GIBBS MEASURES AND SYMBOLIC DYNAMICS

SAMUEL A. HARDER

Abstract. In this paper we outline the arguments put forth by Rufus Bowen in his 1975 monograph, *Equilibrium States and the Ergodic theory of Anasov Diffeomorphisms.* We show the existence of Gibbs measures for Hölder continuous potentials on Markov shift spaces, then use the tools of symbolic dynamics and Markov partitions to apply our results to Axiom A diffeomorphisms.

Contents

1. Introduction 1
2. Gibbs Measures 3
2.1. Gibbs Measures in Finite Systems 3
2.2. Gibbs Measures for Infinite Systems 5
2.3. The Construction of Gibbs Measures 8
2.4. The Variational Principle and Equilibrium States 10
3. Symbolic Dynamics 11
3.1. Axiom A Diffeomorphisms 14
3.2. Rectangles 17
3.3. Existence of Markov Partitions 19
3.4. Encoding 21
4. SRB Measures on Anosov Diffeomorphisms 22
4.1. Physical Measures 24
5. Further Reading 24
6. Acknowledgements 24
References 24

1. Introduction

In the field of Dynamical Systems we study spaces and maps on those spaces. Often we are interested in describing the behavior of a dynamical system exactly, but in many cases this is very hard to do.

In the subset of Dynamical systems known as ergodic theory we introduce the tools of measure theory into dynamics. This is useful because we can obtain statistical results for our dynamical systems. In ergodic theory we study invariant measures. Suppose we have a dynamical system on a space X with $f : X \to X$ measurable. Then a measure μ is f-invariant if

$$\mu(A) = \mu(f^{-1}(A))$$

Date: DEADLINES: Draft AUGUST 14 and Final version AUGUST 26, 2018.

for all measurable sets A in our space. If we think of μ as a probability measure, then this statement is equivalent to stating that the probability of finding a point x in the set A is preserved over time.

In ergodic theory (as the name suggests) we are more concerned with ergodic measures, because these are in some sense "irreducible". Any invariant measure can be reduced into a convex combination of ergodic measures. Therefore it suffices to study the ergodic measures if we want results on invariant measures.

Definition 1.1. A measure μ is ergodic if it is f-invariant and

$$f(A) = A \iff \mu(A) = 1 \text{ or } \mu(A) = 0$$

for A measurable.

This definition tells us that we cannot decompose the space into sets of positive measure such that the map preserves the parts of the decomposition. It then seems clear that most of the points in our space "explore the space," because no set of significance is fixed (besides the whole space). The Birkhoff Ergodic Theorem captures this.

Theorem 1.2 (Birkhoff Ergodic Theorem)**.** *Suppose μ is an ergodic measure. Then for x almost everywhere (w/r/t to μ)*

$$\lim_{k\to\infty} \frac{\phi(x) + \phi(fx) + \cdots + \phi(f^{k-1}x)}{k} = \int_X \phi \; d\mu.$$

for any μ-integrable function ϕ.

The ergodic theorem states that if we follow a "typical" point and average some function along its trajectory, then this average will approach the average value of this function on this space. The time average is equal to the space average.

Unfortunately in order to apply the ergodic theorem we need an ergodic measure. How do we find ergodic measures for an arbitrary dynamical system?

One issue that we might encounter when looking for ergodic measures is ergodic measures with trivial support. Consider the interval with 0 and 1 associated, and the map of multiplication by 3 mod 1. It is clear the set $\{1/2\}$ is an invariant set under this transformation. Consequently the measure μ defined by

$$\mu(A) = \begin{cases} 0, & \frac{1}{2} \in A \\ 1, & \frac{1}{2} \notin A \end{cases}$$

is ergodic. However, we see that μ is of little use when we apply the Ergodic theorem. Why? The theorem states that x *almost-everywhere* has time average equal to the space average. Recall that almost everywhere means on a set whose complement has measure zero. But, with this measure, we see that any set containing $\{1/2\}$ has a complement with measure zero. Thus we see that the theorem could apply to a set of points with zero Lebesgue measure. Therefore we don't glean any significant results in this case.

We would like to define measures that are "physically significant," i.e. for which the ergodic theorem applies to a large number of orbits. First we define the basin of a measure μ as the set of points $x \in X$ such that

$$\lim_{k\to\infty} \frac{\phi(x) + \phi(fx) + \cdots + \phi(f^{k-1}x)}{k} = \int_X \phi \; d\mu.$$

holds. Then we define a measure μ to be a *physical measure* if the basin of μ has positive Lebesgue measure. We would like to find physical ergodic measures for our spaces, because for these measures the Ergodic Theorem applies to a set of positive Lebesgue measure.

In this paper we survey the 1975 monograph *Equilibrium States* by Rufus Bowen. Our eventual goal is to construct physical measures for a class of chaotic dynamical systems, in particular Axiom A diffeomorphisms. Our path to constructing this measure will be slightly roundabout. In Section 1 we will first look to construct "good" measures for simple spaces. We introduce the physical concept of a Gibbs distribution, which is the expected distribution of a system's states at equilibrium. Then we will look at how to extend the Gibbs distribution to a measure on a simple infinite system, namely a Markov shift space or subshift of finite type. We will call the infinite analogue of the Gibbs distribution the Gibbs measure. In section 2 we will examine ways in which to extend our Gibbs measure to more interesting spaces, such as those exhibiting hyperbolic (chaotic) dynamics. We will use Markov partitions, which enable us to symbolically encode the dynamics of a system in such a way that the intricacies of the system are preserved. Once we have a symbolic encoding for a system, we can apply the theorems proved in section 1 to that system.

Finally, in Section 4 we will combine Sections 2 and 3 and briefly describe how to transfer results on shift spaces to Axiom A diffeomorphisms.

In writing this paper my goal was to provide intuition for many of the methods in Bowen, without necessarily proving all of them. In the interest of brevity (and clarity) I omit many proofs and refer the reader to Bowen's text.

2. Gibbs Measures

2.1. **Gibbs Measures in Finite Systems.** We begin our discussion of Gibbs measures with the physical viewpoint. Suppose we have a physical system which has a finite number of possible states $1, \ldots, n$ and energies for each state $E_1, \ldots, E_n$. Suppose also that this system is in equilibrium with a much larger system of temperature T, such that the temperature of the larger system is constant. To be in equilibrium means that macroscopic observables, for example a function averaged over the space, are constant.

When the system is at equilibrium we assume that any of its states can occur, however the probability distribution of these states is fixed. What distribution do we expect among the different states?

Before answering this question, we introduce the concept of entropy. Entropy is a measure of our uncertainty about a system's state. Suppose we have a random variable X with states $1, \ldots, n$ such that the probability of state i is given by p_i. We would like to measure how uncertain we are of the state of X.

Definition 2.1. The *entropy* of the random variable X is given by

$$h(X) = \sum_i -p_i \log(p_i).$$

To see that entropy does capture uncertainty of a random variable (or distribution) we note the following facts. Suppose a random variable is constant, i.e. that there is one state with probability 1 and all the other states have probability 0. Then the entropy of that random variable is 0, because there is no uncertainty about its outcome. Alternatively, if our random variable is equidistributed with

$p_i = \frac{1}{n}$ for all i, then the entropy is maximized. Generally, the more biased a system is towards certain states, the lower the entropy.

It is a fundamental principle in statistical mechanics that a system in equilibrium maximizes entropy. By this we mean that the entropy of the probability distribution of the system's states is maximized.

Let's return to our question of before. What distribution do we expect of the physical system with states $1, \dots, n$ and energies $E_1, \dots, E_n$. Without any constraints on our distribution, we expect the equilibrium distribution to be

$$p_i = \frac{1}{n}$$

because this is the unique entropy maximizing distribution.

To derive the Gibbs distribution we make one assumption. We assume that the average energy

$$E = \sum_{i=1}^{n} p_i E_i$$

is fixed. Then we can use the method of Lagrange multipliers to find the entropy-maximizing distribution subject to this constraint. We present the following theorem.

Theorem 2.2. *Subject to the constraint $\sum_i p_i E_i = E$ the unique entropy maximizing distribution is*

$$p_i = \frac{e^{-\beta E_i}}{\sum_i e^{-\beta E_i}}$$

where β is the Lagrange multiplier when optimizing entropy with respect to the constraint $E = \sum_{i=1}^{n} p_i E_i$.

To prove this one uses the method of Lagrange multipliers and the objective function

$$L(p_1, \dots, p_n, \beta) = \sum_i -p_i \log(p_i) + \beta \left(\sum_i E_i p_i - E \right)$$

with the additional constraint that $\sum p_i = 1$.

We observe, by plugging in our distribution into our objective function L, that the maximum value of L is

$$\log \left(\sum_{i=1}^{n} e^{-\beta E_i} \right).$$

Suppose we have some generic energy function $\phi(i) = a_i$. Then we can see that

$$\sum_i -p_i \log(p_i) + \sum_i a_i p_i$$

is maximized under the gibbs distribution.

It is then a fact of the Gibbs distribution that it maximizes this quantity. When we construct Gibbs measures on infinite systems, it will be important that we retain this useful variational principle.

We conclude this section by restating that the Gibbs measure represents the distribution of states when a system is at equilibrium. Therefore, we would expect a Gibbs measure to be invariant under the evolution of a dynamical system. It is with this motivation that we say that the Gibbs measure is a "natural" choice of invariant measure.

2.2. **Gibbs Measures for Infinite Systems.** Previously, we introduced the notion of a Gibbs measure for finite systems. Recall that we interpreted the Gibbs distribution as the probability distribution of a system at equilibrium. In this section we will extend the Gibbs distribution to a measure on a certain class of infinite dynamical systems.

First let $\Sigma_n = \Pi\{1, \ldots, n\}$. In other words, the set Σ_n consists of all the bi-infinite sequences of symbols in $\{1, \ldots, n\}$.[1] We think of this sequence as a single dimensional lattice, maybe as a simple model of a gas or of magnetic particles. The symbol $\{1, \ldots, n\}$ then represents a state for a particle in each position.

We will need some notion of "energy" for each state $\underline{x} \in \Sigma_n$ of our system. We imagine $\underline{x}$ as representing a single dimensional lattice of particles, each of which assume states in 1 through n. Then for a physical system, such as a model of gas particles or magnetic, we have two forms of energy contributions to the total energy of our system.

(1) For each state k there is some intrinsic contribution to the total energy $\Phi_0(k)$ independent of where it occurs in our system;
(2) There is energy due to the interaction between different positions in our system, dependent on which states we have in those positions. We think of this as a function $\Phi_2 : \mathbb{Z} \times \{1, \ldots, n\} \times \{1, \ldots, n\} \to \mathbb{R}$ where $\mathbb{Z}$ represents the absolute distance between the two points.

Then we can represent the total endergy due to the state x_0 occuring in the 0th position as

$$\phi^*(\underline{x}) = \Phi_0(x_0) + \sum_{j \neq 0} \frac{1}{2} \Phi_2(j; x_j, x_0).$$

(The multiplier of $\frac{1}{2}$ prevents double counting of Φ_2).

We will now make the simplifying assumption that the energy due to interactions between particles "decreases quickly" as the distance between those particles grows. We can state this mathematically by assuming that $\|\Phi_2\|_j = \sup_{k_1,k_2} |\phi(j; k_1, k_2)|$ satisfies

$$\sum_{j=1}^{\infty} \|\phi_2\|_j < \infty.$$

This guarantees that the quantity $\phi^*(\underline{x})$ is a finite real number. Note that in many physical systems in physics, such as a magnetic system, the force between two particles (and therefore the potential energy of their interaction) matches this criterion.

We can also equip the space Σ_n with a topology generated by the sets

$$[a_1 \ldots a_n] = \{\underline{x} \mid x_i = a_i \text{ for } i \in [1, n]\}.$$

which we will call *cylinders*. Then any measure on these cylinders will be a Borel measure. A Borel measure is a measure on a topological space that is defined for all the open sets.

Suppose we wish to compute the energy contributed by the particles in positions $-m$ through m. These particles have states $x_{-m} \ldots x_0 \ldots x_m$. We can see that the

---

[1] A bi-infinite sequence is a sequence indexed by the integers. This is in contrast to a regular sequence, also called a right-sided sequence, which is indexed by the natural numbers.

energy is given by

$$E_m(x_{-m},\dots,x_m)=\sum_{j=-m}^{m}\Phi_0(x_j)+\sum_{-m\le j\le k\le m}\Phi_2(k-j,x_k,x_j)$$

and the Gibbs distribution (measure) on this system would assign probabilities proportional to $\exp(-\beta E_m(x_{-m},\dots,x_m))$.

Suppose (and this is a strong assumption) that

$$\mu(x_{-m},\dots,x_m)=\lim_{k\to\infty}\sum\{\mu_k(x'_{-k}\dots x'_k):x'_i=x_i\forall|i|\le m\}$$

converges.

Then the measure μ on the cylinder $[x_{-m},\dots,x_m]$ represents the limit of the measure of that cylinder in the finite system $x_{-k},\dots,x_k$ as k goes to infinity. If this measure were to exist, then we would be justified in calling it the Gibbs distribution.

Now suppose that instead of E_m we took into account the energy contributions of x_j with $j\in[-m,m]$ with all other x_k's:

$$\sum_{j=-m}^{m}\left(\Phi_0(x_j)+\sum_{k=-\infty}^{\infty}\frac{1}{2}\Phi_2(k-j;x_k,x_j)\right),$$

but this can be rewritten as

$$\sum_{j=-m}^{m}\phi^*(\sigma^j\vec{x}).$$

How much do the quantities $E_m(x_{-m},\dots,x_m)$ and $\sum_{j=-m}^{m}\phi^*(\sigma^j\underline{x})$ differ? It turns out we can bound their difference. Since we assumed

$$C=\sum_{k=1}^{\infty}k\|\Phi_2\|_k<\infty$$

we can see that E_m differs from $\sum_{j=-m}^{m}\phi^*(\sigma^j\vec{x})$ by at most $2C$. Therefore the Gibbs distributions derived from E_m or $\phi^*(\sigma^j\vec{x})$ are within factors of $[e^{-2C},e^{2C}]$ of each other. Thus taking into account interactions beyond the finite system $x_{-m},\dots,x_m$ does not drastically change our distribution.

We now define the k-th variation of a continuous potential $\phi:\Sigma_n\to\mathbb{R}$,

$$\text{var}_k\phi=\sup\{|\phi(x)-\phi(y)|:x_i=y_k,\forall|i|\le k\}.$$

We have finally developed all the requisite definitions to state the central theorem of this section:

Theorem 2.3. *Suppose ϕ is a function from Σ_n to $\mathbb{R}$ and there are $c>0,\alpha\in(0,1)$ so that $var_k\phi\le c\alpha^k$ for all k. Then there is a unique $\mu\in\mathcal{M}_\sigma(\Sigma_n)$ for which one can find constants $c_1>0$, $c_2>0$, and P such that*

$$c_1\le\frac{\mu\{\underline{y}:y_i=x_i\ \forall i=0,\dots,m\}}{\exp\left(-Pm+\sum_{k=0}^{m-1}\phi(\sigma^k\underline{x})\right)}\le c_2$$

for every $\underline{x}\in\Sigma_n$ and $m\ge 0$.

Let's deconstruct the theorem statement slightly.

(1) We think of ϕ as some sort of energy or potential function on Σ_n. If we interpret it as the energy of the particle in the 0th position due to interaction with all the other particles, then the statement $var_k\phi \leq c\alpha^k$ for all k is equivalent to saying that interaction between particles decays rapidly with distance.

(2) The central inequality of the theorem is interpreted as saying that the measure of the cylinder $[x_1, \ldots, x_m]$ is approximately,

$$\frac{\exp\left(\sum_{k=0}^{m-1} \phi(\sigma^k \vec{x})\right)}{e^{-Pm}}$$

within factors of $[c_1, c_2]$. The quantity e^{-Pm} is seen as a normalization constant. As we saw before, the numerator is close to $\exp(E_m)$, and therefore approximates the finite Gibbs measures.

We observe that Theorem 2.3 does not have a dynamical flavor. The set Σ_n does not have any map defined on it. Recall from the introduction that our purposes for developing these symbolic systems is quite different from that of physicists. We are interested in modeling the dynamics of another system using a symbolic dynamical system.

How can a Gibbs measure be relevant to symbolic dynamics? The key idea is in noting that the quantity $\sum_{k=0}^{m-1} \phi(\sigma^k \underline{x})$ can be seen as a partial ergodic sum for the point $\underline{x}$ under the map σ. So we see that the Gibbs measure might have some relation to the shift operator. Conveniently, when we apply symbolic dynamics to diffeomorphisms it will become clear that the shift operator is precisely the right map.

Now, while we could simply consider the properties of the Gibbs measure for the system (Σ_n, σ), when we wish to model other systems with symbolic dynamics we will want greater generality. In Section 3 we will show that if one partitions a space correctly, then we have symbolic representations of points in the space via their *itinerary*, which represents the orbit with respect to the partition. However, if we restricted ourselves to Σ_n, we would only be able to model systems where particles can move from any section of the partition to any other. Thus we would like to expand our models so that they might apply to systems without such nice behavior.

Let A be an $n \times n$ matrix of 0's and 1's. Define

$$\Sigma_A = \{\underline{x} \in \Sigma_n : A_{x_i x_{i+1}} = 1 \ \forall i \in \mathbb{Z}\}.$$

We think of A as an incidence matrix for a directed graph. Then Σ_A is the space of bi-infinite walks on this directed graph. If $A_{x_i x_{i+1}} = 1$ then we say that the sequence $x_i x_{i+1}$ is *allowed*. Then Σ_A consists of all sequences such that $x_i x_{i+1}$ is allowed for all $i \in \mathbb{Z}$.

For a concrete look at the meaning Σ_A, we can think of a directed graph corresponding to A, where we think of A as an incidence matrix. Then Σ_A is the set of all possible infinite walks on our graph.

In order to make Σ_A a dynamical system we need to equip it with some map. We define the left-shift operator σ where

$$\sigma(\underline{x})_i = \underline{x}_{i+1}.$$

We call the dynamical system (Σ_A, σ) a *subshift of finite type* (or in some of the literature, a Topological Markov Shift Space). Note that σ is an invertible map on Σ_A. When we introduce symbolic dynamics, (Σ_A, σ) will model invertible dynamical systems.

Subshifts of finite type are particularly nice because of their simplicity. First, the orbit of any point $\underline{x} \in \Sigma_A$ is very clear, and for this reason periodic points are much easier to identify. These subshifts of finite type also have a "Markov" property. If there is a word $a_1 \dots a_n$ in $\vec{x}$ and a word $a_n \dots a_m$ in $\vec{y}$ then $a_1 \dots a_m$ appears in some $\vec{z}$.

We arrive at the following much more dynamical theorem, which yields a σ-invariant measure for a topologically mixing system Σ_A. This is the most important theorem of this section.

Theorem 2.4. *Let Σ_A be topologically mixing, $\phi \in \mathscr{F}_A$. There is a unique σ-invariant Borel probability measure μ on Σ_A for which one can find constants $c_1 > 0$, $c_2 > 0$ and P such that*

$$c_1 \leq \frac{\mu\{\underline{y} : y_i = x_i \ \forall i = 0, \dots, m\}}{\exp\left(-Pm + \sum_{k=0}^{m-1} \phi(\sigma^k \underline{x})\right)} \leq c_2 \tag{2.5}$$

for every $\underline{x} \in \Sigma_A$ and $m \geq 0$. We denote this measure μ with μ_ϕ and call it the Gibbs measure *of ϕ.*

We will spend the remainder of this section outlining the proof of this theorem, as well as looking at the properties of the Gibbs measure μ_ϕ.

2.3. **The Construction of Gibbs Measures.** We will omit the technical proof of the existence of Gibbs measures here in hopes of reducing it to its most essential conceptual parts. We will highlight the use of Transfer Operators as well as the analogy between the Ruelle-Perron-Frobenius theorem and the Perron-Frobenius Theorem on matrices.

In constructing the Gibbs measure it can be shown that we need only consider functions $\phi(\underline{x})$ that are determined by x_i for $i \geq 0$. We refer the reader to [1], Lemmas 1.5 and 1.6 on pages 7-8. Any Gibbs measure constructed on these functions will extend to a Gibbs measure for any Hölder continuous ϕ. We take $\phi \in \mathscr{F}_A$, where $\mathscr{F}_A$ is the family of continous functions $\phi : \Sigma_A \to \mathbb{R}$ where $\text{var}_k \phi \leq b\alpha^k$ for all $k \geq 0$ and for some constants b and $\alpha \in (0, 1)$.

Consider a dynamical system (X, f). Suppose we want to find a f-invariant measure on X. The method of transfer operators is a useful tool in dynamics because they allow us to use spectral theory to identify invariant measures, as well as determine the rate of decay of correlations.

It is a fact that an equivalent definition for a f-invariant measure μ is that

$$\int_X \phi \circ f \ d\mu = \int_X \phi \ d\mu$$

for any function $\phi : X \to \mathbb{R}$.

We define the transfer operator $\mathcal{L}$, acting on ϕ, such that

$$\int_X \phi \circ f d\mu = \int_X \mathcal{L}(\phi) d\mu.$$

Then it is clear that any fixed point of the transfer operator is an invariant measure for our system. To find such a fixed point, we employ the use of spectral theory from functional analysis. This also allows us to give bounds on convergence of certain measures toward an invariant measure.

In constructing Gibbs measures we use a similar method but we define a very special operator,

$$\mathcal{L}_\phi(f(\underline{x})) = \sum_{\underline{y}\in\sigma^{-1}(\vec{x})} e^{\phi(\underline{y})} f(\underline{y})$$

called the Ruelle transfer operator.

The key theorem for constructing Gibbs measures is the following:

Theorem 2.6 (Ruelle-Perron-Frobenius Theorem)**.** *Let Σ_A be topologically-mixing, $\phi \in \mathscr{F}_A \cap \mathscr{C}(\Sigma_A^+)$ and $\mathcal{L} = \mathcal{L}_\phi$ as above. There are $\lambda > 0, h \in \mathscr{C}(\Sigma_A^+)$, with $h > 0$ and $\nu \in \mathcal{M}(\Sigma_A^+)$ for which $\mathcal{L}h = \lambda h, \mathcal{L}^*\nu = \lambda\nu$, $\nu(h) = 1$ and*

$$\lim_{m\to\infty} \|\lambda^{-m}\mathcal{L}^m g - \nu(g)h\| = 0 \text{ for all } g \in \mathscr{C}(\Sigma_A^+).$$

This theorem is a generalization of the Perron-Frobenius theorem for matrices, which is used to prove that there are unique stationary distributions for finite Markov chains. In this case we are dealing with a linear operator $\mathcal{L}$ operating on an infinite-dimensional vector space. We see that any probability distribution g will converge to h under the repeated application of $\frac{1}{\lambda}\mathcal{L}$.

We will now examine the various properties of the measure $\mu_\phi = h\nu$. that make it desirable for our statistical studies. The first is that $\mu = \mu_\phi$ is invariant under the dynamics of our system.

Lemma 2.7. *μ is invariant under $\sigma : \Sigma_A^+ \to \Sigma_A^+$.*

Proof. To show μ is invariant we want to show that $\mu(f) = \mu(f\circ\sigma)$ for all continuous f (where we are viewing μ as a linear functional). We first note that

$$\begin{aligned}((\mathcal{L}f)\cdot g)(\underline{x}) &= \sum_{y\in\sigma^{-1}\vec{x}} e^{\phi(\underline{y})} f(\underline{y}) g(\underline{x}) \\ &= \sum_{y\in\sigma^{-1}\vec{x}} e^{\phi(\underline{y})} f(\underline{y}) g(\sigma\underline{y}) \\ &= \mathcal{L}(f\cdot(g\circ\sigma))(\underline{x}).\end{aligned}$$

Using this fact we have

$$\begin{aligned}\mu(f) &= \nu(hf) \\ &= \nu(\lambda^{-1}\mathcal{L}h\cdot f) \\ &= \lambda^{-1}\nu(\mathcal{L}(h\cdot(f\circ\sigma))) \\ &= \lambda^{-1}(\mathcal{L}^*\nu)(h\cdot(f\circ\sigma)) \\ &= \nu(h\circ(f\circ\sigma)) \\ &= \mu(f\circ\sigma),\end{aligned}$$

as desired. □

Not only is the measure μ invariant but it is also mixing, which implies ergodicity. See Bowen [1] for details. Most importantly, μ_ϕ is the Gibbs measure for our system.

Theorem 2.8. *μ_ϕ is a Gibbs measure for $\phi \in \mathscr{F}_A \cap \mathscr{C}(\Sigma_A^+)$.*

Proof. See Bowen [1]. □

Finally we can characterize μ using a variational principle, just like in the finite case we examined at the start of this section.

2.4. **The Variational Principle and Equilibrium States.** One of the foremost properties of the Gibbs measure is that it satisfies a similar variational principle to the Gibbs distribution for finite systems. By this we mean that the Gibbs measure maximizes a certain quantity. Furthermore, we know the maximum value that this quantity attains over all measures, so we can verify whether or not a measure is a Gibbs measure.

Before we can state this variational principle we need to define a notion of entropy for our measure μ. We will build up to the entropy of a measure in increments.

First, we define the entropy of a partition $\mathcal{C} = \{C_1, \ldots, C_k\}$ of the measure space $(X, \mathscr{B}, \mu)$ to be

$$H_\mu(\mathcal{C}) = \Sigma_{i=1}^k(-\mu(C_i)\log\mu(C_i)).$$

This definition is directly inspired by the entropy of a random variable or distribution. In fact, we can form a random variable that takes on states $1, \ldots, k$ with probabilities $\mu(C_1), \ldots, \mu(C_k)$. Then the two forms of entropy coincide. Intuitively, the entropy of the partition is a measure of how much uncertainty we have about which section of our partition C_i a point $x \in X$ lies in, and conversely how much information we gain about a point when we are told which section of the partition C_i it inhabits.

Then we can define the refinement of two partitions $\mathcal{C}$ and $\mathcal{D}$ as

$$\mathcal{C} \vee \mathcal{D} = \{C_i \cap D_j : C_i \in \mathcal{C}, D_j \in \mathcal{D}\}.$$

For any point $x \in M$, knowing its position relative to the partition $\mathcal{C} \vee \mathcal{D}$ conveys the same information as observing its position relative to both $\mathcal{C}$ and $\mathcal{D}$.

If $\mathcal{D}$ is a finite partition we define

$$h_\mu(T, \mathcal{D}) = \lim_{m\to\infty} \frac{1}{m} H_\mu(\mathcal{D} \vee T^{-1}\mathcal{D} \vee \cdots \vee T^{-m+1}\mathcal{D}).$$

This quantity $h_\mu(T, \mathcal{D})$ captures the average increase in complexity of the orbit structure of our space relative to the partition $\mathcal{D}$.

Finally we are ready to define the entropy of the measure μ.

Definition 2.9. Let μ be a σ-invariant measure on Σ_A and $\mathcal{U} = \{U_1, \ldots, U_n\}$ where $U_i = \{\underline{x} \in \Sigma_A \mid x_0 = i\}$. Then we define the entropy of the measure μ to be

$$s(\mu) = H_\mu(T, \mathcal{U})$$

This definition of entropy is quite particular to our symbolic dynamics (Σ_A, σ). We are tracking the average increase in complexity of our orbits over time. If many points with the same initial condition (i.e. in the same section U_i) diverge, then our entropy will be higher.

We now look to define a quantity known as the pressure of a function $\phi \in \mathscr{F}_A$. Pressure will also in some sense measure the rate at which complexity of orbits grows, but is more subtle in that it weights orbits differently dependent on our choice of ϕ.

We define

$$\sup_{a_0 \ldots a_{m-1}} S_m\phi = \sup\{\sum_{i=0}^{m-1} \phi(\sigma^k \underline{x}) : \underline{x} \in \Sigma_A, x_i = a_i \text{for all} 0 \leq i \leq m\}$$

and

$$Z_m(\phi) = \sum_{a_0 a_1 \ldots a_{m-1}} \exp\left(\sup_{a_0 \ldots a_{m-1}} S_m\phi\right).$$

For each $\phi \in \mathscr{C}(\Sigma_A)$ we define the *pressure* of ϕ to be

$$P(\phi) = \lim_{m \to \infty} \frac{1}{m} \log Z_m(\phi).$$

As stated earlier, the pressure is a more subtle way to track orbit complexity. Note also the similarities between pressure and the maximum value attained by the Gibbs distribution in the finite case.

Finally we arrive at the following characterization of a Gibbs measure of ϕ:

Theorem 2.10. *Let $\phi \in \mathscr{F}_A$, Σ_A topologically mixing and μ_ϕ the Gibbs measure of ϕ. Then μ_ϕ is the unique measure σ-invariant measure μ such that*

$$s(\mu) + \int \phi \, d\mu = P(\phi).$$

Proof. For a full proof, see [1], p. 20-22. □

This variational principle is an exact analogue of the finite case, with $s(\mu)$ representing entropy and $\int \phi \, d\mu$ representing the average value of our function under the measure μ.

We note that all of the definitions given above can be generalized to any measurable dynamical system. However, the most important thing is that we call μ an *equilibrium state* of a dynamical system for a potential ϕ if

$$s(\mu) + \int \phi \, d\mu = P(\phi).$$

For a general thermodynamic formalism, we refer the reader to chapter 2 of Bowen's monograph [1].

3. Symbolic Dynamics

The discussion of symbolic dynamics which follows is largely drawn from Adler [2].

Our goal in this section is to apply symbolic dynamics to non-symbolic systems. Consider a dynamical system consisting of a space X and a map T on that space. We would like a map π from our space X to some shift space $\sum_A$ with the following properties:

(1) We would like an encoding for each element $x \in X$. In order for our shift space to fully capture the dynamics of the system (T, X) we need to be able to distinguish all points via their symbolic encoding; so we also require that π to be injective. We also want to make π surjective.
(2) We would like σ to be conjugate to the map T. The maps $\sigma : \Sigma_A \to \Sigma_A$ and $T : X \to X$ are conjugate via π if

$$\sigma(\pi(x)) = \pi(Tx)$$

and π is a homeomorphism.

When two dynamical systems are conjugate, we can transfer many dynamical properties from one to the other. For instance the number of fixed points and periodic points of each period are preserved under conjugation.

Given these two properties, we can study our dynamical system using the symbolic structure we have built in the previous sections. Is this wishful thinking? Can such an encoding exist, and if so for which systems?[2] The following examples show that in some cases this is definitely possible and also serve to point us toward a more general method of obtaining a symbolic encoding.

Example 3.1 (Multiplication by 2)**.** We will start with a classic dynamical system map. Let $X = [0, 1)$ and $T(x) = 2x \mod 1$.

Consider the binary decimal expansion of a point $x \in X$,

$$x = 0.a_1a_2a_3\dots$$

It seems natural to let π send $\underline{a} = a_1a_2a_3\dots$ to the number x which has decimal expansion $x = 0.a_1a_2a_3....$ Now under the map T, we see that

$$0.a_1a_2a_3\cdots \to 0.a_2a_3a_4\dots.$$

So the shift operator corresponds to the map T in our symbolic space.

It is worth noting that we don't necessarily have a unique encoding for each real number. For instance the real number $x = 0.5$ is encoded by

$$1000000\dots$$

as well as

$$0111111\dots$$

Example 3.2 (Encoding with Partitions)**.** The method of encoding described in the previous example seems rather cheap. For most dynamical systems we won't have an analogue to the decimal expansion of a number to rely on. So in a general dynamical system (T, X) how might one approach this problem?

One idea is to partition our space X into disjoint sets $X_1, \dots, X_n$ such that $X = \cup_i X_i$.

Then for each point $x \in X$ we can generate a symbolic representation $(s_i)_{i\in\mathbb{N}}$ as follows:

(1) Clearly $x \in X_i$ for some i. Write down this i.
(2) Apply T to x and we now have $Tx \in X_j$ for some j. Write down this j.
(3) Repeatedly apply T, each time writing the corresponding symbol for the part of the partition that $T^k(x)$ is in.

With this procedure we can generate an infinite two sided sequence $(s_n)_{n=-\infty}^{\infty}$ such that $T^k(x) \in X_{s_k}$ for all $k \in \mathbb{Z}$. (This is sometimes called the *itinerary* of the point x). The issue with the partition scheme laid out above is that we might not have uniqueness of symbolic encoding. For instance, we could have two points which follow the same trajectory with respect to our partition. In what partitions is this issue avoided? Can it be avoided? Our next example will show that there exist partitions for which we have a unique point for every symbolic sequence.

Example 3.3 (Encoding Multiplication with a Partition)**.** Let's return to Example 3.1. Recall that our issue with Example 3.1 was that, despite offering a good encoding, it did not feel generalizable. In Example 3.2 we introduced the idea of using a partition to encode points in our dynamics, yet we did not have any clear examples of partitions with *good* encodings.

[2]As an aside, we note that if maps are not sufficiently chaotic then we cannot use this symbolic encoding. For instance if we have an infinite set and we apply the identity map as our dynamics, then under any finite partition we will not obtain injectivity.

As we will show in this example, we can view the binary decimal expansions through the lens of partitions, thereby unifying these two approaches and giving support to the partition approach as a general method.

Again let $T(x) = 2x \mod 1$ and $X = [0, 1)$. Then we will partition X into two sets,

$$X_0 = (0, \frac{1}{2}) \text{ and } X_1 = (\frac{1}{2}, 1).$$

Note that the union of the closure of these two sets is the whole space X.

Then for any point $(s_1, s_2, s_3, \dots)$ we identify the point

$$\{x\} = \bigcap_{n=0}^{\infty} \overline{X_{s_1} \cap f^{-1}(R_{s_2}) \cap \dots \cap f^{-n}(R_{s_{n+1}})}.$$

The next example is particularly important because it will serve as the concrete model with which we interpret many of the more general results in the next section.

Example 3.4 (Toral Automorphism)**.** We define a map on the 2-dimensional torus, $\mathbb{T}^2$. We model the torus with a unit square where opposite sides are associated. We can also think of the torus as $\mathbb{R}^2 \mod \mathbb{Z}^2$. Consequently, any linear transformation on $\mathbb{R}^2$ that fixes the lattice $\mathbb{Z}^2$ yields a valid map on the torus.

We define the map $\Gamma : \mathbb{T}^2 \to \mathbb{T}^2$ where

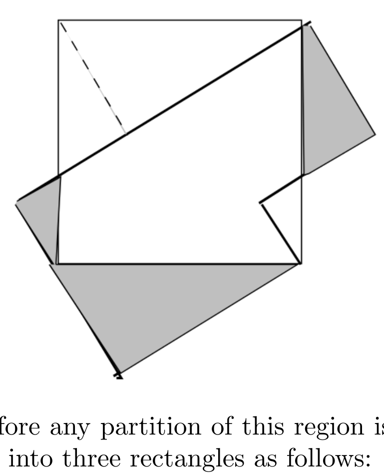
$$\Gamma\left(\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}\right)$$

(In the literature this is often referred to as the cat map on the torus. The etymology is that when V.I. Arnold introduced the map, he demonstrated the result of iterating the map on a square picture of a cat). This linear transformation is diagonalizable into two eigendirections, l_λ and l_μ corresponding to the eigenvalues

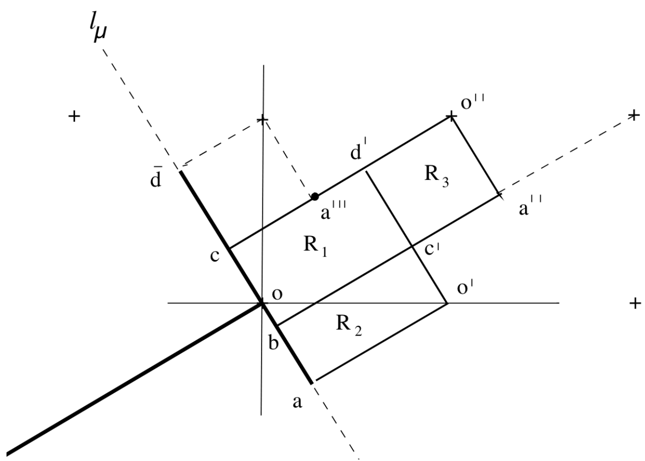
$$\lambda = (1 + \sqrt{5})/2 \text{ and } \mu = (1 - \sqrt{5})/2.$$

In the direction of l_λ our map expands, while in the direction of l_μ it contracts points.

Observe that the region outlined below



covers the torus. Therefore any partition of this region is a partition of the torus. We can split this region into three rectangles as follows:



We observe that the parts of the partition have a simple behavior under application of Γ. The map Γ flips over the expanding axis, and then stretches in that direction, while squishing in the contracting direction. Thus rectangles aligned with the eigendirections stay rectangles.

We will see that this partition is a "good" partition in that any sequence of symbols in $\{1, 2, 3\}$ corresponds to exactly one orbit in our dynamics.

We note several characteristics of this partition that we will hope to generalize.

(1) These rectangles are in some way "aligned" with our dynamics, in that their boundaries have expanding and contracting directions. This is formalized by the product structure being preserved, which we will discuss later.
(2) Second, the boundaries are preserved. When the rectangles are iterated under the map, we see that the boundaries aligned with the expanding direction are mapped onto themselves. When iterating the inverse the same is true of the boundaries aligned with the compressing direction. For this reason rectangles map across each other in the expanding direction as we iterate the map, and stretch across in the contracting direction when we iterate the inverse.

Consider a sequence $(s_k)_{k\in\mathbb{Z}}$ where $s_k \in \{1, 2, 3\}$. We get unique representation because the rectangles stretch across the entire space in both directions. Therefore they must intersect. In other words

$$\cap_{k=-n}^{n} \phi^{-k} R_{s_k} \neq \emptyset$$

so

$$\cap_{k=-\infty}^{\infty} \phi^{-k} R_{s_k}$$

has exactly one point in its intersection.

3.1. **Axiom A Diffeomorphisms.** For constructing Markov partitions we restrict ourselves to hyperbolic systems. There is a practical reason for this. Hyperbolic systems lend themselves to an encoding. Recall that we intend to use a finite partition to generate symbolic dynamics for our systems. In order to give a meaningful symbolic dynamics, we need different points to be distinguishable at some iteration

$k \in \mathbb{Z}$ of our map T. For a hyperbolic system we know that for any point, points around it are either expanding as we iterate the map, or expanding as we iterate its inverse. Therefore we see that points move farther away from each other at some time, leading to them being distinguishable under our partition.

In this section we will construct Markov Partitions on certain subsets of Axiom A diffeomorphisms using the methods of Bowen [1].

The previous examples are a step in the direction of successfully encoding dynamics. We would now like to focus our attention on a certain sub-class of maps where we think a general method would be attainable. These will be the *hyperbolic maps*.

In particular we will be looking at *hyperbolic differentiable dynamical systems*. These encompass both of our preceding examples which were hyperbolic and locally linear. The underlying space for a differentiable dynamical system is a *differentiable manifold*. For our purposes it suffices to think of a differentiable manifold M^n in n-dimensions as a n-dimensionable differentiable surface, or *submanifold*, of $\mathbb{R}^N$, where $N > n$. For each point $x \in M$ we have some *local coordinate system* identifying a neighborhood of the point x with a neighborhood of zero in $\mathbb{R}^n$. We define a *tangent space* $T_xM \subset \mathbb{R}^n$ as the set of all the vectors tangent to M at x. A one-to-one map with a differentiable inverse is called a *diffeomorphism*. For complete generalization, see [1], Chapter 3.

Let $f : M \to M$ be a diffeomorphism on a compact C^∞ Riemannian manifold M. We define a hyperbolic set as follows:

Definition 3.5. A closed subset $\Lambda \subseteq M$ is hyperbolic if $f(\Lambda) = \Lambda$ and each tangent space T_xM with $x \in \Lambda$ can be written as a direct sum

$$T_xM = E_x^u \oplus E_x^s$$

of subspaces so that

(1) $Df(E_x^s) = E_{f(x)}^s$, $Df(E_x^u) = E_{f(x)}^u$;
(2) there exist constants $c > 0$ and $\lambda \in (0, 1)$ such that

$$\|Df^n(v)\| \leq c\lambda^n\|v\| \quad \text{when } v \in E_x^s, n \geq 0$$

and

$$\|Df^{-n}(v)\| \leq c\lambda^n\|v\| \quad \text{when } v \in E_x^u, n \geq 0$$

(3) E_x^s, E_x^u vary continuously with x.

In words, at every point x in a hyperbolic set Λ, we can decompose the derivative (linear approximation) of f into expanding and contracting directions. The derivative preserves these directions. Also, the rate of expansion and contraction is controlled by a single value λ, which is strictly smaller than 1.

In building up to defining Axiom A diffeomorphisms we make the following definitions.

Definition 3.6 (Non-wandering Set)**.** A point $x \in M$ is called *non-wandering* if

$$U \cap \bigcup_{n>0} f^n U \neq \emptyset$$

for every neighborhood U of x. We denote the set of non-wandering points by $\Omega = \Omega(f)$.

Remark 3.7. (1) If a point x is *periodic* then x is in the non-wandering set.

(2) The set Ω is closed and f-invariant.

The Axiom A diffeomorphisms are characterized by their behavior on the nonwandering set. This is analogous to how in one-dimensional dynamics, fixed and periodic points capture much of the dynamics in many cases (asymptotically).

Definition 3.8. We say that a diffeomorphism f satisfies *Axiom A* if the nonwandering set $\Omega(f)$ is hyperbolic and $\Omega(f)$ is the closure of the set periodic points.

For any point $x \in M$ we can define the stable set $W^s(x)$ and the unstable set $W^u(x)$.

Definition 3.9. For a point $x \in M$ let

$$\begin{aligned} W^s(x) &= \{y \in M : d(f^n x, f^n y) \to 0 \text{ as } n \to \infty\} \\ W^s_\varepsilon(x) &= \{y \in M : d(f^n x, f^n y) \leq \varepsilon \text{ as } n \to \infty\} \\ W^u(x) &= \{y \in M : d(f^{-n} x, f^{-n} y) \to 0 \text{ as } n \to \infty\} \\ W^u_\varepsilon(x) &= \{y \in M : d(f^{-n} x, f^{-n} y) \leq \varepsilon \text{ as } n \to \infty\}. \end{aligned}$$

In some texts the sets $W^s(x)$ and $W^u(x)$ are referred to as the stable and unstable manifolds of x respectively. We note that the sets $W^s_\varepsilon(x)$ and $W^u_\varepsilon(x)$ are approximations of $W^s(x)$ and $W^u(x)$ respectively.

We can then prove the following stable manifold theorem for Axiom A diffeomorphisms:

Theorem 3.10 (Stable Manifold Theorem)**.** *Let Λ be a hyperbolic set for a C^r diffeomorphism f. For small $\varepsilon > 0$*

(1) $W^s_\varepsilon(x), W^u_\varepsilon(x)$ are C^r disks for $x \in \Lambda$ with $T_x W^s_\varepsilon(x) = E^s_x$, $T_x W^u_\varepsilon(x) = E^u_x$;
(2) $d(f^n x, f^n y) \leq \lambda^n d(x, y)$ for $y \in W^s_\varepsilon$, $n \geq 0$ and
$d(f^{-n} x, f^{-n} y) \leq \lambda\lambda^n d(x, y)$ for $y \in W^u_\varepsilon(x)$, $N \geq 0$;
(3) $W^s_\varepsilon(x), W^u_\varepsilon(x)$ vary continuously with x (in C^r topology).

Proof. See Bowen [1], p.48. □

This theorem shows that for small enough ε we have $W^s_\varepsilon(x) \subset W^s(x)$. In other words, once a point gets within ε of x, it will converge to x under iteration of f. A similar property holds for $W^u_\varepsilon(x)$ and $W^u(x)$ except with respect to iterations of f^{-1}.

The following theorem introduces a product operator for points in the nonwandering set.

Theorem 3.11. *Suppose f is an Axiom A diffeomorphism. For any small $\varepsilon > 0$ (as in Theorem 3.10) there is a $\delta > 0$ such that $W^s_\varepsilon(x) \cap W^u_\varepsilon(y)$ consists of a single point whenever $x, y \in \Omega(f)$. We define this point to be $[x, y]$. Then*

$$[\cdot, \cdot] : \{(x, y) \in \Omega(f) \times \Omega(f) : d(x, y) \leq \delta\} \to \Omega(f)$$

is continuous.

Proof. See Bowen p.49 for a rigorous proof.

The proof of the first sentence hinges on $W^s_\varepsilon(x)$ and $W^u_\varepsilon(x)$ intersecting transversely. Transversality is basically the opposite of tangency. It is a fact that transverse intersections are preserved under small perturbations.

For the proof that $[x, y] \in \Omega(f)$, use the density of periodic points in $\Omega(f)$ (Axiom A). □

We arrive finally at the following very important characterization of a hyperbolic set.

Theorem 3.12. *Let Λ be a hyperbolic set. Then there is an $\varepsilon > 0$ such that Λ is* expansive in M*, i.e. if $x \in \Lambda$ and $y \in M$ with $y \neq x$, then*

$$d(f^k x, f^k y) > \varepsilon \text{ for some } k \in \mathbb{Z}.$$

Proof. We prove the contrapositive. Suppose $d(f^k x, f^k y) \leq \varepsilon$ for all $k \in \mathbb{Z}$. Then $y \in W^s_\varepsilon(x) \cap W^u_\varepsilon(x)$. So $y = x$ by the previous theorem. □

Remark 3.13. This statement formalizes the exact intuition we had earlier about why hyperbolic sets were good candidates for symbolic encoding. Even if there are points that are very close together, we can find some point in the future or past of the map during which they were distinguishable by some fixed ε. Thus we can be hopeful that a partition can distinguish between even the most closely associated points.

The preceding theorems show that we can get uniqueness of symbolic encodings. In the next section we will look at the appropriate conditions for a partition to generate "Markov" encodings, i.e. encodings that form a Markov shift space.

Assume from now on that f is an Axiom A diffeomorphism.

Theorem 3.14 (Spectral Decomposition)**.** *One can write $\Omega(f) = \Omega_1 \cup \Omega_2 \cup \cdots \cup \Omega_s$ where Ω_i are pairwise disjoint closed sets with*

(a) $f(\Omega_i) = \Omega_i$ and $f|_{\Omega_i}$ is topologically transitive;
(b) $\Omega_i = X_{1,i} \cup \cdots \cup X_{n_i,i}$ with the $X'_{j,i}s$ disjoint closed sets,

$$f(X_{j,i}) = X_{j+1,i}(X_{n_j+1,i} = X_{1,i}$$

and $f^{n_i}|_{X_{j,i}}$ topologically mixing.

Proof. For the proof see Bowen [1] p. 58-60. □

We call the sets Ω_i *basic sets*. We will eventually construct equilibrium states on these basic states.

3.2. **Rectangles.** We will now define an abstract notion of a rectangle. These "rectangles" will be the sections that make up our partitions. They are called rectangles because they are seen to have a product structure. They are also generalizations of the "aligned" rectangles that we saw in Example 3.

Definition 3.15. A subset $R \subseteq \Omega_s$ is called a *rectangle* if it has small diameter and

$$[x, y] \in R \text{ whenever } x, y \in R.$$

We say that R is *proper* if R is closed and $R = \overline{\text{int}(R)}$.

For $x \in R$ let

$$W^s(x, R) = W^s_\varepsilon(x) \cap R \text{ and } W^u(x, R) = W^u_\varepsilon(x) \cap R.$$

As with the torus example, we can express the boundary as a union of expanding and contracting sets.

Lemma 3.16. *Let R be a closed rectangle. As a subset of Ω_s, R has boundary $\delta R = \delta^s R \cup \delta^u R$ where*

$$\delta^s R = \{x \in R : x \not\in int(W^u(x, R))\}$$
$$\delta^u R = \{x \in R : x \not\in int(W^s(x, R))\}.$$

Proof. See Bowen. □

Before giving the definition of a Markov partition, we give some intuition for why this is the *correct* definition.

Suppose have partitioned our space into rectangles, $\mathcal{R} = \{R_1, \ldots, R_n\}$. Suppose we have a point $R_1 \cap f(R_2) \neq \emptyset$ and a point in $R_1 \cap f^{-1}(R_3)$. This is equivalent to saying that in our symbolic encoding the sequence 21 is allowed, and the sequence 13 is allowed. If our symbolic encodings form a Markov shift space then naturally the sequence 213 should occur in some encoding. In other words $f(R_2) \cap R_1 \cap f^{-1}(R_3)$ is nonempty. In a Markov shift space if the sequence $a_1 \ldots a_n$ is allowed, and $a_n \ldots a_m$ is allowed, then $a_1 \ldots a_m$ is allowed.

We say that a Markov partition should have the n-fold intersection property,

$$R_{s_k} \cap \phi^{-1} R_{s_{k+1}} \neq \emptyset, 1 \leq k \leq n-1 \implies \bigcap_{k=1}^{n} \phi^{-k} R_{s_k}.$$

for sequences $s_1, \ldots, s_n$ of length $n \geq 3$. This is called the Markov property of a partition.

However, while this property clearly communicates its meaning, it is hard to check because it is essentially an infinite condition. Our goal is to find an easier but equivalent condition for a partition to have the Markov property.

Suppose we have partition such that

$$fW^u(x, R_i) \supset W^u(fx, R_j)$$

and

$$fW^s(x, R_i) \subset W^s(fx, R_j)$$

when $x \in \text{int}(R_i), fx \in \text{int}(R_j)$. We will call this Property M (alluding to the fact that this is actually a Markov-like property).

Suppose we were looking at the symbolic encodings generated by the orbits of the point x. Suppose a point y is in $W^s(x, R_i)$. This means that after no iterations of our map f, y and x lie in the same rectangle. Then by the second property, after an iteration of our map we see that fy and fx again lie in the same partition. By iterating the second equation we see that $f^k x$ and $f^k y$ lie in the same rectangle for all $k \geq 0$. Similarly we can see that if $z \in W^u(x, R_i)$ then $f^k z$ and $f^k x$ lie in the same rectangle for all $k \leq 0$.

In fewer words, y looks like x in the future, z looks like x in the past. Suppose $\mathcal{R} = \{R_1, \ldots, R_n\}$ is a partition with Property M. Then it satisfies the 3-fold intersection property.

Proof. Suppose $p \in fR_i \cap R_j \neq \emptyset$ and $q \in R_j \cap f^{-1} R_k$. Then consider their product $[p, q]$. We know that $[p, q] \in R_j$ because of the product structure of rectangles. Then we see that $[p, q] \in f^{-1} W^s(fq)$ □

The following theorem allows us to extend this result to the n-fold intersection property.

Theorem 3.17. *For a partition $\mathcal{R} = \{R_1, \dots, R_n\}$ with property M the following two properties hold*

$$f^n W^u(x, R_i) \supset W^u(f^n x, R_j) \text{ and } f^n W^s(x, R_i) \subset W^s(f^n x, R_i).$$

Proof. Induction on Property M leads to this result. □

It follows from this theory that n-fold intersection property holds for any n.

Therefore we see that Property M implies the Markov Property, and it is therefore a suitable replacement condition.

We are finally able to state the definition of a Markov partition (for a basic set Ω_s).

Definition 3.18. A Markov partition of Ω_s is a finite covering $\mathcal{R} = \{R_1, \dots, R_m\}$ of Ω_s by proper rectangles with

1. $\text{int}(R_i) \cap \text{int}(R_j) = \emptyset$ for $i \neq j$,
2. $fW^u(x, R_i) \supset W^u(fx, R_j)$ and
   $fW^s(x, R_i) \subset W^s(fx, R_j)$ when $x \in \text{int}(R_i), fx \in \text{int}(R_j)$.

3.3. **Existence of Markov Partitions.** In this section we will prove that every basic set Ω_s of an Axiom A diffeomorphism has a Markov partition $\mathcal{R}$ of arbitrarily small diameter. Before doing so, we introduce pseudo orbits and shadowing.

Definition 3.19. A sequence of points $\vec{x} = \{x_i\}_{i=a}^b$ of points in M is an *α-pseudo-orbit* if

$$d(fx_i, x_{i+1}) < \alpha \text{ for all } i \in [a, b-1).$$

A point $x \in M$ $\beta-$shadows $\underline{x}$ if

$$d(f^i x, x_i) \leq \beta \text{ for all } i \in [a, b].$$

And we provide the following property of the nonwandering set.

Theorem 3.20. *For every $\beta > 0$ there is an $\alpha > 0$ such that every α-pseudo orbit $\{x_i\}_{i=a}^b$ in Ω (i.e. every $x_i \in \Omega$) is β-shadowed by a point $x \in \Omega$.*

Proof. See Bowen [1] p.51. □

Theorem 3.21. *Let Ω_s be a basic set for an Axiom A diffeomorphism f. Then Ω_s has Markov partitions $\mathcal{R}$ of arbitrarily small diameter.*

Proof. Let $\beta > 0$ be small and choose $\alpha > 0$ such that every α-pseudo orbit is approximated within β by an actual orbit. Since f is continuous we can choose $\gamma < \alpha/2$ such that if $d(x, y) < \gamma$, we have $d(fx, fy) < \alpha/2$.

Now we will construct a rough cover and clean it up after we have attained approximately good encoding.

So let $P = \{p_1, \dots, p_r\}$ be a γ dense subset of Ω_s. See the diagram below. This means that every point $x \in \Omega_s$ is contained in at least one ball.

Now let

$$\Sigma(P) = \left\{ \underline{q} \in \Pi_{-\infty}^{\infty} P : d(f_{q_j}, q_{j+1}) < \alpha \text{ for all } j \right\}.$$

These are all the sequences of points in P that are α-pseudo orbits. We can think about that seqeunce of points as a symbolic representation. We would now like to find an actual point x that "looks like" this sequence of points. By Theorem 3.20 we know that we can find x that β shadows $\underline{q}$. We will denote $x = \theta(\underline{q})$. We can see that this x is actually unique, i.e. the symbolism corresponds to only one point.

Conversely, for each x is there $\underline{q} \in \Sigma(P)$ such that $\theta(\underline{q}) = x$? Well, for each k, choose q_k such that x is in the γ ball around q_k. Then we have

$$d(fq_j, q_{j+1}) \leq d(fq_j, f(f^j(x))) + d(f^{j+1}(x), q_{j+1}) < 2(\alpha/2) = \alpha.$$

We have $d(fq_j, f(f^j(x))) < \alpha/2$ because $d(q_j, f^j(x)) < \gamma$. We know that

$$d(f^{j+1}(x), q_{j+1}) < \gamma$$

because we chose q_{j+1} to be γ-close to $f^{j+1}(x)$, and $\gamma < \alpha/2$ as chosen earlier.

Then we can define a product operator on the set of symbolic sequences that is equivalent to the product operator. For $\underline{q}, \underline{q}' \in \Sigma(P)$ with $q_0 = q_0'$ we define $q^* = [\underline{q}, \underline{q}'] \in \Sigma(P)$ such that

$$q_j^* = \begin{cases} q_j & , j \geq 0 \\ q_j' & , j \leq 0. \end{cases}$$

We would like to show that the element $\theta(q^*) = [\theta(\underline{q}), \theta(\underline{q}')]$. Naturally

$$d(f^j\theta(\underline{q}^*), f^j\theta(\underline{q})) \leq 2\beta$$

for $j \geq 0$ and $d(f^j\theta(q^*), f^j\theta(q')) \leq 2\beta$ for $j \leq 0$. Thus

$$\theta(q^*) \in W^s_{2\beta}(\theta(\underline{q})) \cap W^u_{2\beta}(\theta(\underline{q}'))$$

which implies that

$$\theta[q, q'] = [\theta(\underline{q}), \theta(\underline{q}')]$$

as desired.

We would now like to make some rectangles. Let

$$T_s = \{\theta(\underline{q}) : \underline{q} \in \Sigma(P),\ q_0 = p_s\}.$$

We will show that T_s is actually a rectangle. Choose $x, y \in T_s$. Then we write $x = \theta(\underline{q})$ and $y = \theta(\underline{q}')$ where $q_0 = p_s = q_0'$. Then

$$[x, y] = \theta[\underline{q}, \underline{q}'] \in T_s.$$

Therefore we have

$$fW^s(x, T_s) \subset W^s(fx, T_t)$$

A similar proof shows that

$$fW^u(x, T_s) \sup W^u(fx, T_t).$$

So we see that we almost have a Markov partition except for the fact that our rectangles are not disjoint (so we don't have a partition). The rest of the proof serves to amend this fact.

Bowen shows that θ is continuous in his book, as well as how to get rid of overlaps between the rectangles T_s to get a partition. We refer the reader to Bowen [1] for this part of the proof. □

3.4. **Encoding.** We have a Markov Partition. Now we define a corresponding symbolic space in the way we alluded to earlier in our discussion of partitions. What we will end up showing is that for each symbolic sequence we have only one point in our basic set corresponding to it.

For the Markov partition $\mathcal{R} = \{R_1, \dots, R_m\}$ we define the transition matrix $A = A(\mathcal{R})$ where

$$A_{ij} = \begin{cases} 1 & \text{if } \operatorname{int}(R_i) \cap f^{-1}\operatorname{int}(R_j) \neq \emptyset, \\ 0 & \text{otherwise.} \end{cases}$$

We define the notion of a subrectangle. This is more than just a subset that is rectangle.

Definition 3.22 (u-subrectangle)**.** We say that S is a u-subrectangle of a rectangle R if

(1) $S \neq \emptyset, S \subset R, S$ is proper, and
(2) $W^u(y, S) = W^u(y, R)$ for $y \in S$.

Then we have the following (interesting) Lemma from Bowen (p. 59).

Lemma 3.23. *Suppose S is a u-subrectangle of R_i and $A_{ij} = 1$. Then $f(S) \cap R_j$ is a u-subrectangle of R_j.*

Using this lemma we have our desired theorem.

Theorem 3.24. *For each $\underline{a} \in \Sigma_A$ the set $\cap_{j\in\mathbb{Z}} f^{-j} R_{a_j}$ consists of a single point, denoted $\pi(\underline{a})$. The map $\pi : \Sigma_A \to \Omega_s$ is a continuous surjection, $\pi \circ \sigma = f \circ \pi$, and π is one-to-one over the residual set*

$$Y = \Omega_s \setminus \bigcup_{j\in\mathbb{Z}} f^j(\delta^s\mathcal{R} \cup \delta^u\mathcal{R}).$$

Proof. Suppose $a_1a_2\dots a_n$ is an allowable word. Then by induction we see that

$$\bigcap_{j=1}^{n} f^{n-j} R_{a_j} = R_{a_n} \cap f\left(\bigcap_{j=1}^{n-1} f^{n-1-j} R_{a_j}\right)$$

is a u-subrectangle of R_{a_n}. So we see that

$$K_n(\underline{a}) = \cap_{j=-n}^{n} f^{-j} R_{a_j}$$

is nonempty and the closure of its interior. As $K_n(\underline{a}) \supset K_{n+1}(\underline{a}) \supset \dots$ we have

$$K(\underline{a}) = \bigcap_{j=-\infty}^{\infty} f^{-j} R_{a_j} = \bigcap_{n=1}^{\infty} K_n(\underline{a}) \neq \emptyset.$$

Then if $x, y \in K(\underline{a})$ we have $f^j x, f^j y \in R_{a_j}$ for all $j \in \mathbb{Z}$. So we see that $x = y$ by expansiveness.

It is clear that $\pi\circ\sigma = f\circ\pi$. The map π is continuous, and this is proved similarly to how we proved that θ was cont. earlier. □

We would now like to show that some fundamental properties are maintained by $\sigma : \Sigma_A \to \Sigma_A$. We know that f restricted to Ω_s is topologically transitive. Does this hold for Σ_A under σ? Similarly, does topological mixing transfer? The next theorem answers these questions.

Theorem 3.25. *The dynamical system* $\sigma : \Sigma_A \to \Sigma_A$ *is topologically transitive. If* $f|_{\Sigma_s}$ *is topologically mixing so is* $\sigma : \Sigma_A \to \Sigma_A$.

Proof. See Bowen [1]. □

4. SRB Measures on Anosov Diffeomorphisms

In this section we will use Markov Partitions to transfer our results on symbolic spaces to Axiom A Diffeomorphisms using Markov partitions.

We will see that the dynamics of an Axiom A Diffeomorphism are in some sense captured by the basic sets. As we saw earlier, the manifold M can be written as a union of the stable sets for each basic set:

$$M = \bigcup_{n=1}^{s} W^s(\Omega_s).$$

We will then see that for almost every point in the basin of attraction for a basic set Ω_s, the time average of an observable ϕ is equal to the space average on the basic set under the SRB measure.

Recall that a function ϕ is *Hölder continuous* if there are constants $a, \theta > 0$ such that

$$|\phi(x) - \phi(y)| \leq a d(x, y)^{\theta}.$$

Then for Hölder continuous potentials ϕ we have the following result:

Theorem 4.1. *Let* Ω_s *be a basic set for an Axiom A diffeomorphism* f *and* $\phi : \Omega_s \to \mathbb{R}$ *Hölder continuous. Then*

(1) ϕ *has a unique equilibrium state* μ_ϕ *with respect to* f *on* Ω_s*;*
(2) μ_ϕ *is* ergodic*;*
(3) μ_ϕ *is Bernoulli if* $f\ |_{\Omega_s}$ *is topologically mixing.*

Proof. Let $\mathcal{R}$ be a Markov partition for Ω_s with diameter at most ε (as in the lemma). Let A be the transition matrix for $\mathcal{R}$ and π be the factor map, $\pi : \Sigma_A \to \Omega_s$.

We can find a function ϕ^* on Ω_s by composing ϕ with π, $\phi^* = \phi \circ \pi$. We would like to show that $\phi^* \in \mathscr{F}_A$. Recall that $\phi^* \in \mathscr{F}_A$ if it is continuous and $\text{var}_k \phi \leq b\alpha^k$ for some positive constant b and $\alpha \in (0, 1)$.

Consider $x, y \in \Sigma_A$ and suppose that $x_k = y_k$ for $k \in [-N, N]$. Then we can conclude that $f^k\pi(x), f^k\pi(y)$ are in $R_{x_k} = R_{y_k}$ for $k \in [-N, N]$. In other words, with respect to the partition, the points $\pi(x)$ and $\pi(y)$ are identical for f^k where $k \in [-N, N]$. We also chose $\mathcal{R}$ such that each rectangle has diameter less than ε, so

$$d(f^k\pi(x), f^k\pi(y)) < \varepsilon \text{ for } k \in [-N, N]$$

thus by our Lemma, we have $d(\pi(x), \pi(y)) < \alpha^N$. Since ϕ is Holder continuous we have

$$|\phi(\pi(x)) - \phi(\pi(y))| = |\phi^*(\underline{x}) - \phi^*(\underline{y})| \leq a(\alpha^N)^{\phi} = a(\alpha^{\theta})^N.$$

Thus $\phi^* \in \mathscr{F}_A$.

Now suppose $f\ |_{\Omega_s}$ is mixing. We proved in the previous section that this implies that $\sigma\ |_{\Sigma_A}$ is mixing. Consequently we have a Gibbs measure μ_{ϕ^*}. Recall that we know that μ_{ϕ^*} is ergodic.

Now recall that uniqueness of encoding for points in Ω_s only failed on points that at some point in time landed on the boundary of our partition. This makes intuitive sense, since when we land on a boundary we have two choices of which partition the point is in. We would like to show that this non-uniqueness is irrelevant in measure theory because the set of points with non-unique encoding has measure zero.

So let $D_s = \pi^{-1}(\delta^s \mathcal{R})$ and $D_u = \pi^{-1}(\delta^u \mathcal{R})$. These are closed subsets of Σ_A (because they are the preimage of closed sets), and each is a strict subset of Σ_A. Furthermore we know that $\sigma D_s \subset D_s$ and $\sigma^{-1} D_u \subset D_u$. Since μ_{ϕ^*} is σ-invariant, we see that $\mu_{\phi^*}(\sigma^n D_s) = \mu_{\phi^*}(D_s)$; Since $\sigma^{n+1} D_s \subset \sigma^n D_s$ we have that

$$\sigma^n D_s = \cap_{k=1}^n \sigma^k D_s$$

so

$$\mu_{\phi^*}\left(\bigcap_{n\geq 0} \sigma^n D_s\right) = \mu_{\phi^*}(D_s).$$

We can then see that $\bigcap_{n\geq 0} \sigma^n D_s$ is invariant under σ, so by the ergodicity of μ_{ϕ^*} it either has measure zero or one. But it's complement is open, and therefore has positive measure, so

$$\mu_{\phi^*}\left(\bigcap_{n\geq 0} \sigma^n D_s\right) = \mu_{\phi^*}(D_s) = 0.$$

We can see similarly that $\mu_{\phi^*}(D_u) = 0$.

Now define the measure μ_ϕ on Ω_s where

$$\mu_\phi(E) = \mu_{\phi^*}(\pi^{-1} E).$$

We want to show that μ_ϕ is f-invariant. Well,

$$\begin{aligned}
\mu_\phi(f^{-1}E) &= \mu_{\phi^*}(\pi^{-1}(f^{-1}(E))) \\
&= \mu_{\phi^*}((f \circ \pi)^{-1}(E)) \\
&= \mu_{\phi^*}((\pi \circ \sigma)^{-1}(E)) \\
&= \mu_{\phi^*}(\sigma^{-1}(\pi^{-1}(E))) \\
&= \mu_{\phi^*}(\pi^{-1}(E)) = \mu_\phi(E).
\end{aligned}$$

We also see that σ and f are conjugate automorphisms because π is one-to-one except on a null-set $\bigcup_{n\in\mathbb{Z}} \sigma^n(D_s \cup D_u)$ (i.e. with respect to integration these two things are equivalent under conjugation).

So $h_{\mu_\phi}(f) = h_{\mu_{\phi^*}}(\sigma)$ because entropy is preserved under conjugation.

Then we know that

$$h_{\mu_\phi}(f) + \int \phi \, \mathrm{d}\mu_\phi \leq P_f(\phi)$$

by the variational principle for pressure.

However,

$$\begin{aligned}
h_{\mu_\phi}(f) + \int \phi \, \mathrm{d}\mu_\phi &= h_{\mu_\phi^*}(\sigma) + \int \phi \, \mathrm{d}\mu_{\phi^*} \\
&= P_\sigma(\phi^*) \geq P_f(\phi).
\end{aligned}$$

The last inequality is from Proposition 2.13 from Bowen [1].

Thus $P_\sigma(\phi^*) = P_f(\phi)$ and μ_ϕ is an equilibrium state for ϕ.

We leave uniqueness and generalization to Bowen [1]. □

4.1. **Physical Measures.** In the previous section we outlined how we construct unique equilibrium states for basic sets. However, basic sets do not concern the entire Manifold M, and we would like very much to describe dynamics on a set beyond our basic sets. As is seen in examples, the non-wandering set often has Lebesgue measure zero. However, if we can show that in some way our measure "extends" to an open set beyond the basic sets, then we will have a physical measure on our space M.

Since M is Riemannian locally we have a volume measure m on M. We will assume that $f : M \to M$ is a C^2 Axiom A diffeomorphism and Ω_s is basic set for f.

For $x \in \Omega_s$ let $\phi^{(u)}(x) = -\log \lambda(x)$ where $\lambda(x)$ is the Jacobian of the linear map

$$Df : E^u_x \to E^u_{fx}.$$

It is shown in Bowen that $\phi^{(u)}$ is Hölder continuous. We denote the corresponding equilibrium state for $\phi^{(u)}$ as μ^+.

An attracting basic set is a set Ω_s such that $W^s(\Omega_s)$ has positive Lebesgue measure (or is an open set).

We arrive at the following final theorem. We know $M = \bigcup_{k=1}^r W^s(\Omega_k)$.

Theorem 4.2. *Let Ω_s be a C^2 attractor. For m-almost all $x \in W^s(\Omega_s)$ one has*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} g(f^k x) = \int g \; d\mu^+$$

for all continuous $g : M \to \mathbb{R}$.

Therefore we see that the measure μ^+ on attractors yields a physical measure from which we can derive statistical properties about our Axiom A diffeomorphism. We call this μ^+ the SRB measure on Ω_s.

5. Further Reading

As stated in the abstract, this survey is centered around Bowen's text, [1]. In order to keep this survey relatively brief, I excluded an extensive discussion of the Thermodynamic Formalism described in Chapter 2 of [1]. For an extensive exposition of Markov Partitions, see [2]. For intuition on the transfer operators and the Ruelle Perron Frobenius operator, I recommend Vaughn Climenhaga's notes [8].

6. Acknowledgements

I would like to foremost thank my mentor Jon Dewitt for helping me understand my "unknown unknowns." I would also like to thank Mario Micheli, Bill Barker, and Michael King for inspiring me to pursue mathematics. Finally I would like to thank Peter May for organizing the UChicago REU for the Summer of 2018.

References

[1] Rufus Bowen. Equilibrium States and the Ergodic thoery of Anasov Diffeomorphisms, 2nd Ed. Springer, 2013.
[2] Roy Adler. Symbolic Dynamics and Markov Partitions. Bulletin of the American Mathematical Society, 1998.
[3] Omri Sarig, Penn State Notes on Thermodynamic Formalism
[4] Omri Sarig, Penn State Notes on the Transfer Operator

[5] C. Beck & F. Schogl. Thermodynamics of Chaotic Systems: An Introduction.Cambridge Nonlinear Science Series, 1993
[6] Brinn and Stucke. Introduction to Dynamical Systems. Cambridge University Press, 2004.
[7] Chazottes. Pressure and Equilibrium States in Ergodic Theory. ncyclopedia of Complexity and System Science (Springer), 2008.
[8] Vaughn Climenhaga, Lecture Notes on Spectral Gap in Dynamics. 2013.
[9] Lai-Sang Young. What are SRB Measures, and Which Dynamical Systems Have Them? Journal of Statistical Physics, 2002.