)}(\mathcal{L}_{\bar{w}^k} \cdots \mathcal{L}_{\bar{w}^1} (\mathcal{L}_{y_1 y_0} 1), \mathcal{L}_{\bar{w}^k} \cdots \mathcal{L}_{\bar{w}^1} 1) \\
&\leq \gamma^{k-1} \Theta_{\mathcal{C}(\text{in}, \bar{w}^2, d_1)}(R_{\bar{w}_{\text{in}}^2} \mathcal{L}_{\bar{w}^1} (\mathcal{L}_{y_1 y_0} 1), R_{\bar{w}_{\text{in}}^2} \mathcal{L}_{\bar{w}^1} 1) \\
&\leq \gamma^{k-1} \Theta_{\mathcal{C}(\text{out}, \bar{w}^1, d_1)}(\mathcal{L}_{\bar{w}^1} (\mathcal{L}_{y_1 y_0} 1), \mathcal{L}_{\bar{w}^1} 1) \leq \gamma^{k-1} D < \varepsilon.
\end{aligned}$$

Therefore $\{f_n\}$ is Cauchy in $C(Y)$ and thus converges by completeness. To see that ψ is Bowen notice that

$$\begin{aligned} S_n\psi(y) &= \sum_{k=0}^{n-1} \lim_{m \rightarrow \infty} \log \frac{\langle \mathcal{L}_{y_m y_{m-1}} \cdots \mathcal{L}_{y_{k+2} y_{k+1}} \mathcal{L}_{y_{k+1} y_k} 1, \nu \rangle}{\langle \mathcal{L}_{y_m y_{m-1}} \cdots \mathcal{L}_{y_{k+2} y_{k+1}} 1, \nu \rangle} \\ &= \lim_{m \rightarrow \infty} \log \frac{\langle \mathcal{L}_{y_m y_{m-1}} \cdots \mathcal{L}_{y_{n+1} y_n} \mathcal{L}_{y_n y_{n-1}} \cdots \mathcal{L}_{y_1 y_0} 1, \nu \rangle}{\langle \mathcal{L}_{y_m y_{m-1}} \cdots \mathcal{L}_{y_{n+1} y_n} 1, \nu \rangle}. \end{aligned}$$

Thus for any n and $y, y' \in Y$ such that $y_i = y'_i$ for all $0 \leq i \leq n-1$ we have that

$$\begin{aligned} &|S_n\psi(y) - S_n\psi(y')| \\ &= \lim_{m \rightarrow \infty} \left| \log \frac{\langle \mathcal{L}_{y_m y_{m-1}} \cdots \mathcal{L}_{y_{n+1} y_n} \mathcal{L}_{y_n y_{n-1}} \cdots \mathcal{L}_{y_1 y_0} 1, \nu \rangle \langle \mathcal{L}_{y'_m y'_{m-1}} \cdots \mathcal{L}_{y'_{n+1} y'_n} 1, \nu \rangle}{\langle \mathcal{L}_{y_m y_{m-1}} \cdots \mathcal{L}_{y_{n+1} y_n} 1, \nu \rangle \langle \mathcal{L}_{y'_m y'_{m-1}} \cdots \mathcal{L}_{y'_{n+1} y'_n} \mathcal{L}_{y'_n y'_{n-1}} \cdots \mathcal{L}_{y'_1 y'_0} 1, \nu \rangle} \right| \\ &= \lim_{m \rightarrow \infty} \left| \log \frac{\langle \mathcal{L}_{y_n y_{n-1}} \cdots \mathcal{L}_{y_1 y_0} 1, \nu_{y,m} \rangle \langle 1, \nu_{y',m} \rangle}{\langle 1, \nu_{y,m} \rangle \langle \mathcal{L}_{y_n y_{n-1}} \cdots \mathcal{L}_{y_1 y_0} 1, \nu_{y',m} \rangle} \right| \\ &\leq \Theta_{\mathcal{C}(\text{out}, \bar{w}, +)}(\mathcal{L}_{y_n y_{n-1}} \cdots \mathcal{L}_{y_1 y_0} 1, 1). \end{aligned}$$

Notice that for any word $w = w_0 \cdots w_n$ admissible in Σ_A^+ and $x, x' \in [w_n]$ we have that

$$\begin{aligned} L_w 1(x) &= e^{S_n \varphi(w_0 \cdots w_{n-1} x) - S_n \varphi(w_0 \cdots w_{n-1} x') + S_n \varphi(w_0 \cdots w_{n-1} x')} \\ &\leq e^K L_w 1(x'). \end{aligned}$$

Where K is the constant in the Bowen property for φ . Thus for any word \bar{w} admissible in Y we have that $\mathcal{L}_{\bar{w}} 1 \in \mathcal{C}(\text{out}, \bar{w}, +B)$ (note that $K < 2 + \frac{K}{\sigma} = B$). Therefore restricting $\mathcal{L}_{\bar{w}} 1$ as necessary (in the same way as the paragraph

preceding equation (3.5)) we have that for $n \geq N + 1$

$$\Theta_{\mathcal{C}(\text{out}, \bar{w}, +)}(\mathcal{L}_{y_{n-1}y_{n-2}} \cdots \mathcal{L}_{y_{n-N}y_{n-N-1}}(\mathcal{L}_{y_{n-N-1}y_{n-N-2}} \cdots \mathcal{L}_{y_1y_0}1), 1) \leq C$$

where N and C are as in lemma 3.4.4. Hence ψ is Bowen.

2. Notice that for any probability measure η and cylinder set $[I]$ we have that

$$e^{-K}C_1\mu_\varphi([I]) \leq e^{-K} \sup_{z:Iz \in \Sigma_A} e^{S_n\varphi(Iz)} \leq \int e^{S_n\varphi(Iz)} d\eta(z) \leq \sup_{z:Iz \in \Sigma_A} e^{S_n\varphi(Iz)} \leq C_2\mu_\varphi([I]).$$

Thus

$$\begin{aligned} e^{S_n\psi(y)} &= \lim_{m \rightarrow \infty} \frac{\langle \mathcal{L}_{y_m y_{m-1}} \cdots \mathcal{L}_{y_{n+1} y_n} \mathcal{L}_{y_n y_{n-1}} \cdots \mathcal{L}_{y_1 y_0} 1, \nu \rangle}{\langle \mathcal{L}_{y_m y_{m-1}} \cdots \mathcal{L}_{y_{n+1} y_n} 1, \nu \rangle} \\ &= \lim_{m \rightarrow \infty} \left\langle \mathcal{L}_{y_n y_{n-1}} \cdots \mathcal{L}_{y_1 y_0} 1, \frac{\nu_{y,m}}{\langle 1, \nu_{y,m} \rangle} \right\rangle \\ &= \lim_{m \rightarrow \infty} \sum_{I:\pi(I)=y_0 \cdots y_{n-1}} \int_{\Sigma_A} e^{S_n\varphi(Iz)} d\left(\frac{\nu_{y,m}}{\langle 1, \nu_{y,m} \rangle}\right) \\ &\geq e^{-K}C_1\pi_*\mu_\varphi[y_0 \cdots y_{n-1}] \end{aligned}$$

where

$$\nu_{y,m} = \mathcal{L}_{y_{n+1}y_n}^* \cdots \mathcal{L}_{y_m y_{m-1}}^* \nu.$$

Similarly

$$e^{S_n\psi(y)} \leq C_2\pi_*\mu_\varphi[y_0 \cdots y_{n-1}].$$

Observe that this implies that the pressure of ψ is 0. The result is then simply an application of the Gibbs inequality for μ_ψ .

3. By 2 we have that μ_ψ and $\pi_*\mu$ are mutually absolutely continuous. As they are ergodic this implies they are equal.

□

Remark 4. Observe that if φ is the logarithm of a g -function then $\nu = \mu_\varphi$ and the function ψ is the logarithm of the g function for $\pi_*\mu_\varphi$. Thus the proof of theorem 3.1.2 will also yield corollary 3.1.3.

What remains to prove theorem 3.1.2 is the following.

Theorem 3.4.7. Suppose that φ is Walters (respectively Hölder). Then the function

$$\psi(y) = \lim_{m \rightarrow \infty} \log \frac{\langle \mathcal{L}_{y_m y_{m-1}} \cdots \mathcal{L}_{y_2 y_1} \mathcal{L}_{y_1 y_0} 1, \nu \rangle}{\langle \mathcal{L}_{y_m y_{m-1}} \cdots \mathcal{L}_{y_2 y_1} 1, \nu \rangle}$$

is Walters (respectively Hölder).

Proof. We will prove the result in the case that φ is Walters. The Hölder case is similar to the previous section [35]. Let $\alpha_k = \sup_{n \geq 1} \text{var}_{n+k} S_n \varphi$ which by the assumption that φ is Walters converges to 0 (and thus can be used as an input for procedure 3.4.1). Define the sequence $\{n_j\}$ based on procedure 3.4.1. Set d_j to be the sequences of metrics based on $\{\alpha_i\}$ and $\{n_j\}$ as in (3.3).

Let $\varepsilon > 0$. To show that ψ has the Walters property we need to show that there exists N such that for all $j \geq N$, $\text{var}_{n+j} S_n \psi < \varepsilon$ for all $n \geq 1$. To do so suppose that $n \geq 1$, $j \geq 0$ and that $y, y' \in Y$ are such that $y_i = y'_i$ for all $0 \leq i \leq n + j - 1$. In the same way as the proof of proposition 3.4.6 we have that

$$\begin{aligned} & |S_n \psi(y) - S_n \psi(y')| \\ &= \lim_{m \rightarrow \infty} \left| \log \frac{\langle \mathcal{L}_{y_{n+j-1} y_{n+j-2}} \cdots \mathcal{L}_{y_{n+1} y_n} \mathcal{L}_{y_n y_{n-1}} \cdots \mathcal{L}_{y_1 y_0} 1, \nu_{y,m} \rangle \langle \mathcal{L}_{y_{n+j-1} y_{n+j-2}} \cdots \mathcal{L}_{y_{n+1} y_n} 1, \nu_{y',m} \rangle}{\langle \mathcal{L}_{y_{n+j-1} y_{n+j-2}} \cdots \mathcal{L}_{y_{n+1} y_n} 1, \nu_{y,m} \rangle \langle \mathcal{L}_{y_{n+j-1} y_{n+j-2}} \cdots \mathcal{L}_{y_{n+1} y_n} \mathcal{L}_{y_n y_{n-1}} \cdots \mathcal{L}_{y_1 y_0} 1, \nu_{y',m} \rangle} \right| \end{aligned}$$

where

$$\nu_{y,m} = \mathcal{L}_{y_{n+j}y_{n+j-1}}^* \cdots \mathcal{L}_{y_m y_{m-1}}^* \nu$$

and similarly for $\nu_{y',m}$. Thus by lemma A.1.4

$$|S_n \psi(y) - S_n \psi(y')| \leq \Theta_+(\mathcal{L}_{y_{n+j-1}y_{n+j-2}} \cdots \mathcal{L}_{y_{n+1}y_n} (\mathcal{L}_{y_n y_{n-1}} \cdots \mathcal{L}_{y_1 y_0} 1), \mathcal{L}_{y_{n+j-1}y_{n+j-2}} \cdots \mathcal{L}_{y_{n+1}y_n} 1).$$

Let D and γ be as in lemma 3.4.5, take k such that $\gamma^{k-1}D < \varepsilon$ and set $N = 1 + \sum_{i=1}^k n_i$ (note N depends only on ε and not on n, y or y'). Assume that $j \geq N$. For notational purposes divide $y_n y_{n+1} \cdots y_{n+N}$ into words $\bar{w}^1 = y_n \cdots y_{n+n_1}$, $\bar{w}^2 = y_{n+n_1} \cdots y_{n+n_1+n_2}$ and so on. Notice that by our choice of the sequence $\{\alpha_k\}$ we have that for any $x \in \Sigma_A^+$

$$L_{x_n x_{n-1}} \cdots L_{x_1 x_0} 1 \in \mathcal{C}([x_n], d_0)$$

and thus

$$R_{\bar{w}_{\text{in}}^1} 1, R_{\bar{w}_{\text{in}}^1} \mathcal{L}_{y_n y_{n-1}} \cdots \mathcal{L}_{y_1 y_0} 1 \in \mathcal{C}(\text{in}, \bar{w}^1, d_0).$$

The picture to have in mind through the remainder of the proof is the following

$$\cdots \xleftarrow{R_{\bar{w}_{\text{in}}^3}} \mathcal{C}(\text{out}, \bar{w}^2, d_2) \xleftarrow{\mathcal{L}_{\bar{w}^2}} \mathcal{C}(\text{in}, \bar{w}^2, d_1) \xleftarrow{R_{\bar{w}_{\text{in}}^2}} \mathcal{C}(\text{out}, \bar{w}^1, d_1) \xleftarrow{\mathcal{L}_{\bar{w}^1}} \mathcal{C}(\text{in}, \bar{w}^1, d_0).$$

By lemma 3.4.5 each $L_{\bar{w}^i}$ contracts the relevant cones by a factor of γ and thus we have

$$|S_n \psi(y) - S_n \psi(y')| \leq \Theta_+(\mathcal{L}_{y_{n+j-1}y_{n+j-2}} \cdots \mathcal{L}_{y_{n+1}y_n} (\mathcal{L}_{y_n y_{n-1}} \cdots \mathcal{L}_{y_1 y_0} 1), \mathcal{L}_{y_{n+j-1}y_{n+j-2}} \cdots \mathcal{L}_{y_{n+1}y_n} 1)$$

$$\begin{aligned}
&\leq \Theta_{\mathcal{C}(\text{out}, \bar{w}^k, d_k)}(\mathcal{L}_{\bar{w}^k} \cdots \mathcal{L}_{\bar{w}^1}(\mathcal{L}_{y_n y_{n-1}} \cdots \mathcal{L}_{y_1 y_0} 1), \mathcal{L}_{\bar{w}^k} \cdots \mathcal{L}_{\bar{w}^1} 1) \\
&\leq \gamma^{k-1} \Theta_{\mathcal{C}(\text{in}, \bar{w}^2, d_1)}(R_{\bar{w}_{\text{in}}^2} \mathcal{L}_{\bar{w}^1}(\mathcal{L}_{y_n y_{n-1}} \cdots \mathcal{L}_{y_1 y_0} 1), R_{\bar{w}_{\text{in}}^2} \mathcal{L}_{\bar{w}^1} 1) \\
&\leq \gamma^{k-1} \Theta_{\mathcal{C}(\text{out}, \bar{w}^1, d_1)}(\mathcal{L}_{\bar{w}^1}(\mathcal{L}_{y_n y_{n-1}} \cdots \mathcal{L}_{y_1 y_0} 1), \mathcal{L}_{\bar{w}^1} 1) \leq \gamma^{k-1} D < \varepsilon.
\end{aligned}$$

Thus taking the supremum over y, y' we have that $\text{var}_{n+j} S_n \psi < \varepsilon$ and taking the supremum over n gives $\sup_{n \geq 1} \text{var}_{n+j} S_n \psi < \varepsilon$. Therefore $\sup_n \text{var}_{n+j} S_n \psi \xrightarrow{j \rightarrow \infty} 0$ and ψ is Walters. \square

3.5 Final remarks

1. Similar results hold on countable state topological Markov shifts with a strong “mixing in fibers” assumption on the factor map.
2. The theory of factors of Gibbs measures in dimension greater than 1 is well known to be filled with strange phenomena such as loss of Gibbsianness. This is in stark contrast to the results we have presented in this chapter. It is worth pointing out that the uniqueness regimes we have considered in this chapter do not exhibit many pathologies which are prevalent in higher dimensions. In particular they do not undergo phase transitions.
3. One interesting direction to move in an attempt to bridge the gap between our results and those in higher dimension would be to consider factors of Markov measures on countable state topological Markov shifts. These represent an intriguing class with respect to the current theory. On the one hand they can exhibit phase transitions and on the other they are fairly tractable objects with a well developed theory. It seems likely that one could understand in a fairly explicit way the existence of so called “hidden phase transitions” in these models and their connection to loss of Gibbsianness. We leave this for future work.

Appendix A

Cone Theory

A.1 Cones in $C(\Sigma_A^+)$

Definition A.1.1. Let V be a real vector space. A subset $\mathcal{C} \subseteq V$ is called a *cone* if

1. $a\mathcal{C} = \mathcal{C}$ for all $a > 0$.
2. \mathcal{C} is convex.
3. $\mathcal{C} \cap (-\mathcal{C}) = \emptyset$

if \mathcal{C} satisfies only (1) and (2) we say that \mathcal{C} is a *wedge*. We say that \mathcal{C} is a *closed cone* if $\mathcal{C} \cup \{0\}$ is closed.

Every cone induces a partial ordering on the vector space V by

$$x \leq_{\mathcal{C}} y \iff y - x \in \mathcal{C} \cup \{0\}.$$

For $x \in \mathcal{C}$ and $y \in V$ we say that x *dominates* y if there exists α, β such that

$$\alpha x \leq_{\mathcal{C}} y \leq_{\mathcal{C}} \beta x.$$

If x dominates y and y dominates x then x, y are said to be *comparable*. A cone \mathcal{C} is called *almost Archimedean* if whenever $x \in \mathcal{C}$ and $y \in V$ are such that $-\varepsilon x \leq_{\mathcal{C}} y \leq_{\mathcal{C}} \varepsilon x$ for all $\varepsilon > 0$ it follows that $y = 0$. If $x \geq_{\mathcal{C}} 0$ dominates y then define

$$m(y/x; \mathcal{C}) = \sup \{ \alpha : \alpha x \leq_{\mathcal{C}} y \}$$

$$M(y/x; \mathcal{C}) = \inf \{ \beta : y \leq_{\mathcal{C}} \beta x \}$$

if x and y are comparable we define

$$\Theta_{\mathcal{C}}(x, y) = \log \left(\frac{M(x/y; \mathcal{C})}{m(x/y; \mathcal{C})} \right).$$

This is known as *Hilbert's metric* and has been used in a number of different contexts. In dynamics it is most often used to prove upper bounds on the rate of convergence for transfer operators. This function $\Theta_{\mathcal{C}}$ is a projective psuedo metric in the following sense.

Proposition A.1.2. Let V be a real vector space and $\mathcal{C} \subseteq V$ a cone. Suppose that x, y, z are comparable and $a, b > 0$. Then

1. $\Theta_{\mathcal{C}}(x, y) \geq 0$.
2. $\Theta_{\mathcal{C}}(x, y) = \Theta_{\mathcal{C}}(y, x)$.
3. $\Theta_{\mathcal{C}}(x, y) \leq \Theta_{\mathcal{C}}(x, z) + \Theta_{\mathcal{C}}(z, y)$.
4. $\Theta_{\mathcal{C}}(ax, by) = \Theta_{\mathcal{C}}(x, y)$.

If \mathcal{C} is almost Archimedean then

$$\Theta_{\mathcal{C}}(x, y) = 0 \implies x = ay$$

for some $a > 0$.

Proof. These follow by direct computation. See for example [14, lemma 2.7]. \square

The true utility of Hilbert's metric is the following theorem.

Theorem A.1.3. (Birkhoff [1]) Let $\mathcal{C}_1, \mathcal{C}_2$ be closed cones and $L : V_1 \rightarrow V_2$ a linear map such that $L\mathcal{C}_1 \subseteq \mathcal{C}_2$. Then for all $\phi, \psi \in \mathcal{C}_1$

$$\Theta_{\mathcal{C}_2}(L\phi, L\psi) \leq \tanh\left(\frac{\text{diam}_{\mathcal{C}_2}(L\mathcal{C}_1)}{4}\right) \Theta_{\mathcal{C}_1}(\phi, \psi)$$

where

$$\text{diam}_{\mathcal{C}_2}(L\mathcal{C}_1) = \sup \{\Theta_{\mathcal{C}_2}(f, g) : f, g \in L\mathcal{C}_1\}$$

and $\tanh \infty = 1$.

Given a closed cone \mathcal{C} in a Banach space X we define its dual

$$\mathcal{C}^* = \{\psi \in X^* : \langle x, \psi \rangle \geq 0 \text{ for all } x \in \mathcal{C}\}$$

Notice that if $L\mathcal{C}_1 \subseteq \mathcal{C}_2$ then $L^*\mathcal{C}_2 \subseteq \mathcal{C}_1^*$. To see this suppose that $\psi \in \mathcal{C}_2^*$ and $x \in \mathcal{C}_1$ then

$$\langle x, L^*\psi \rangle = \langle Lx, \psi \rangle \geq 0.$$

A word of caution: in general \mathcal{C}^* is not a cone but a wedge.

Proposition A.1.4. Let \mathcal{C} be a closed cone and $x, y \in \mathcal{C}$ such that x and y are comparable. Then for any $\phi \in \mathcal{C}^*$, $\langle x, \phi \rangle = 0$ if and only if $\langle y, \phi \rangle = 0$. Furthermore

$$\Theta_{\mathcal{C}}(x, y) = \log \sup \left\{ \frac{\langle x, \phi \rangle \langle y, \psi \rangle}{\langle y, \phi \rangle \langle x, \psi \rangle} : \psi, \phi \in \mathcal{C}^* \text{ and } \langle y, \phi \rangle \langle x, \psi \rangle \neq 0 \right\}$$

Proof. The proof can be found in [13, Lemma 1.4]. □

Let Σ_A^+ be a topological Markov shift over a countable alphabet (we will work in this generality so that the industrious reader can convince themselves that under the appropriate assumptions the results about factors can be carried over to countable state spaces). Define

$$C_b(\Sigma_A^+) = \left\{ f \in C(\Sigma_A^+) : \sup_x |f(x)| < \infty \right\}.$$

When Σ_A^+ is a shift space it is often nice to think of $C_b(\Sigma_A)$ as a “ ℓ^∞ ” direct sum of the spaces

$$C([a]) = \left\{ f \in C_b(\Sigma_A^+) : f(x) = 0 \text{ for all } x \notin [a] \right\}$$

where a is a symbol in the alphabet. Recall that given a collection of Banach spaces $\{X_i\}_{i=0}^\infty$ we can define

$$\bigoplus_{i=0}^\infty X_i = \left\{ (a_i)_{i=0}^\infty : \sup_{i \geq 0} \|a_i\|_{X_i} < \infty \right\}$$

and

$$\|(a_i)_{i=0}^\infty\| = \sup_{i \geq 0} \|a_i\|_{X_i}.$$

This along with coordinate-wise addition and scalar multiplication is one of a number

of possible ways to take the direct sum of Banach spaces.

Proposition A.1.5. The function $\|(a_i)_{i=0}^\infty\| = \sup_{i \geq 0} \|a_i\|_{X_i}$ defines a norm on $\bigoplus_{i=0}^\infty X_i$ which makes $\bigoplus_{i=0}^\infty X_i$ into a Banach space.

It is not hard to see that

$$C_b(\Sigma_A^+) \cong \bigoplus_{a \in \Sigma} C([a]).$$

Moreover given a subset of symbols S we identify

$$\bigoplus_{a \in S} C([a]) = \left\{ f \in C_b(\Sigma_A) : f(x) = 0 \text{ for all } x \notin \bigcup_{a \in S} [a] \right\}.$$

We will also be concerned with cones which arise in the following way.

Proposition A.1.6. Suppose that $\{X_i\}_{i=0}^\infty$ is a sequence of Banach spaces and $\mathcal{C}_i \subseteq X_i$ are closed cones. Then

$$\bigoplus_{i=0}^\infty \mathcal{C}_i = \left\{ (x_i)_{i=0}^\infty \in \bigoplus_{i=0}^\infty X_i : x_i \in \mathcal{C}_i \cup \{0\} \right\} \setminus \{(0)_{i=0}^\infty\}$$

is a closed cone in $\bigoplus_{i=0}^\infty X_i$.

Proof. As addition and scalar multiplication are coordinate wise it is clear that $\bigoplus_{i=0}^\infty \mathcal{C}_i$ is a cone. To see that it is closed suppose that x is in the closure of $\bigoplus_{i=0}^\infty \mathcal{C}_i$ and take a net x^α converging to x . By the definition of the norm we have that for any i x_i^α converges to x_i , as $\mathcal{C}_i \cup \{0\}$ is closed this implies that $x_i \in \mathcal{C}_i \cup \{0\}$. Therefore $x \in \bigoplus_{i=0}^\infty \mathcal{C}_i$. \square

Lemma A.1.7. Suppose that $\{X_i\}_{i=0}^{\infty}$ is a set of Banach spaces and $\mathcal{C}_i \subseteq X_i$ are closed cones. Then

$$m\left(x/y; \bigoplus_i \mathcal{C}_i\right) = \inf_i m(x_i/y_i; \mathcal{C}_i)$$

and

$$M\left(x/y; \bigoplus_i \mathcal{C}_i\right) = \sup_i M(x_i/y_i; \mathcal{C}_i).$$

Thus

$$\Theta_{\bigoplus_i \mathcal{C}_i}(x, y) = \log\left(\sup_{i,j} \frac{M(x_i/y_i; \mathcal{C}_i)}{m(x_j/y_j; \mathcal{C}_j)}\right)$$

Proof. We will show this for m the proof for M is similar. Notice that

$$\begin{aligned} m\left(x/y; \bigoplus_i \mathcal{C}_i\right) &= \sup\{\alpha : \alpha x_i \leq_{\mathcal{C}_i} y_i \text{ for all } i\} \\ &= \sup \bigcap_i \{\alpha : \alpha x_i \leq_{\mathcal{C}_i} y_i\} \\ &= \sup \bigcap_i [0, m(x_i/y_i; \mathcal{C}_i)] \\ &= \inf_i m(x_i/y_i; \mathcal{C}_i) \end{aligned} \quad \square$$

For a metric space (Z, d) define the cones

$$C_b(Z)^+ = \{f \in C_b(Z) : f \geq 0\}$$

and

$$\mathcal{C}(Z, d) = \left\{f \in C_b(Z) : f > 0 \text{ and } f(x) \leq e^{d(x,y)} f(y) \text{ for all } x, y \in Z\right\}.$$

Lemma A.1.8. Assume that Z has no isolated points. For $f, g \in \mathcal{C}(Z, d)$ we have

$$m(g/f; \mathcal{C}(Z, d)) = \inf_{x, y \in Z, x \neq y} \frac{e^{d(x, y)} g(y) - g(x)}{e^{d(x, y)} f(y) - f(x)}$$

and

$$M(g/f; \mathcal{C}(Z, d)) = \sup_{x, y \in Z, x \neq y} \frac{e^{d(x, y)} g(y) - g(x)}{e^{d(x, y)} f(y) - f(x)}.$$

Proof. This is a mild generalization of the proof of lemma 2.2 in [29] we provide the details for the reader's convenience. We will compute $m(f/g; \mathcal{C}(Z, d))$, $M(f/g; \mathcal{C}(Z, d))$ is analogous. Notice $\alpha f \leq_{\mathcal{C}(Z, d)} g$ if and only if

$$\alpha \leq \frac{g(x)}{f(x)} \quad \forall x \in X \quad \text{and} \quad \alpha \leq \frac{e^{d(x, y)} g(y) - g(x)}{e^{d(x, y)} f(y) - f(x)} \quad \forall x \neq y.$$

Thus

$$\begin{aligned} m(g/f; \mathcal{C}(Z, d)) &= \sup \left\{ \alpha : \alpha f \leq_{\mathcal{C}(Z, d)} g \right\} \\ &= \min \left\{ \inf_{x \in Z} \frac{g(x)}{f(x)}, \inf_{x, y \in Z, x \neq y} \frac{e^{d(x, y)} g(y) - g(x)}{e^{d(x, y)} f(y) - f(x)} \right\}. \end{aligned}$$

So the claim is that

$$\inf_{x, y \in Z, x \neq y} \frac{e^{d(x, y)} g(y) - g(x)}{e^{d(x, y)} f(y) - f(x)} \leq \inf_{x \in Z} \frac{g(x)}{f(x)}.$$

Let $z \in Z$ and take $x \neq z$ such that $\frac{g(x)f(z)}{f(x)g(z)} \leq 1$. Note this can always be done unless $\frac{g}{f}$ has a unique minimum at z . Then

$$\inf_{x, y \in Z, x \neq y} \frac{e^{d(x, y)} g(y) - g(x)}{e^{d(x, y)} f(y) - f(x)} \leq \frac{e^{d(x, z)} g(x) - g(z)}{e^{d(x, z)} f(x) - f(z)}$$

$$\begin{aligned}
&= \frac{e^{d(x,z)} \frac{g(x)}{f(x)} f(x) - \frac{g(z)}{f(z)} f(z)}{e^{d(x,z)} f(x) - f(z)} \\
&= \frac{g(z)}{f(z)} \frac{e^{d(x,z)} \frac{g(x)f(z)}{f(x)g(z)} f(x) - f(z)}{e^{d(x,z)} f(x) - f(z)} \\
&\leq \frac{g(z)}{f(z)} \text{ as } \frac{g(x)f(z)}{f(x)g(z)} \leq 1.
\end{aligned}$$

If z is the unique minimum of $\frac{g}{f}$ then take a sequence which is not eventually z converging to z and the same inequality follows (hence the assumption that Z contains no isolated points). \square

We are concerned with the case that Z is a cylinder set in Σ_A^+ and d some metric. Given a metric and a number $\sigma > 0$ we can define a new metric by the function $(x, y) \mapsto \sigma d(x, y)$, the new metric is called σd . The next corollary will be one of our fundamental tools, it has been noticed previously, see for example [32] and [29], although we state it in a somewhat different way.

Corollary A.1.9. Suppose that $\{(Z_i, d_i)\}_i$ is a countable collection of metric spaces without isolated points. If $f, g \in \bigoplus_i \mathcal{C}(Z_i, \sigma d_i)$ then

$$\Theta_{\bigoplus_i \mathcal{C}(Z_i, d_i)}(f, g) \leq 2 \log \frac{1 + \sigma}{1 - \sigma} + \Theta_{\bigoplus_i \mathcal{C}(Z_i)}(f, g)$$

Proof. Notice that for and i, j we have that

$$\begin{aligned}
&\frac{M(f_i/g_i; \mathcal{C}(Z_i, d))}{m(f_j/g_j; \mathcal{C}(Z_j, d))} = \\
&\sup \left\{ \frac{e^{d(x,y)} g(y) - g(x)}{e^{d(x,y)} f(y) - f(x)} \cdot \frac{e^{d(w,z)} f(z) - f(w)}{e^{d(w,z)} g(z) - g(w)} : x, y \in Z_i, w, z \in Z_j, x \neq y, w \neq z \right\}
\end{aligned}$$

and

$$\frac{e^{d(x,y)} g_i(y) - g_i(x)}{e^{d(x,y)} f_i(y) - f_i(x)} \cdot \frac{e^{d(w,z)} f_j(z) - f_j(w)}{e^{d(w,z)} g_j(z) - g_j(w)}$$

$$\begin{aligned}
&= \frac{e^{d(x,y)} - \frac{g_i(x)}{g_i(y)}}{e^{d(x,y)} - \frac{f_i(x)}{f_i(y)}} \cdot \frac{e^{d(w,z)} - \frac{f_j(w)}{f_j(z)}}{e^{d(w,z)} - \frac{g_j(w)}{g_j(z)}} \cdot \frac{g_i(y)f_j(z)}{f_i(y)g_j(z)} \\
&\leq \frac{e^{d(x,y)} - e^{-\sigma d(x,y)}}{e^{d(x,y)} - e^{\sigma d(x,y)}} \cdot \frac{e^{d(w,z)} - e^{-\sigma d(w,z)}}{e^{d(w,z)} - e^{\sigma d(w,z)}} \cdot \frac{g_i(y)f_j(z)}{f_i(y)g_j(z)} \\
&= \frac{e^{d(x,y)}(1 - e^{-(1+\sigma)d(x,y)})}{e^{d(x,y)}(1 - e^{-(1-\sigma)d(x,y)})} \cdot \frac{e^{d(w,z)}(1 - e^{-(1+\sigma)d(w,z)})}{e^{d(w,z)}(1 - e^{-(1-\sigma)d(w,z)})} \cdot \frac{g_i(y)f_j(z)}{f_i(y)g_j(z)} \\
&= \frac{1 - e^{-(1+\sigma)d(x,y)}}{1 - e^{-(1-\sigma)d(x,y)}} \cdot \frac{1 - e^{-(1+\sigma)d(w,z)}}{1 - e^{-(1-\sigma)d(w,z)}} \cdot \frac{g_i(y)f_j(z)}{f_i(y)g_j(z)} \\
&\leq \frac{(1 + \sigma)d(x,y)}{(1 - \sigma)d(x,y)} \cdot \frac{(1 + \sigma)d(w,z)}{(1 - \sigma)d(w,z)} \cdot \frac{g_i(y)f_j(z)}{f_i(y)g_j(z)} \\
&= \frac{(1 + \sigma)}{(1 - \sigma)} \cdot \frac{(1 + \sigma)}{(1 - \sigma)} \cdot \frac{g_i(y)f_j(z)}{f_i(y)g_j(z)}
\end{aligned}$$

where in line 6 we have used that for $x \geq y > 0$

$$\frac{1 - e^{-x}}{1 - e^{-y}} \leq \frac{x}{y}. \quad \square$$

A semi-norm $|\cdot|$ is said to be *adapted* to a cone \mathcal{C} if

$$f + g \in \mathcal{C} \text{ and } g - f \in \mathcal{C} \implies |f| \leq |g|.$$

Lemma A.1.10. Suppose that $|\cdot|_1$ and $|\cdot|_2$ are semi norms adapted to a closed cone \mathcal{C} . If $|f|_1 = |g|_1 \neq 0$ then

$$|f - g|_2 \leq (e^{\Theta_{\mathcal{C}}(f,g)} - 1) |f|_2.$$

Proof. The proof is essentially that in [29], this exact statement appears in [32]. \square

A.2 Ruelle-Perron-Frobenius theorems

The Ruelle-Perron-Frobenius (RPF) theorem is one of the most studied results in dynamical systems. It has many variants for different types of dynamical systems. Here we will prove a version of the original RPF theorem and an extension of the original for potentials on one-sided shift spaces satisfying the Bowen property. This result is largely known but we include this section for two reasons. First as we have developed the machinery of cones in $C(\Sigma_A^+)$ we can give a relatively simple proof, and second it serves to point out that the Bowen property is the natural assumption required to make use of cones. In this section we will prove the following theorems.

Theorem A.2.1. Suppose that Σ is a set of symbols and $\varphi : \Sigma^{\mathbb{N}} \rightarrow \mathbb{R}$ is Hölder. Let L_φ be the Ruelle operator and assume without loss that L_φ has spectral radius $\rho(L_\varphi) = 1$ and let $0 < \theta < 1$ and $|\varphi|_\theta$ be as in the definition of Hölder.

1. There exists ν a probability measure on $\Sigma^{\mathbb{N}}$ and a Hölder continuous function $h > 0$ such that $L_\varphi^* \nu = \nu$, $L_\varphi h = h$ and $\int h d\nu = 1$.
2. Let

$$\mathcal{C} = \left\{ f \in C(\Sigma^{\mathbb{N}}) : f(x) > 0, f(x) \leq f(y) e^{K\theta k(x,y)} \right\}$$

where $k(x, y) = \min \{i \geq 0 : x_i \neq y_i\}$ and $K = \frac{|\varphi|_\theta}{\sigma - \theta}$ and $0 < \theta < \sigma < 1$. There exist constants $A > 0$ and $0 < \gamma < 1$ such that

$$\left\| L_\varphi^n f - \langle f, \nu \rangle h \right\|_\infty \leq A \langle f, \nu \rangle \gamma^n$$

for all $f \in \mathcal{C}$ and $n \geq 0$.

Theorem A.2.2. Suppose that φ has the Bowen property and Σ_A^+ is a one-sided topologically mixing shift of finite type. Let L_φ be the Ruelle operator and assume without loss that $\rho(L_\varphi) = 1$.

1. There exists ν a probability measure on Σ_A and a measurable function $h > 0$ such that $L_\varphi^* \nu = \nu$, $L_\varphi h = h$ and $\int h d\nu = 1$.
2. For any probability measure η

$$\frac{(L_\varphi^*)^n \eta}{\langle 1, (L_\varphi^*)^n \eta \rangle} \xrightarrow[\text{weak}^*]{n \rightarrow \infty} \nu.$$

In particular ν is the unique probability measure for which $L_\varphi^* \nu = \nu$.

3. If $h \in C(\Sigma_A^+)$ then for all $f \in C(\Sigma_A^+)$

$$\|L_\varphi^n f - \langle f, \nu \rangle h\|_\infty \xrightarrow{n \rightarrow \infty} 0.$$

As far as the author is aware it is unknown if there are examples in which the function h is not continuous, see [43, section 5] for some discussion of this. Fundamentally this is a question of whether or not the sequence $\{L_\varphi^n 1\}_{n=1}^\infty$ is equicontinuous. This is assured if for instance the function φ satisfies the Walters property but it is unclear if it holds for potentials with Bowen property.

Proof (of Theorem A.2.1). 1. The existence of ν is standard the proof is sketched at the beginning of the proof of Theorem A.2.3. Next we claim that $L_\varphi^n 1$ is Cauchy. By the same argument as lemma 3.3.1 (2) we have that $D = \text{diam}_C(L_\varphi \mathcal{C}) < \infty$. Set $\gamma = \tanh(D/4)$. Notice that the semi-norms $\|\cdot\|_\infty$ and $|\langle \cdot, \nu \rangle|$ are adapted to $C(\Sigma_A^+)$ and that $\langle L_\varphi^n 1, \nu \rangle = 1$ for all $n \geq 0$. Suppose

that $n \leq m$

$$\begin{aligned} \|L_\varphi^n 1 - L_\varphi^m 1\|_\infty &\leq \|L_\varphi^n 1\|_\infty (e^{\theta_+(L_\varphi^n 1, L_\varphi^m 1)} - 1) \text{ by lemma A.1.10} \\ &\leq C(e^{\theta_+(L_\varphi^n 1, L_\varphi^m 1)} - 1) \text{ by equation (A.1)} \\ &\leq C(e^{\gamma^{n-1}D} - 1). \end{aligned}$$

Take $h = \lim_{n \rightarrow \infty} L_\varphi^n 1$.

2. Again as $\|\cdot\|_\infty$ and $|\langle \cdot, \nu \rangle|$ are adapted to $C(\Sigma_A^+)^+$ we have that

$$\begin{aligned} \|L_\varphi^n f - \langle f, \nu \rangle h\|_\infty &= \langle f, \nu \rangle \left\| L_\varphi^n \frac{f}{\langle f, \nu \rangle} - h \right\|_\infty \\ &\leq \langle f, \nu \rangle \|h\|_\infty (e^{\theta_+(L_\varphi^n f, h)} - 1) \text{ by lemma A.1.10} \\ &\leq C \|h\|_\infty \langle f, \nu \rangle \theta_+(L_\varphi^n f, h) \\ &\leq C \|h\|_\infty \langle f, \nu \rangle \theta_C(L_\varphi^n f, h) \\ &\leq C \|h\|_\infty \langle f, \nu \rangle \gamma^{n-1} D. \quad \square \end{aligned}$$

The convergence properties in theorem A.2.2 will result from the following lemma.

Lemma A.2.3. Suppose that $\Sigma_A \subseteq \Sigma^{\mathbb{N}}$ is the topologically mixing shift of finite type defined by the matrix A and that φ is Bowen. Let $\varepsilon > 0$. For any $f, g \in C(\Sigma_A^+)$ such that $f, g > 0$ there exists an N such that

$$\Theta_+(L_\varphi^n f, L_\varphi^n g) < \varepsilon$$

for all $n \geq N$.

Proof. Take M such that $A^M > 0$ and $\alpha_n = \max\{\text{var}_n \log f, \text{var}_n \log g\}$. Construct a sequence of numbers $\{n_j\}_{j=0}^\infty$ as in section 3.4 such that $n_j \geq M$ for all j and metrics

$\{d_j\}_{j=0}^\infty$ as in equation (3.3). As in section 3.4 define the cone

$$\mathcal{C}([a], d_j) = \left\{ f \in C([a]) : f(x) \leq f(y)e^{d_j(x,y)} \text{ for all } x, y \in [a] \right\}$$

which for any j and a is contained in the cone

$$\mathcal{C}([a], +B) = \left\{ f \in X_a : f \geq 0 \text{ and } f(x) \leq e^B f(y) \right\}$$

where $B = \frac{2+K}{\sigma}$. This is where the Bowen property is crucial. Finally take

$$\begin{aligned} \mathcal{C}_j &= \bigoplus_{a \in \Sigma} \mathcal{C}([a], d_j) \\ &= \left\{ f \in C(\Sigma_A) : f \geq 0, f(x) \leq f(y)e^{d_j(x,y)} \text{ for all } x, y \in \Sigma_A \text{ with } x_0 = y_0 \right\}. \end{aligned}$$

By an argument similar to lemma 3.4.2 we have that $L_\varphi^{n_{j+1}}$ maps $\bigoplus_{a \in \Sigma} \mathcal{C}([a], d_j)$ to $\bigoplus_{a \in \Sigma} \mathcal{C}([a], \sigma d_{j+1})$. Thus by corollary A.1.9 we have that for any $f_1, f_2 \in \mathcal{C}_j$

$$\begin{aligned} \Theta_{\mathcal{C}_{j+1}}(L_\varphi^{n_{j+1}} f_1, L_\varphi^{n_{j+1}} f_2) &\leq 2 \log \frac{1+\sigma}{1-\sigma} + \Theta_+(L_\varphi^{n_{j+1}} f_1, L_\varphi^{n_{j+1}} f_2) \\ &\leq 2 \log \frac{1+\sigma}{1-\sigma} + \Theta_+(L_\varphi^M f_1, L_\varphi^M f_2). \end{aligned}$$

Notice $f_1 \in \mathcal{C}_j \subseteq \bigoplus_{a \in \Sigma} \mathcal{C}([a], +B)$. Let $x \in \Sigma_A^+$. Take z such that $f_1(z) = \|f_1\|_\infty$. By the assumption that Σ_A^+ is topologically mixing there exists an I such that $i_0 = z_0$, $|I| = M$, and $Ix \in \Sigma_A^+$ thus

$$L^M f_1(x) = \sum_{I: |I|=M, Ix \in \Sigma_A^+} e^{S_M \varphi(Ix)} f(Ix) \geq e^{-M\|\varphi\|} \|f_1\|_\infty e^{-B}$$

Thus

$$\Theta_+(L_\varphi^M f_1, 1) \leq \log \left(\frac{\|L_\varphi^M\|_{\text{op}}}{e^{-B-M\|\varphi\|_\infty}} \right)$$

and therefore

$$\Theta_{\mathcal{C}_{j+1}}(L_\varphi^{n_{j+1}} f_1, L_\varphi^{n_{j+1}} f_2) \leq 2 \log \frac{1+\sigma}{1-\sigma} + 2 \log \left(\frac{\|L_\varphi^M\|_{\text{op}}}{e^{-B-M\|\varphi\|_\infty}} \right).$$

Set

$$D = 2 \log \frac{1+\sigma}{1-\sigma} + 2 \log \left(\frac{\|L_\varphi^M\|_{\text{op}}}{e^{-B-M\|\varphi\|_\infty}} \right) \text{ and } \gamma = \tanh \left(\frac{D}{4} \right).$$

Take k such that $\gamma^{k-1}D < \varepsilon$ and set $N = \sum_{i=1}^k n_i$ if $n \geq N$ then by theorem A.1.3 we have that

$$\Theta_+(L_\varphi^n f, L_\varphi^n g) \leq \Theta_+(L_\varphi^N f, L_\varphi^N g) \leq \Theta_{\mathcal{C}_k}(L_\varphi^N f, L_\varphi^N g) \leq \gamma^{k-1} \Theta_{\mathcal{C}_1}(L_\varphi^{n_1} f, L_\varphi^{n_1} g) \leq \gamma^{k-1} D < \varepsilon.$$

□

Proof (of Theorem A.2.2). 1. This is standard see [43]. We will sketch the details.

Define the map

$$\eta \mapsto \frac{L_\varphi^* \eta}{\langle 1, L_\varphi^* \eta \rangle}$$

and notice that this is weak* continuous. By Schauder-Tychonoff there exists a probability measure ν with $L_\varphi^* \nu = \langle 1, L_\varphi^* \nu \rangle \nu$. Notice that

$$\log \langle 1, L_\varphi^* \nu \rangle = \frac{1}{n} \log \langle 1, L_\varphi^* \nu \rangle^n = \frac{1}{n} \log \langle 1, (L_\varphi^*)^n \nu \rangle = \frac{1}{n} \log \langle L_\varphi^n 1, \nu \rangle.$$

It follows from the Bowen property that

$$e^{-K} \|L_\varphi^n\|_{\text{op}} \leq \langle L_\varphi^n \mathbf{1}, \nu \rangle \leq \|L_\varphi^n\|_{\text{op}}. \quad (\text{A.1})$$

Thus

$$\log \langle \mathbf{1}, L_\varphi^* \nu \rangle = \lim_{n \rightarrow \infty} \frac{1}{n} \log \langle L_\varphi^n \mathbf{1}, \nu \rangle = \frac{1}{n} \log \|L_\varphi^n\|_{\text{op}} = \log \rho(L_\varphi) = 0$$

in other words $\langle \mathbf{1}, L_\varphi^* \nu \rangle = 1$. Define

$$h_1(x) = \limsup_{n \rightarrow \infty} L_\varphi^n \mathbf{1}(x).$$

The Bowen property implies that there exists a constant $D > 0$ such that $D^{-1} \leq L_\varphi^n \mathbf{1}(x) \leq D$ for all $x \in \Sigma_A$, see [43, theorem 2.16]. Notice that

$$\begin{aligned} L_\varphi h_1(x) &= \sum_{i:ix \in \Sigma_A} e^{\varphi(ix)} \limsup_{n \rightarrow \infty} L_\varphi^n \mathbf{1}(ix) \\ &\geq \limsup_{n \rightarrow \infty} \sum_{i:ix \in \Sigma_A} e^{\varphi(ix)} L_\varphi^n \mathbf{1}(ix) \\ &= \limsup_{n \rightarrow \infty} L_\varphi^{n+1} \mathbf{1}(x) \\ &= h_1(x) \end{aligned}$$

Set $h(x) = \lim_{n \rightarrow \infty} L_\varphi^n h_1(x)$ which exists by monotone convergence and normalize so that $\int h d\nu = 1$.

2. By proposition A.1.4 we have that for any $f \in C(\Sigma_A)$ with $f \geq 0$

$$\frac{\left\langle f, \frac{(L_\varphi^*)^n \eta}{\langle \mathbf{1}, (L_\varphi^*)^n \eta \rangle} \right\rangle}{\langle f, \nu \rangle} = \frac{\langle \mathbf{1}, \nu \rangle \langle f, (L_\varphi^*)^n \eta \rangle}{\langle f, \nu \rangle \langle \mathbf{1}, (L_\varphi^*)^n \eta \rangle}$$

$$\begin{aligned}
&= \frac{\langle 1, (L_\varphi^*)^n \nu \rangle \langle f, (L_\varphi^*)^n \eta \rangle}{\langle f, (L_\varphi^*)^n \nu \rangle \langle 1, (L_\varphi^*)^n \eta \rangle} \\
&= \frac{\langle L_\varphi^n 1, \nu \rangle \langle L_\varphi^n f, \eta \rangle}{\langle L_\varphi^n f, \nu \rangle \langle L_\varphi^n 1, \eta \rangle} \\
&\leq e^{\Theta_+(L_\varphi^n f, L_\varphi^n 1)}
\end{aligned}$$

and similarly

$$\frac{\langle f, \nu \rangle}{\left\langle f, \frac{(L_\varphi^*)^n \eta}{\langle 1, (L_\varphi^*)^n \eta \rangle} \right\rangle} \leq e^{\Theta_+(L_\varphi^n f, L_\varphi^n 1)} \implies \frac{\left\langle f, \frac{(L_\varphi^*)^n \eta}{\langle 1, (L_\varphi^*)^n \eta \rangle} \right\rangle}{\langle f, \nu \rangle} \geq e^{-\Theta_+(L_\varphi^n f, L_\varphi^n 1)}.$$

Therefore by lemma A.2.3 we have that

$$\left\langle f, \frac{(L_\varphi^*)^n \eta}{\langle 1, (L_\varphi^*)^n \eta \rangle} \right\rangle \xrightarrow{n \rightarrow \infty} \langle f, \nu \rangle.$$

It follows that $\frac{(L_\varphi^*)^n \eta}{\langle 1, (L_\varphi^*)^n \eta \rangle}$ converges weak* to ν , as $\frac{(L_\varphi^*)^n \eta}{\langle 1, (L_\varphi^*)^n \eta \rangle}$ is bounded and $\text{span} \{f \in C(\Sigma_A^+) : f > 0\}$ is dense.

3. We have that

$$\begin{aligned}
\|L_\varphi^n f - \langle f, \nu \rangle h\|_\infty &= \langle f, \nu \rangle \left\| L_\varphi^n \frac{f}{\langle f, \nu \rangle} - h \right\|_\infty \\
&\leq \langle f, \nu \rangle \|h\|_\infty (e^{\Theta_+(L_\varphi^n f, h)} - 1) \text{ by lemma A.1.10} \\
&\xrightarrow{n \rightarrow \infty} 0 \text{ by lemma A.2.3} \quad \square
\end{aligned}$$

A.3 Finite dimensional Perron-Frobenius Theory

Here we collect some facts as well as the basic definitions and properties of cones in finite dimensional vector spaces. Most of material on cones will be familiar from the

classical Perron-Frobenius theory for non-negative matrices. In addition we collect some elementary propositions and lemmas which we use in the thesis.

Definition A.3.1. Let $A : \mathbb{R}^d \rightarrow \mathbb{R}^d$ be a linear map and $K \subseteq \mathbb{R}^d$ be a cone.

- We say that A is K -non-negative provided $AK \subseteq K$ and we write $A \geq^K 0$.
- We say that A is K -positive if $A(K \setminus \{0\}) \subseteq \text{int}(K)$ and we write $A >^K 0$.
- We say A is K -primitive if $A \geq^K 0$ and there exists an N such that A^N is K -positive.
- We say $A0$ is K -irreducible if $A \geq^K 0$ and $\sum_{k=0}^{d-1} A^k$ is K -positive.

Often when K is understood we suppress the notation.

There are various definitions of K -irreducible in the literature. It is known that these conditions are all equivalent however finding a complete proof in the literature is surprising difficult. Thus for the sake of completeness we include we prove these equivalences in Proposition A.3.3.

Lemma A.3.2. Suppose that $K \subseteq \mathbb{R}^d$ is a cone. Then

$$\text{int}(K) = \{u : \langle u, v \rangle > 0 \text{ for all } v \in K^* \setminus \{0\}\}.$$

Proof. First we recall that

$$K = \{u : \langle u, v \rangle \geq 0 \text{ for all } v \in K^* \text{ with } \|v\| = 1\}$$

this is a very general fact for closed cones in Banach spaces which follows from a suitable Hahn-Banach theorem, see [13] for a nice discussion. Let u be such that $\langle u, v \rangle > 0$ for all $v \in K^* \setminus \{0\}$. Now take $\delta > 0$ such that $\langle u, v \rangle > \delta$ for all $v \in K^*$

with $\|v\| = 1$. Suppose that $\|w - u\| < \delta/2$ then for any $v \in K^*$ with $\|v\| = 1$ we have

$$|\langle u, v \rangle - \langle w, v \rangle| = |\langle u - w, v \rangle| \leq \|w - u\| < \delta/2.$$

This implies that $\langle w, v \rangle \geq \delta/2 > 0$ and thus $w \in K$. As $B(u, \delta/2) \subseteq K$ we have that $u \in \text{int}(K)$.

Now suppose that $u \in \text{int}(K)$. Take $\delta > 0$ such that $B(u, \delta) \subseteq K$ if $\|w\| = \delta/2$ then

$$\|(u + w) - u\| = \|w\| < \delta \text{ and } \|(u - w) - u\| = \|w\| < \delta$$

implies that $u - w, u + w \in K$. Thus for any $v \in K^*$

$$0 \leq \langle u + w, v \rangle = \langle u, v \rangle + \langle w, v \rangle \text{ and } 0 \leq \langle u - w, v \rangle = \langle u, v \rangle - \langle w, v \rangle$$

which implies

$$-\langle w, v \rangle \leq \langle u, v \rangle \leq \langle w, v \rangle.$$

Then we have

$$\|v\| = (2/\delta) \sup_{\|w\|=\delta/2} |\langle w, v \rangle| \leq (2/\delta) \langle u, v \rangle.$$

In particular if $v \neq 0$ then $\langle u, v \rangle > 0$. □

Proposition A.3.3. Suppose that A preserves a non-void cone $K \subseteq \mathbb{R}^d$. The following are equivalent:

1. A has no eigenvector contained in ∂K .

2. A has no invariant faces. (See [40] for the definition of a face.)
3. $(I + A)^{d-1}$ is K -positive.
4. $\sum_{k=0}^{d-1} A^k$ is K -positive. That is A is K -irreducible in the sense of Definition A.3.1.

Proof. (1 \iff 2) is [40, Theorem 4.1].

(2 \implies 3) is [40, Lemma 4.2].

(4 \implies 1) is clear if A has an eigenvector contained in ∂K then so does $\sum_{k=0}^{d-1} A^k$.

(3 \implies 4) Suppose that $u \in K \setminus \{0\}$ by the assumption that $(I + A)^{d-1} \succ^K 0$ for any $v \in K^* \setminus \{0\}$ we have that

$$0 < \langle (I + A)^{d-1} u, v \rangle = \sum_{k=0}^{d-1} \binom{d-1}{k} \langle A^k u, v \rangle$$

by lemma A.3.2. This implies that

$$0 < \sum_{k=0}^{d-1} \langle A^k u, v \rangle = \left\langle \left(\sum_{k=0}^{d-1} A^k \right) u, v \right\rangle$$

and hence $\sum_{k=0}^{d-1} A^k u \in \text{int}(K)$ by Lemma A.3.2. So $\sum_{k=0}^{d-1} A^k$ is K -positive. \square

The Perron-Frobenius theorem holds for abstract finite dimensional cones just as it does for the positive orthant.

Theorem A.3.4. Suppose that $A : \mathbb{R}^d \rightarrow \mathbb{R}^d$ is linear and K is a closed cone with non-void interior. Let $\{\lambda_i\}_i$ denote the eigenvalues of A ordered such that we have $\rho(A) = |\lambda_1| \geq |\lambda_2| \geq \dots$.

1. If A is K -non-negative then

- (a) $\rho(A)$ is an eigenvalue.
- (b) K contains an eigenvector corresponding to $\rho(A)$.

2. If A is K -irreducible then

- (a) $\rho(A)$ is a simple eigenvalue, and any other eigenvalue with the same modulus is simple.
- (b) Suppose that u is an eigenvector for A corresponding to $\rho(A)$ and v is an eigenvector of A^T corresponding to $\rho(A)$ normalized so that $\langle u, v \rangle = 1$.

Then

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \rho(A)^{-k} A^k = P$$

where $P = uv^T$, that is, $Pw = \langle w, v \rangle u$.

3. If A is K -primitive then

- (a) $\rho(A)$ is a simple eigenvalue, which is greater in modulus than any other eigenvalue.
- (b) Suppose that u is an eigenvector for A corresponding to $\rho(A)$ and v is an eigenvector of A^T corresponding to $\rho(A)$ normalized so that $\langle u, v \rangle = 1$.

Then for all small $\varepsilon > 0$ there exists $C > 0$ such that for all $n \geq 0$

$$\left\| \rho(A)^{-n} A^n - P \right\| \leq C \left(\frac{|\lambda_2| + \varepsilon}{\rho(A)} \right)^n$$

where $Pw = \langle w, v \rangle u$.

Proof. 1. This is contained in [40, Theorem 3.1].

2. Part (a) can be found in [40, Theorem 4.3]. Part (b) follows from (a) using the same proof as for non-negative matrices, which can be found in [22, Theorem 8.6.1].
3. This result is well known. It can be proved for example using Hilbert's projective metric and holds in significant generality; see for example [13, theorem 2.3] (although the result in [13] is significantly more powerful than needed here). The article [40] contains a proof of this result when A is assumed K -positive.

□

It is clear from the definition of irreducibility and primitivity that the eigenvector corresponding to $\rho(A)$ is contained in the interior of the cone K . This agrees with the fact from classical Perron-Frobenius theory the the eigenvector has all positive entries. Notice also that for K -irreducible/primitive matrices there always exist vectors u and v with u an eigenvector for A corresponding to $\rho(A)$ and v is an eigenvector for A^T corresponding to $\rho(A)$ such that $\langle u, v \rangle = 1$ (by 1(b) and the observation that $u \in \text{int}(K)$ by irreducibility/primitivity). This ensures that 2(b) and 3(b) are never vacuous.

We need the following to produce the Gibbs inequality.

Lemma A.3.5. Suppose that K is a cone and that $D \subset \text{int}(K)$ and $D^* \subset \text{int}(K^*)$ are non-empty and compact. Then there exists a constant $C > 0$ such that

$$C^{-1} \|A\| \leq \langle Au, v \rangle \leq C \|A\|$$

for all $u \in D$, $v \in D^*$, and $A \geq^K 0$.

Proof. Suppose $A \geq^K 0$, and that for some $u \in \text{int}(K)$ and $v \in \text{int}(K^*)$ we have $\langle Au, v \rangle = 0$. Then $Au = 0$ by an argument similar to Lemma A.3.2. Thus for any

$w \in K^*$ we have that $\langle u, A^*w \rangle = 0$ which implies that $A^*w = 0$ by lemma A.3.2. Therefore $A^* = 0$ and of course $A = 0$. Thus the function

$$(A, u, v) \mapsto \langle Au, v \rangle$$

is continuous and $\langle Au, v \rangle > 0$ for all $A \geq^K 0$, $A \neq 0$ and $u \in D, v \in D^*$. As the set of norm one K non-negative matrices cross $D \times D^*$ is compact we can find a $C > 0$ such that

$$C^{-1} \leq \left\langle \frac{Au}{\|A\|}, v \right\rangle \leq C$$

for all $A \geq^K 0$, $A \neq 0$ and $u \in D, v \in D^*$. Clearly the inequality holds for $A = 0$ hence we have the result. \square

We will use the proceeding lemma to relate the definitions of irreducible and primitive for a collection of matrices in chapter 2 to irreducibility/ primitivity of the operator $L_{\mathcal{A}}$ from example 2.2.3.

Proposition A.3.6. Let $\mathcal{A} = (A_0, \dots, A_{M-1}) \in M_d(\mathbb{R})$ and define $L_{\mathcal{A}}$ as in example 2.2.3, that is, $L_{\mathcal{A}}B = \sum_i A_i B A_i^*$.

1. If $L_{\mathcal{A}}$ is irreducible then \mathcal{A} satisfies inequality (2.3).
2. If $L_{\mathcal{A}}$ is primitive then \mathcal{A} satisfies inequality (2.4).

Proof. We will prove (2); then (1) will be similar. Let L_i be as in example 2.2.3. Take N such that $L_{\mathcal{A}}^N >^K 0$ and U, V positive definite matrices. Set $D = \{U\}$ and $D^* = (L_{\mathcal{A}}^*)^N(\{W \in K^* : \langle U, W \rangle_{\text{HS}} = 1\})$. Notice that $\{W \in K^* : \langle U, W \rangle_{\text{HS}} = 1\}$ is closed and bounded (by lemma A.3.5) hence compact. Take $C > 0$ as in lemma

A.3.5. Then

$$\begin{aligned}
\sum_{|K|=N} \|A_I A_K A_J\|^2 &\geq C^{-1} \langle L_I L_{\mathcal{A}}^N L_J U, V \rangle_{\text{HS}} \\
&= C^{-1} \langle L_I U, V \rangle_{\text{HS}} \left\langle L_J U, (L_{\mathcal{A}}^*)^N \frac{L_I^* V}{\langle L_I U, V \rangle} \right\rangle_{\text{HS}} \\
&\geq C^{-3} \|A_I\|^2 \|A_J\|^2,
\end{aligned}$$

where we have used the fact that $\|L_I\| = \|A_I\|^2$. Therefore

$$\sum_{|K|=N} \|A_I A_K A_J\| \geq \left(\sum_{|K|=N} \|A_I A_K A_J\|^2 \right)^{1/2} \geq C^{-3/2} \|A_I\| \|A_J\| \square$$

Proposition A.3.7. Let k be an even number and define

$$S = \text{span} \{v^{\otimes k} : v \in \mathbb{R}^d\}$$

and

$$K = \left\{ w \in S^* : \langle v^{\otimes k}, w \rangle_{(\mathbb{R}^d)^{\otimes k}} \geq 0 \text{ for all } v \in \mathbb{R}^d \right\}.$$

Then K is a closed cone with non-void interior.

Proof. That K is a closed cone is trivial. Thus we turn our attention to showing that K has a non-void interior. First we note that there exist elements $w \in K$ such that $\langle v^{\otimes k}, w \rangle > 0$ for all $v \in \mathbb{R}^d \setminus \{0\}$. For example define a multi-linear map $f : (\mathbb{R}^d)^k \rightarrow \mathbb{R}$ by

$$f(v^1, v^2, \dots, v^k) = \sum_{i=1}^d v_i^1 v_i^2 \cdots v_i^k$$

this gives a linear map $f : (\mathbb{R}^d)^{\otimes k} \rightarrow \mathbb{R}$ such that

$$f(v^{\otimes k}) = \sum_{i=1}^d v_i^k > 0.$$

Now if $v^n \xrightarrow{n \rightarrow \infty} w$ then

$$\begin{aligned} (v^n)^{\otimes k} &= \sum_{i_1 \cdots i_k} v_{i_1}^n \cdots v_{i_k}^n e_{i_1} \otimes \cdots \otimes e_{i_k} \\ &\xrightarrow{n \rightarrow \infty} \sum_{i_1 \cdots i_k} w_{i_1} \cdots w_{i_k} e_{i_1} \otimes \cdots \otimes e_{i_k} \\ &= w^{\otimes k}. \end{aligned}$$

Thus $\{v^{\otimes k} : \|v^{\otimes k}\| = 1\}$ is compact. Take $\delta > 0$ such that $f(v^{\otimes k}) > \delta$ for all v for which $\|v^{\otimes k}\| = 1$. Now if $g \in S^*$ is such that $\|f - g\| < \delta/2$ then $g \in K$. Hence $\text{int}(K) \neq \emptyset$. \square

Proposition A.3.8. Suppose that (A_0, \dots, A_{M-1}) is irreducible. Then $(A_0^*, \dots, A_{M-1}^*)$ is also irreducible.

Proof. Notice that if $A_i^*W \subseteq W$ then $A_iW^\perp \subseteq W^\perp$. To see this consider for any $u \in W^\perp$ and $w \in W$ we have

$$0 = \langle A_i^*w, u \rangle = \langle w, A_iu \rangle$$

which implies that $A_iu \in W^\perp$. If $A_iW \subseteq W$ for all $0 \leq i \leq M-1$ then $W^\perp = \{0\}$ or \mathbb{R}^d hence $W = \{0\}$ or \mathbb{R}^d . \square

Bibliography

- [1] Garrett Birkhoff. Extensions of Jentzsch's theorem. *Trans. Amer. Math. Soc.*, 85:219–227, 1957.
- [2] Jairo Bochi and Ian D. Morris. Equilibrium states of generalised singular value potentials and applications to affine iterated function systems. *Geom. Funct. Anal.*, 28(4):995–1028, 2018.
- [3] Rufus Bowen. Markov partitions for Axiom A diffeomorphisms. *Amer. J. Math.*, 92:725–747, 1970.
- [4] Rufus Bowen. Symbolic dynamics for hyperbolic flows. *Amer. J. Math.*, 95:429–460, 1973.
- [5] Rufus Bowen. Bernoulli equilibrium states for Axiom A diffeomorphisms. *Math. Systems Theory*, 8(4):289–294, 1974/75.
- [6] Rufus Bowen. *Equilibrium states and the ergodic theory of Anosov diffeomorphisms*, volume 470 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, revised edition, 2008. With a preface by David Ruelle, Edited by Jean-René Chazottes.
- [7] Mike Boyle and Karl Petersen. Hidden Markov processes in the context of symbolic dynamics. In *Entropy of hidden Markov processes and connections to dy-*

- dynamical systems*, volume 385 of *London Math. Soc. Lecture Note Ser.*, pages 5–71. Cambridge Univ. Press, Cambridge, 2011.
- [8] Richard C. Bradley. Basic properties of strong mixing conditions. A survey and some open questions. *Probab. Surv.*, 2:107–144, 2005. Update of, and a supplement to, the 1986 original.
- [9] Richard C. Bradley, Jr. On the ψ -mixing condition for stationary random sequences. *Trans. Amer. Math. Soc.*, 276(1):55–66, 1983.
- [10] Yong-Luo Cao, De-Jun Feng, and Wen Huang. The thermodynamic formalism for sub-additive potentials. *Discrete Contin. Dyn. Syst.*, 20(3):639–657, 2008.
- [11] Jean-René Chazottes and Edgardo Ugalde. Projection of Markov measures may be Gibbsian. *J. Stat. Phys.*, 111(5-6):1245–1272, 2003.
- [12] Jean-René Chazottes and Edgardo Ugalde. On the preservation of Gibbsianness under symbol amalgamation. In *Entropy of hidden Markov processes and connections to dynamical systems*, volume 385 of *London Math. Soc. Lecture Note Ser.*, pages 72–97. Cambridge Univ. Press, Cambridge, 2011.
- [13] Simon P. Eveson and Roger D. Nussbaum. Applications of the Birkhoff-Hopf theorem to the spectral theory of positive linear operators. *Math. Proc. Cambridge Philos. Soc.*, 117(3):491–512, 1995.
- [14] Simon P. Eveson and Roger D. Nussbaum. An elementary proof of the Birkhoff-Hopf theorem. *Math. Proc. Cambridge Philos. Soc.*, 117(1):31–55, 1995.
- [15] Kenneth J. Falconer. The Hausdorff dimension of self-affine fractals. *Math. Proc. Cambridge Philos. Soc.*, 103(2):339–350, 1988.

- [16] De-Jun Feng. The limited Rademacher functions and Bernoulli convolutions associated with Pisot numbers. *Adv. Math.*, 195(1):24–101, 2005.
- [17] De-Jun Feng. Equilibrium states for factor maps between subshifts. *Adv. Math.*, 226(3):2470–2502, 2011.
- [18] De-Jun Feng and Antti Käenmäki. Equilibrium states of the pressure function for products of matrices. *Discrete Contin. Dyn. Syst.*, 30(3):699–708, 2011.
- [19] De-Jun Feng and Ka-Sing Lau. The pressure function for products of non-negative matrices. *Math. Res. Lett.*, 9(2-3):363–378, 2002.
- [20] N. A. Friedman and D. S. Ornstein. On isomorphism of weak Bernoulli transformations. *Advances in Math.*, 5:365–394 (1970), 1970.
- [21] Yves Guivarc’h and Émile Le Page. Simplicité de spectres de Lyapounov et propriété d’isolation spectrale pour une famille d’opérateurs de transfert sur l’espace projectif. In *Random walks and geometry*, pages 181–259. Walter de Gruyter, Berlin, 2004.
- [22] Roger A. Horn and Charles R. Johnson. *Matrix analysis*. Cambridge University Press, Cambridge, second edition, 2013.
- [23] Anders Johansson, Anders Öberg, and Mark Pollicott. Ergodic theory of Kusuoka measures. *J. Fractal Geom.*, 4(2):185–214, 2017.
- [24] Anders Johansson, Anders Öberg, and Mark Pollicott. Phase transitions in long-range ising models and an optimal condition for factors of g -measures. *Ergodic Theory and Dynamical Systems*, to appear.
- [25] Thomas Kempton. Factors of Gibbs measures for subshifts of finite type. *Bull. Lond. Math. Soc.*, 43(4):751–764, 2011.

- [26] Abdelaziz Kondah, Véronique Maume, and Bernard Schmitt. Vitesse de convergence vers l'état d'équilibre pour des dynamiques markoviennes non höldériennes. *Ann. Inst. H. Poincaré Probab. Statist.*, 33(6):675–695, 1997.
- [27] Shigeo Kusuoka. Dirichlet forms on fractals and products of random matrices. *Publ. Res. Inst. Math. Sci.*, 25(4):659–680, 1989.
- [28] François Ledrappier. Principe variationnel et systèmes dynamiques symboliques. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 30:185–202, 1974.
- [29] Carlangelo Liverani. Decay of correlations. *Ann. of Math. (2)*, 142(2):239–301, 1995.
- [30] Ian D. Morris. Ergodic properties of matrix equilibrium states. *Ergodic Theory Dynam. Systems*, 38(6):2295–2320, 2018.
- [31] Ian D. Morris. A necessary and sufficient condition for a matrix equilibrium state to be mixing. *Ergodic Theory and Dynamical Systems*, to appear.
- [32] Frédéric Naud. Birkhoff cones, symbolic dynamics and spectrum of transfer operators. *Discrete Contin. Dyn. Syst.*, 11(2-3):581–598, 2004.
- [33] D. S. Ornstein. Bernoulli shifts with the same entropy are isomorphic. *Advances in Math.*, 4:337–352, 1970.
- [34] Donald S. Ornstein. On the root problem in ergodic theory. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971)*, Vol. II: Probability theory, pages 347–356. Univ. California Press, Berkeley, Calif., 1972.
- [35] Mark Piraino. Projections of Gibbs states for Hölder potentials. *J. Stat. Phys.*, 170(5):952–961, Mar 2018.

- [36] Mark Piraino. The weak Bernoulli property for matrix Gibbs states. *Ergodic Theory and Dynamical Systems*, to appear.
- [37] Mark Pollicott and Thomas Kempton. Factors of Gibbs measures for full shifts. In *Entropy of hidden Markov processes and connections to dynamical systems*, volume 385 of *London Math. Soc. Lecture Note Ser.*, pages 246–257. Cambridge Univ. Press, Cambridge, 2011.
- [38] F. Redig and F. Wang. Transformations of one-dimensional Gibbs measures with infinite range interaction. *Markov Process. Related Fields*, 16(4):737–752, 2010.
- [39] D. Ruelle. Statistical mechanics of a one-dimensional lattice gas. *Comm. Math. Phys.*, 9:267–278, 1968.
- [40] James S. Vandergraft. Spectral properties of matrices which have invariant cones. *SIAM J. Appl. Math.*, 16:1208–1222, 1968.
- [41] Evgeny Verbitskiy. On factors of g -measures. *Indag. Math. (N.S.)*, 22(3-4):315–329, 2011.
- [42] Peter Walters. *An introduction to ergodic theory*, volume 79 of *Graduate Texts in Mathematics*. Springer-Verlag, New York-Berlin, 1982.
- [43] Peter Walters. Convergence of the Ruelle operator for a function satisfying Bowen’s condition. *Trans. Amer. Math. Soc.*, 353(1):327–347, 2001.
- [44] Peter Walters. Regularity conditions and Bernoulli properties of equilibrium states and g -measures. *J. London Math. Soc. (2)*, 71(2):379–396, 2005.
- [45] Jisang Yoo. On factor maps that send Markov measures to Gibbs measures. *J. Stat. Phys.*, 141(6):1055–1070, 2010.