Rufus Bowen

Equilibrium States and the Ergodic Theory of Anosov Diffeomorphisms

New edition of

Lect. Notes in Math. **470**, Springer, 1975.

April 14, 2013

Springer

Preface

The Greek and Roman gods, supposedly, resented those mortals endowed with superlative gifts and happiness, and punished them. The life and achievements of Rufus Bowen (1947–1978) remind us of this belief of the ancients. When Rufus died unexpectedly, at age thirty-one, from brain hemorrhage, he was a very happy and successful man. He had great charm, that he did not misuse, and superlative mathematical talent. His mathematical legacy is important, and will not be forgotten, but one wonders what he would have achieved if he had lived longer. Bowen chose to be simple rather than brilliant. This was the hard choice, especially in a messy subject like smooth dynamics in which he worked. Simplicity had also been the style of Steve Smale, from whom Bowen learned dynamical systems theory.

Rufus Bowen has left us a masterpiece of mathematical exposition: the slim volume *Equilibrium States and the Ergodic Theory of Anosov Diffeomorphisms* (Springer Lecture Notes in Mathematics **470** (1975)). Here a number of results which were new at the time are presented in such a clear and lucid style that Bowen's monograph immediately became a classic. More than thirty years later, many new results have been proved in this area, but the volume is as useful as ever because it remains the best introduction to the basics of the ergodic theory of hyperbolic systems.

The area discussed by Bowen came into existence through the merging of two apparently unrelated theories. One theory was equilibrium statistical mechanics, and specifically the theory of states of infinite systems (Gibbs states, equilibrium states, and their relations as discussed by R.L. Dobrushin, O.E. Lanford, and D. Ruelle). The other theory was that of hyperbolic smooth dynamical systems, with the major contributions of D.V. Anosov and S. Smale. The two theories came into contact when Ya.G. Sinai introduced Markov partitions and symbolic dynamics for Anosov diffeomorphisms. This allowed the poweful techniques and results of statistical mechanics to be applied to smooth dynamics, an extraordinary development in which Rufus Bowen played a major role. Some of Bowen's ideas were as follows. First, only one-dimensional statistical mechanics is discussed: this is a richer theory, which yields what is needed for applications to dynamical systems, and makes use of the powerful analytic tool of transfer operators. Second, Smale's Axiom A dynamical systems are studied rather than the less general Anosov systems. Third, Sinai's Markov partitions are reworked to apply to Axiom A systems and their construction is simplified by the use of *shadowing*. The combination of simplifications and generalizations just outlined led to Bowen's concise and lucid monograph. This text has not aged since it was written and its beauty is as striking as when it was first published in 1975.

Jean-René Chazottes has had the idea to make Bowen's monograph more easily available by retyping it. He has scrupulously respected the original text and notation, but corrected a number of typos and made a few other minor corrections, in particular in the bibliography, to improve usefulness and readability. In his enterprise he has been helped by Jerôme Buzzi, Pierre Collet, and Gerhard Keller. For this work of love all of them deserve our warmest thanks.

Bures sur Yvette, mai 2007 *David Ruelle*

Contents

These notes came out of a course given at the University of Minnesota and were revised while the author was on a Sloan Fellowship.

0

INTRODUCTION

The main purpose of these notes is to present the ergodic theory of Anosov and Axiom A diffeomorphisms. These diffeomorphisms have a complicated orbit structure that is perhaps best understood by relating them topologically and measure theoretically to shift spaces. This idea of studying the same example from different viewpoints is of course how the subjects of topological dynamics and ergodic theory arose from mechanics. Here these subjects return to help us understand differentiable systems.

These notes are divided into four chapters. First we study the statistical properties of Gibbs measures. These measures on shift spaces arise in modern statistical mechanics; they interest us because they solve the problem of determining an invariant measure when you know it approximately in a certain sense. The Gibbs measures also satisfy a variational principle. This principle is important because it makes no reference to the shift character of the underlying space. Through this one is led to develop a "thermodynamic formalism" on compact spaces; this is carried out in chapter two. In the third chapter Axiom A diffeomorphisms are introduced and their symbolic dynamics constructed: this states how they are related to shift spaces. In the final chapter this symbolic dynamics is applied to the ergodic theory of Axiom A diffeomorphisms.

The material of these notes is taken from the work of many people. I have attempted to give the main references at the end of each chapter, but no doubt some are missing. On the whole these notes owe most to D. Ruelle and Ya. Sinai.

To start, recall that $(X, \mathscr{B}, \mu)$ is a *probability space* if $\mathscr{B}$ is a σ-field of subsets of X and μ is a nonnegative measure on $\mathscr{B}$ with $\mu(X) = 1$. By an *automorphism* we mean a bijection $T : X \to X$ for which

$$\begin{aligned} &\text{(i)} \quad E \in \mathscr{B} \quad \text{iff} \quad T^{-1}E \in \mathscr{B}, \\ &\text{(ii)} \quad \mu(T^{-1}E) = \mu(E) \ \text{ for } E \in \mathscr{B}. \end{aligned}$$

If $T : X \to X$ is a homeomorphism of a compact metric space, a natural σ-field $\mathscr{B}$ is the family of Borel sets. A probability measure on this σ-field is called a *Borel* probability measure. Let $\mathscr{M}(X)$ be the set of Borel probability measures on X and $\mathscr{M}_T(X)$ the subset of invariant ones, i.e. $\mu \in \mathscr{M}_T(X)$ if $\mu(T^{-1}E) = \mu(E)$ for all Borel sets E. For any $\mu \in \mathscr{M}(X)$ one can define $T^*\mu \in \mathscr{M}(X)$ by $T^*\mu(E) = \mu(T^{-1}E)$.

Remember that the real-valued continuous functions $\mathscr{C}(X)$ on the compact metric space X form a Banach space under $\|f\| = \max_{x\in X} |f(x)|$. The weak $*$-topology on the space $\mathscr{C}(X)^*$ of continuous linear functionals $\alpha : \mathscr{C}(X) \to \mathbb{R}$ is generated by sets of the form

$$U(f, \varepsilon, \alpha_0) = \{\alpha \in \mathscr{C}(X)^* : |\alpha(f) - \alpha_0(f)| < \varepsilon\}$$

with $f \in \mathscr{C}(X)$, $\varepsilon > 0$, $\alpha_0 \in \mathscr{C}(X)^*$.

Riesz Representation. For each $\mu \in \mathscr{M}(X)$ define $\alpha_\mu \in \mathscr{C}(X)^*$ by $\alpha_\mu(f) = \int f d\mu$. Then $\mu \leftrightarrow \alpha_\mu$ is a bijection between $\mathscr{M}(X)$ and

$$\{\alpha \in \mathscr{C}(X)^* : \alpha(1) = 1 \text{ and } \alpha(f) \geq 0 \text{ whenever } f \geq 0\}\,.$$

We identify α_μ with μ, often writing μ when we mean $\alpha(\mu)$. The weak $*$-topology on $\mathscr{C}(X)^*$ carries over by this identification to a topology on $\mathscr{M}(X)$ (called the weak topology).

Proposition. *$\mathscr{M}(X)$ is a compact convex metrizable space.*

Proof. Let $\{f_n\}_{n=1}^\infty$ be a dense subset of $\mathscr{C}(X)$. The reader may check that the weak topology on $\mathscr{M}(X)$ is equivalent to the one defined by the metric

$$d(\mu, \mu') = \sum_{n=1}^{\infty} 2^{-n}\, \|f_n\|^{-1} \left| \int f_n d\mu - \int f_n d\mu' \right| . \quad \square$$

Proposition. *$\mathscr{M}_T(X)$ is a nonempty closed subset of $\mathscr{M}(X)$.*

Proof. Check that $T^* : \mathscr{M}(X) \to \mathscr{M}(X)$ is a homeomorphism and note that $\mathscr{M}_T(X) = \{\mu \in \mathscr{M}(X) : T^*\mu = \mu\}$. Pick $\mu \in \mathscr{M}(X)$ and let $\mu_n = \frac{1}{n}(\mu + T^*\mu + \cdots + (T^*)^{n-1}\mu)$. Choose a subsequence μ_{n_k} converging to $\mu' \in \mathscr{M}(X)$. Then $\mu' \in \mathscr{M}_T(X)$. $\square$

Proposition. *$\mu \in \mathscr{M}_T(X)$ if and only if*

$$\int (f \circ T)\; d\mu = \int f d\mu \quad \text{for all } f \in \mathscr{C}(X)\,.$$

Proof. This is just what the Riesz Representation Theorem says about the statement $T^*\mu = \mu$. $\square$

1

GIBBS MEASURES

A. Gibbs distribution

Suppose a physical system has possible states $1, \ldots, n$ and the energies of these states are $E_1, \ldots, E_n$. Suppose that this system is put in contact with a much larger "heat source" which is at temperature T. Energy is thereby allowed to pass between the original system and the heat source, and the temperature T of the heat source remains constant as it is so much larger than our system. As the energy of our system is not fixed any of the states could occur. It is a physical fact derived in statistical mechanics that the probability p_j that state j occurs is given by the *Gibbs distribution*

$$p_j = \frac{e^{-\beta E_j}}{\sum_{i=1}^{n} e^{-\beta E_i}},$$

where $\beta = \frac{1}{kT}$ and k is a physical constant.

We shall not attempt the physical justification for the Gibbs distribution, but we will state a mathematical fact closely connected to the physical reasoning.

1.1. Lemma. *Let real numbers $a_1, \ldots, a_n$ be given. Then the quantity*

$$F(p_1, \ldots, p_n) = \sum_{i=1}^{n} -p_i \log p_i + \sum_{i=1}^{n} p_i a_i$$

has maximum value $\log \sum_{i=1}^{n} e^{a_i}$ as $(p_1, \ldots, p_n)$ ranges over the simplex $\{(p_1, \ldots, p_n) : p_i \geq 0,\ p_1 + \cdots + p_n = 1\}$ and that maximum is assumed only by

$$p_j = e^{a_j} \left(\sum_i e^{a_i} \right)^{-1}.$$

This is proved by calculus. The quantity $H(p_1,\dots,p_n)=\sum_{i=1}^n -p_i\log p_i$ is called the *entropy* of the distribution $(p_1,\dots,p_n)$ (*note*: $\varphi(x)=-x\log x$ is continuous on $[0,1]$ if we set $\varphi(0)=0$.) The term $\sum_{i=1}^n p_ia_i$ is of course the average value of the function $a(i)=a_i$. In the statistical mechanics case $a_i=-\beta E_i$, entropy is denoted S and average energy E. The Gibbs distribution then maximizes

$$S-\beta E=S-\frac{1}{kT}E,$$

or equivalently minimizes $E-kTS$. This is called the free energy. The principle that "nature minimizes entropy" applies when energy is fixed, but when energy is not fixed "nature minimizes free energy." We will now look at a simple *infinite* system, the one-dimensional lattice. Here one has for each integer a physical system with possible states $1,2,\dots,n$. A configuration of the system consists of assigning an $x_i\in\{1,\dots,n\}$ for each i:

$$\cdot \quad \cdot \quad \overset{x_{-2}}{\cdot} \quad \overset{x_{-1}}{\cdot} \quad \overset{x_0}{\cdot} \quad \overset{x_1}{\cdot} \quad \overset{x_2}{\cdot} \quad \overset{x_3}{\cdot} \quad \cdot \quad \cdot$$

Thus a configuration is a point

$$\underline{x}=\{x_i\}_{i=-\infty}^{+\infty}\in\prod_{\mathbb{Z}}\{1,\dots,n\}=\Sigma_n\,.$$

We now make assumptions about energy:

(1) associated with the occurrence of a state k is a contribution $\Phi_0(k)$ to the total energy of the system independent of which position it occurs at;
(2) if state k_1 occurs in place i_1, and k_2 in i_2, then the potential energy due to their interaction $\Phi_2^*(i_1,i_2,k_1,k_2)$ depends only on their relative position, *i.e.*, there is a function $\Phi_2:\mathbb{Z}\times\{1,\dots,n\}\times\{1,\dots,n\}\to\mathbb{R}$ so that

$$\Phi_2^*(i_1,i_2,k_1,k_2)=\Phi_2(i_1-i_2;k_1,k_2)$$

(also: $\Phi_2(j;k_1,k_2)=\Phi_2(-j;k_2,k_1)$).
(3) all energy is due to contributions of the form (1) and (2).

Under these hypotheses the energy contribution due to x_0 being in the 0th place is

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

(We "give" each of x_0 and x_k half the energy due to their interaction). We now assume that $\|\Phi_2\|_j=\sup_{k_1,k_2}|\Phi(j;k_1,k_2)|$ satisfies

$$\sum_{j=1}^{\infty}\|\Phi_2\|_j<\infty\,.$$

Then $\phi^*(\underline{x}) \in \mathbb{R}$ and depends continuously on $\underline{x}$ when $\{1,\dots,n\}$ is given the discrete topology and $\Sigma_n = \prod_{\mathbb{Z}}\{1,\dots,n\}$ the product topology.

If we just look at $x_{-m}\dots x_0\dots x_m$ we have a finite system (n^{2m+1} possible configurations) and an energy

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

and the Gibbs distribution μ_m assigns probabilities proportional to $e^{-\beta E_m(x_{-m},\dots,x_m)}$. Now just suppose that for each $x_{-m},\dots,x_m$ the limit

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

exists. Then $\mu \in \mathscr{M}(\Sigma_n)$ and it would be natural to call μ the Gibbs distribution on Σ_n (for the given energy and β). If we are given $\underline{x} = \{x_i\}_{i=-\infty}^{\infty}$, then instead of $E_m(x_{-m},\dots,x_m)$ one might add in the contributions by interactions of x_j $(-m\le j\le m)$ with *all* other x_k's, *i.e.*,

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

If we define the (left) *shift* homeomorphism $\sigma : \Sigma_n \to \Sigma_n$ by $\sigma\{x_i\}_{i=-\infty}^{\infty} = \{x_{i+1}\}_{i=-\infty}^{\infty}$, then this expression is just $\sum_{j=-m}^{m}\phi^*(\sigma^j\underline{x})$. This expression differs from $E_m(x_{-m},\dots,x_m)$ by at most

$$\sum_{j=-m}^{m}\left(\sum_{k=j+m+1}^{\infty}\frac{1}{2}\|\Phi_2\|_k \quad + \sum_{k=m-j+1}^{\infty}\frac{1}{2}\|\Phi_2\|_k\right) \le \sum_{k=1}^{\infty} k\|\Phi_2\|_k\,.$$

Thus, if $C = \sum_{k=1}^{\infty} k\|\Phi_2\|_k < \infty$ then $E_m(x_{-m},\dots,x_m)$ differs from $\sum_{j=-m}^{m}\phi^*(\sigma^j\underline{x})$ by at most C. If we used $\sum_{j=-m}^{m}\phi^*(\sigma^j\underline{x})$ instead of $E_m(x_{-m},\dots,x_m)$ in the Gibbs distribution μ_m, the probabilities would change by factors in $[e^{-2C},e^{2C}]$. The point is that taking x_i into consideration for $i\notin[-m,m]$ may change μ_m, but not drastically if one assumes $\sum_{k=1}^{\infty} k\|\Phi_2\|_k < \infty$.

We want now to state a theorem we have been leading up to. For $\phi : \Sigma_n \to \mathbb{R}$ continuous define

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

As ϕ is uniformly continuous, $\lim_{k\to\infty}\mathrm{var}_k\phi = 0$.

1.2. Theorem. *Suppose $\phi : \Sigma_n \to \mathbb{R}$ and there are $c>0$, $\alpha\in(0,1)$ so that $\mathrm{var}_k\phi \le c\alpha^k$ for all k. Then there is a unique $\mu \in \mathscr{M}_\sigma(\Sigma_n)$ for which one can find constants $c_1>0$, $c_2>0$, and P such that*

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

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

This measure μ is written μ_ϕ and called *Gibbs measure* of ϕ. Up to constants in $[c_1, c_2]$ the relative probabilities of the $x_0 \ldots x_m$'s are given by $\exp \sum_{k=0}^{m-1} \phi(\sigma^k \underline{x})$. For the physical system discussed above one takes $\phi = -\beta\phi^*$. In statistical mechanics Gibbs states are not *defined* by the above theorem. We have ignored many subtleties that come up in more complicated systems (*e.g.*, higher dimensional lattices), where the theorem will not hold. Our discussion was a gross one intended to motivate the theorem; we refer to Ruelle [9] or Lanford [6] for a refined outlook.

For later use we want to make a small generalization of Σ_n before we prove the theorem. If A is an $n \times n$ matrix of 0's and 1's, let

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

That is, we consider all $\underline{x}$ in which A says that $x_i x_{i+1}$ is allowable for every i. One easily sees that Σ_A is closed and $\sigma\Sigma_A = \Sigma_A$. *We will always assume that* A *is such that each* k *between* 1 *and* n *occurs at* x_0 *for some* $\underline{x} \in \Sigma_A$. (Otherwise one could have $\Sigma_A = \Sigma_B$ with B an $m \times m$ matrix and $m < n$.)

1.3. Lemma. $\sigma : \Sigma_A \to \Sigma_A$ *is topologically mixing (*i.e.*, when* U, V *are nonempty open subsets of* Σ_A*, there is an* N *so that* $\sigma^m U \cap V \ne \emptyset\ \forall m \ge N$*) if and only if* $A^M > 0$ (i.e., $A^M_{i,j} > 0\ \forall i, j$) *for some* M.

Proof. One sees inductively that $A^m_{i,j}$ is the number of $(m+1)$-strings $a_0 a_1 \ldots a_m$ of integers between 1 and n with

(a) $A_{a_k a_{k+1}} = 1 \quad \forall k$,
(b) $a_0 = i$, $a_m = j$.

Let $U_i = \{\underline{x} \in \Sigma_A : x_0 = i\} \ne \emptyset$.

Suppose Σ_A is mixing. Then $\exists N_{i,j}$ with $U_i \cap \sigma^n U_j \ne \emptyset\ \forall n \ge N_{i,j}$. If $\underline{a} \in U_i \cap \sigma^n U_j$, then $a_0 a_1 \ldots a_n$ satisfies (a) and (b); so $A^m_{i,j} > 0\ \forall i, j$ when $m \ge \max_{i,j} N_{i,j}$.

Suppose $A^M > 0$ for some M. As each number between 1 and n occurs as x_0 for some $\underline{x} \in \Sigma_A$, each row of A has at least one positive entry. From this it follows by induction that $A^m > 0$ for all $m \ge M$.

Consider open subsets U, V of Σ_A with $\underline{a} \in U$, $\underline{b} \in V$. There is an r so that

$$\begin{aligned} U &\supset \{\underline{x} \in \Sigma_A : x_k = a_k \ \ \forall |k| \le r\} \\ V &\supset \{\underline{x} \in \Sigma_A : x_k = b_k \ \ \forall |k| \le r\}\,. \end{aligned}$$

For $t \ge 2r + M$, $m = t - 2r \ge M$ and $A^m > 0$. Hence find $c_0, \ldots, c_m$ with $c_0 = b_r$, $c_m = a_{-r}$, $A_{c_k c_{k+1}} = 1$ for all k. Then

$$\underline{x} = \cdots b_{-2} b_{-1} b_0 \cdots b_r c_1 \cdots c_{m-1} a_{-r} \cdots a_0 a_1 \cdots$$

is in Σ_A and $\underline{x} \in \sigma^t U \cap V$. So Σ_A is topologically mixing. $\square$

Let $\mathscr{F}_A$ be the family of all continuous $\phi : \Sigma_A \to \mathbb{R}$ for which $\mathrm{var}_k \phi \leq b\alpha^k$ (for all $k \geq 0$) for some positive constants b and $\alpha \in (0,1)$. For any $\beta \in (0,1)$ one can define the metric d_β on Σ_A by $d_\beta(\underline{x}, \underline{y}) = \beta^N$ where N is the largest nonnegative integer with $x_i = y_i$ for every $|i| < N$. Then $\mathscr{F}_A$ is just the set of functions which have a positive Hölder exponent with respect to d_β. The theorem we are interested in then reads

1.4. Existence of Gibbs measures. *Suppose Σ_A is topologically mixing and $\phi \in \mathscr{F}_A$. There is 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*

$$c_1 \leq \frac{\mu\{\underline{y} : y_i = x_i \ \text{ for all } \ i \in [0, m)\}}{\exp\left(-Pm + \sum_{k=0}^{m-1} \phi(\sigma^k \underline{x})\right)} \leq c_2$$

for every $\underline{x} \in \Sigma_A$ and $m \geq 0$.

This theorem will not be proved for some time. The first step is to reduce the ϕ's one must consider.

Definition. *Two functions ψ, $\phi \in \mathscr{C}(\Sigma_A)$ are* homologous *with respect to σ (written $\psi \sim \phi$) if there is a $u \in \mathscr{C}(\Sigma_A)$ so that*

$$\psi(\underline{x}) = \phi(\underline{x}) - u(\underline{x}) + u(\sigma \underline{x}) .$$

1.5. Lemma. *Suppose $\phi_1 \sim \phi_2$ and Theorem 1.4 holds for ϕ_1. Then it holds for ϕ_2 and $\mu_{\phi_1} = \mu_{\phi_2}$.*

Proof.

$$\begin{aligned} \left| \sum_{k=0}^{m-1} \phi_1(\sigma^k \underline{x}) - \sum_{k=0}^{m-1} \phi_2(\sigma^k \underline{x}) \right| &= \left| \sum_{k=0}^{m-1} u(\sigma^{k+1} \underline{x}) - u(\sigma^k \underline{x}) \right| \\ &= |u(\sigma^m \underline{x}) - u(\underline{x})| \leq 2\|u\| . \end{aligned}$$

The exponential in the required inequality changes by at most a factor of $e^{2\|u\|}$ when ϕ_1 is replaced by ϕ_2. Thus the inequality remains valid with c_1, c_2 changed and P, μ unchanged. □

1.6. Lemma. *If $\phi \in \mathscr{F}_A$, then ϕ is homologous to some $\psi \in \mathscr{F}_A$ with $\psi(\underline{x}) = \psi(\underline{y})$ whenever $x_i = y_i$ for all $i \geq 0$.*

Proof. For each $1 \leq t \leq n$ pick $\{a_{k,t}\}_{k=-\infty}^{\infty} \in \Sigma_A$ with $a_{0,t} = t$. Define $r : \Sigma_A \to \Sigma_A$ by $r(\underline{x}) = \underline{x}^*$ where

$$x_k^* = \begin{cases} x_k & \text{for} \quad k \geq 0 \\ a_{k,x_0} & \text{for} \quad k \leq 0 . \end{cases}$$

Let

$$u(\underline{x}) = \sum_{j=0}^{\infty} (\phi(\sigma^j \underline{x}) - \phi(\sigma^j r(\underline{x}))) \,.$$

Since $\sigma^j \underline{x}$ and $\sigma^j r(\underline{x})$ agree in places from $-j$ to $+\infty$,

$$|\phi(\sigma^j \underline{x}) - \phi(\sigma^j r(\underline{x}))| \le \mathrm{var}_j \phi \le b\alpha^j.$$

As $\sum_{j=0}^{\infty} b\alpha^j < \infty$, u is defined and continuous. If $x_i = y_i$ for all $|i| \le n$, then, for $j \in [0, n]$,

$$|\phi(\sigma^j \underline{x}) - \phi(\sigma^j \underline{y})| \le \mathrm{var}_{n-j} \phi \le b\alpha^{n-j}$$

and

$$|\phi(\sigma^j r(\underline{x})) - \phi(\sigma^j r(\underline{y}))| \le b\alpha^{n-j} \,.$$

Hence

$$\begin{aligned} |u(\underline{x}) - u(\underline{y})| &\le \sum_{j=0}^{\left[\frac{n}{2}\right]} |\phi(\sigma^j \underline{x}) - \phi(\sigma^j \underline{y}) + \phi(\sigma^j r(\underline{x})) - \phi(\sigma^j r(\underline{y}))| + 2 \sum_{j > \left[\frac{n}{2}\right]} \alpha^j \\ &\le 2b \left(\sum_{j=0}^{\left[\frac{n}{2}\right]} \alpha^{n-j} + \sum_{j > \left[\frac{n}{2}\right]} \alpha^j \right) \le \frac{4b \ \alpha^{\left[\frac{n}{2}\right]}}{1-\alpha} \cdot \end{aligned}$$

This shows that $u \in \mathscr{F}_A$. Hence $\psi = \phi - u + u \circ \sigma$ is in $\mathscr{F}_A$ also. Furthermore

$$\begin{aligned} \psi(\underline{x}) &= \phi(\underline{x}) + \sum_{j=-1}^{\infty} \big(\phi(\sigma^{j+1} r(\underline{x})) - \phi(\sigma^{j+1} \underline{x})\big) + \sum_{j=0}^{\infty} \big(\phi(\sigma^{j+1} \underline{x}) - \phi(\sigma^j r(\sigma \underline{x}))\big) \\ &= \phi(r(\underline{x})) + \sum_{j=0}^{\infty} \big(\phi(\sigma^{j+1} r(\underline{x})) - \phi(\sigma^j r(\sigma \underline{x}))\big) \,. \end{aligned}$$

The final expression depends only on $\{x_i\}_{i=0}^{\infty}$, as we wanted. D. Lind cleaned up the above proof for us. $\square$

Lemmas 1.5 and 1.6 tell us that in looking for a Gibbs measure μ_ϕ for $\phi \in \mathscr{F}_A$ (*i.e.*, proving Theorem 1.4) we can restrict our attention to functions ϕ for which $\phi(\underline{x})$ depends only on $\{x_i\}_{i=0}^{\infty}$.

B. Ruelle's Perron-Frobenius Theorem

We introduce now one-sided shift spaces. One writes $\underline{x}$ for $\{x_i\}_{i=0}^{\infty}$ (we will continue to write $\underline{x}$ for $\{x_i\}_{i=-\infty}^{\infty}$ but never for both things at the same time). Let

$$\Sigma_A^+ = \left\{ \underline{x} \in \prod_{i=0}^{\infty} \{1, \ldots, n\} \ : \ A_{x_i, x_{i+1}} = 1 \text{ for all } i \ge 0 \right\}$$

and define $\sigma : \Sigma_A^+ \to \Sigma_A^+$ by $\sigma(\underline{x})_i = x_{i+1}$. σ is a finite-to-one continuous map of Σ_A^+ onto itself. If $\phi \in \mathscr{C}(\Sigma_A^+)$ we get $\phi \in \mathscr{C}(\Sigma_A)$ by $\phi(\{x_i\}_{i=-\infty}^{\infty}) = \phi(\{x_i\}_{i=0}^{\infty})$. Suppose $\phi \in \mathscr{C}(\Sigma_A)$ satisfies $\phi(\underline{x}) = \phi(\underline{y})$ whenever $x_i = y_i$ for all $i \geq 0$. Then one can think of ϕ as belonging to $\mathscr{C}(\Sigma_A^+)$ as follows: $\phi(\{x_i\}_{i=0}^{\infty}) = \phi(\{x_i\}_{i=-\infty}^{\infty})$ where x_i for $i \leq 0$ are chosen in any way subject to $\{x_i\}_{i=-\infty}^{\infty} \in \Sigma_A$. The functions $\mathscr{C}(\Sigma_A^+)$ are thus identified with a certain subclass of $\mathscr{C}(\Sigma_A)$. We saw in Lemmas 1.5 and 1.6 that one only needs to get Gibbs measures for $\phi \in \mathscr{C}(\Sigma_A^+) \cap \mathscr{F}_A$ in order to get them for all $\phi \in \mathscr{F}_A$.

In this section we will prove a theorem of Ruelle that will later be used to construct and study Gibbs measures. For $\phi \in \mathscr{C}(\Sigma_A^+)$ define the operator $\mathcal{L} = \mathcal{L}_\phi$ on $\mathscr{C}(\Sigma_A^+)$ by

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

It is the fact that σ is not one-to-one on Σ_A^+ that will make this operator useful.

1.7. Ruelle's Perron-Frobenius Theorem [10, 11]. *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 $\nu \in \mathscr{M}(\Sigma_A^+)$ for which $\mathcal{L}h = \lambda h$, $\mathcal{L}^*\nu = \lambda\nu$, $\nu(h) = 1$ and*

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

Proof. Because $\mathcal{L}$ is a positive operator and $\mathcal{L}1 > 0$, one has that $G(\mu) = (\mathcal{L}^*\mu(1))^{-1}\mathcal{L}^*\mu \in \mathscr{M}(\Sigma_A^+)$ for $\mu \in \mathscr{M}(\Sigma_A^+)$. There is a $\nu \in \mathscr{M}(\Sigma_A^+)$ with $G(\nu) = \nu$ by the Schauder-Tychonoff Theorem (see Dunford and Schwartz, Linear Operators I, p. 456): Let E be a nonempty compact convex subset of a locally convex topological vector space. Then any continuous $G : E \to E$ has a fixed point. In our case $G(\nu) = \nu$ gives $\mathcal{L}^*\nu = \lambda\nu$ with $\lambda > 0$.

We will prove 1.7 via a sequence of lemmas. Let $b > 0$ and $\alpha \in (0, 1)$ be any constants so that $\mathrm{var}_k\phi \leq b\alpha^k$ for all $k \geq 0$. Set $B_m = \exp\left(\sum_{k=m+1}^{\infty} 2b\alpha^k\right)$ and define

$$\Lambda = \{f \in \mathscr{C}(\Sigma_A^+) :\ f \geq 0, \nu(f) = 1\,,\ f(\underline{x}) \leq B_m f(\underline{x}'), \\ \text{whenever } x_i = x_i' \text{ for all } i \in [0, m]\}\,.$$

1.8. Lemma. *There is an $h \in \Lambda$ with $\mathcal{L}h = \lambda h$ and $h > 0$.*

Proof. One checks that $\lambda^{-1}\mathcal{L}f \in \Lambda$ when $f \in \Lambda$. Clearly $\lambda^{-1}\mathcal{L}f \geq 0$ and

$$\nu(\lambda^{-1}\mathcal{L}f) = \lambda^{-1}\mathcal{L}^*\nu(f) = \nu(f) = 1\,.$$

Assume $x_i = x_i'$ for $i \in [0, m]$. Then

$$\mathcal{L}f(\underline{x}) = \sum_j e^{\phi(j\underline{x})} f(j\underline{x})$$

where the sum ranges over all j with $A_{jx_0} = 1$. For $\underline{x}'$ the expression runs over the same j; as $j\underline{x}$ and $j\underline{x}'$ agree in places 0 to $m+1$

$$e^{\phi(j\underline{x})} f(j\underline{x}) \le e^{\phi(j\underline{x}')} e^{b\alpha^{m+1}} B_{m+1} f(j\underline{x}') \le B_m e^{\phi(j\underline{x}')} f(j\underline{x}')$$

and so

$$\mathcal{L}f(\underline{x}) \le B_m \mathcal{L}f(\underline{x}') \,.$$

Consider any $\underline{x}, \underline{z} \in \Sigma_A^+$. Since $A^M > 0$ there is a $\underline{y}' \in \sigma^{-M}\underline{x}$ with $y_0' = z_0$. For $f \in \Lambda$

$$\begin{aligned}\mathcal{L}^M f(\underline{x}) &= \sum_{\underline{y} \in \sigma^{-M}\underline{x}} \exp\left(\sum_{k=0}^{M-1} \phi(\sigma^k \underline{y}) f(\underline{y})\right) \\ &\ge e^{-M\|\phi\|} f(\underline{y}') \,.\end{aligned}$$

Let $K = \lambda^M e^{M\|\phi\|} B_0$. Then $1 = \nu(\lambda^{-M}\mathcal{L}^M f) \ge K^{-1} f(\underline{z})$ gives $\|f\| \le K$ as $\underline{z}$ is arbitrary. As $\nu(f) = 1$, $f(\underline{z}) \ge 1$ for some $\underline{z}$ and we get $\inf \lambda^{-M}\mathcal{L}^M f \ge K^{-1}$.

If $x_i = x_i'$ for $i \in [0, m]$ and $f \in \Lambda$, one has

$$|f(\underline{x}) - f(\underline{x}')| \le (B_m - 1)K \to 0$$

as $m \to \infty$, since $B_m \to 1$. Thus Λ is equicontinuous and compact by the Arzela-Ascoli Theorem. $\Lambda \ne \emptyset$ as $1 \in \Lambda$. Applying Schauder-Tychonoff Theorem to $\lambda^{-1}\mathcal{L} : \Lambda \to \Lambda$ gives us $h \in \Lambda$ with $\mathcal{L}h = \lambda h$. Furthermore $\inf h = \inf \lambda^{-M}\mathcal{L}^M h \ge K^{-1}$. $\square$

1.9. Lemma. *There is an $\eta \in (0,1)$ so that for $f \in \Lambda$ one has $\lambda^{-M}\mathcal{L}^M f = \eta h + (1-\eta) f'$ with $f' \in \Lambda$.*

Proof. Let $g = \lambda^{-M}\mathcal{L}^M f - \eta h$ where η is to be determined. Provided $\eta\|h\| \le K^{-1}$ we will have $g \ge 0$. Assume $x_i = x_i'$ for all $i \in [0, m]$. We want to pick η so that $g(\underline{x}) \le B_m g(\underline{x}')$, or equivalently

$$(\star) \quad \eta(B_m h(\underline{x}') - h(\underline{x})) \le B_m \lambda^{-M}\mathcal{L}^M f(\underline{x}') - \lambda^{-M}\mathcal{L}^M f(\underline{x}) \,.$$

We saw above that $\mathcal{L}f_1(\underline{x}) \le B_{m+1}\ e^{b\alpha^{m+1}} \mathcal{L}f_1(\underline{x}') \le B_{m+1}\ e^{b\alpha^m} \mathcal{L}f_1(\underline{x}')$ for any $f_1 \in \Lambda$. Applying this to $f_1 = \lambda^{-M+1}\mathcal{L}^{M-1} f$ one has

$$\lambda^{-M}\mathcal{L}^M f(\underline{x}) \le B_{m+1}\ e^{b\alpha^m} \lambda^{-M}\mathcal{L}^M f(\underline{x}') \,.$$

Now $h(\underline{x}) \ge B_m^{-1} h(\underline{x}')$ because $h \in \Lambda$. To get $(\star)$ it is therefore enough to have

$$\eta(B_m - B_m^{-1}) h(\underline{x}') \le (B_m - B_{m+1} e^{b\alpha^m})\ \lambda^{-M}\mathcal{L}^M f(\underline{x}')$$

or

$$\eta(B_m - B_m^{-1})\|h\| \le (B_m - B_{m+1} e^{b\alpha^m}) K^{-1} \,.$$

There is an L so that the logarithms of B_m, B_m^{-1} and $B_{m+1}e^{b\alpha^m}$ are in $[-L, L]$ for all m. Let u_1, u_2 be positive constants such that

$$u_1(x-y) \leq e^x - e^y \leq u_2(x-y) \text{ for all } x, y \in [-L, L],\ x > y\,.$$

For $(\star)$ to hold it is enough for $\eta > 0$ to satisfy

$$\eta\|h\|u_1(\log B_m + \log B_m) \leq K^{-1}u_2(\log B_m - \log(B_{m+1}e^{b\alpha^m}))$$

or

$$\eta\|h\|u_1\left(\frac{4b\alpha^{m+1}}{1-\alpha}\right) \leq K^{-1}u_2 b\alpha^m$$

or

$$\eta \leq u_2(1-\alpha)(4\alpha u_1\|h\|K)^{-1}\,. \qquad \square$$

1.10. Lemma. *There are constants $A > 0$ and $\beta \in (0,1)$ so that*

$$\|\lambda^{-n}\mathcal{L}^n f - h\| \leq A\beta^n$$

for all $f \in \Lambda$, $n \geq 0$.

Proof. Let $n = Mq + r$, $0 \leq r < M$. Inductively one sees from Lemma 1.9 and $\mathcal{L}h = \lambda h$ that, for $f \in \Lambda$,

$$\lambda^{-Mq}\mathcal{L}^{Mq}f = (1-(1-\eta)^q)h + (1-\eta)^q f'_q$$

where $f'_q \in \Lambda$. As $\|f'_q\| \leq K$ one has

$$\|\lambda^{-Mq}\mathcal{L}^{Mq}f - h\| \leq (1-\eta)^q(\|h\| + K)$$

and

$$\begin{aligned}\|\lambda^{-n}\mathcal{L}^n f - h\| &= \|\lambda^{-r}\mathcal{L}^r(\lambda^{-Mq}\mathcal{L}^{Mq}f - h)\| \\ &\leq A(1-\eta)^{q+1} \\ &\leq A\beta^n,\end{aligned}$$

where

$$A = (1-\eta)^{-1}(\|h\| + K)\sup_{0\leq r<M}\|\lambda^{-r}\mathcal{L}^r\|$$

and

$$\beta^M = 1 - \eta\,. \qquad \square$$

1.11. Lemma. *Let $\mathscr{C}_r = \{f \in \mathscr{C}(\Sigma_A^+) : \mathrm{var}_r f = 0\}$. If $F \in \Lambda$, $f \in \mathscr{C}_r$, $f \geq 0$ and $fF \not\equiv 0$, then $\nu(fF)^{-1}\lambda^{-r}\mathcal{L}^r(fF) \in \Lambda$.*

Proof. Assume $x_i = x_i'$ for $i \in [0,m]$. Then

$$\mathcal{L}^r(fF)(\underline{x}) = \sum_{j_1\cdots j_r\underline{x}} \exp\left(\sum_{k=0}^{r-1} \phi(\sigma^k(j_1\cdots j_r\underline{x}))\right) f(j_1\cdots j_r\underline{x})F(j_1\cdots j_r\underline{x})$$

where $j_1\cdots j_r$ runs over all r-strings of symbols for which $j_1\cdots j_r\underline{x} \in \Sigma_A^+$. In the expression for $\mathcal{L}^r(fF)(\underline{x}')$ one has $j_1\cdots j_r$ running over the same r-strings. Now $f(j_1\cdots j_r\underline{x}) = f(j_1\cdots j_r\underline{x}')$ as $f \in \mathscr{C}_r$, $F(j_1\cdots j_r\underline{x}) \le B_{m+r}F(j_1\cdots j_r\underline{x}')$, and $\phi(\sigma^k(j_1\cdots j_r\underline{x})) \le \phi(\sigma^k(j_1\cdots j_r\underline{x}')) + \mathrm{var}_{m+r-k}\phi$. Since

$$B_{m+r}\exp\left(\sum_{k=0}^{r-1}\mathrm{var}_{m+r-k}\phi\right) \le B_{m+r}\exp\left(\sum_{j=m+1}^{m+r} b\alpha^j\right) \le B_m,$$

each term in the above expression for $\mathcal{L}^r(fF)(\underline{x})$ is bounded by B_m times the corresponding term for $\mathcal{L}^r(fF)(\underline{x}')$. Hence $\mathcal{L}^r(fF)(\underline{x}) \le B_m\mathcal{L}^r(fF)(\underline{x}')$.

One must still show $\nu(fF) > 0$. Reasoning as in the proof of 1.8 (with $\mathcal{L}^r(fF)$ in place of f) we get

$$\lambda^r\nu(fF) = \nu(\lambda^{-M}\mathcal{L}^{M+r}(fF)) \ge K^{-1}\mathcal{L}^r(fF)(\underline{z}),$$

for any $\underline{z}$. But $(fF)(\underline{w}) > 0$ gives $\mathcal{L}^r(fF)(\sigma^r\underline{w}) > 0$ and so $\nu(fF) > 0$. $\square$

1.12. Lemma. *For $f \in \mathscr{C}_r$, $F \in \Lambda$ and $n \ge 0$,*

$$\|\lambda^{-n-r}\mathcal{L}^{n+r}(fF) - \nu(fF)h\| \le A\nu(|fF|)\beta^n\,.$$

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

Proof. Write $f = f^+ - f^-$ with f^+, $f^- \ge 0$ and f^+, $f^- \in \mathscr{C}_r$. Then

$$\|\lambda^{-n-r}\mathcal{L}^{n+r}(f^\pm F) - \nu(f^\pm F)h\| \le A\nu(|f^\pm F|)\beta^n.$$

For $f^\pm F \equiv 0$, this is obvious; for $f^\pm F \not\equiv 0$ we apply Lemmas 1.11 and 1.10. These inequalities add up to give us the first statement of the lemma.

Given g and $\varepsilon > 0$ one can find r and f_1, $f_2 \in \mathscr{C}_r$ so that $f_1 \le g \le f_2$ and $0 \le f_2 - f_1 \le \varepsilon$. As $|\nu(f_i) - \nu(g)| < \varepsilon$, the first statement of the lemma with $F = 1$ gives

$$\|\lambda^{-m}\mathcal{L}^m(f_i) - \nu(g)h\| \le \varepsilon(1 + \|h\|)$$

for large m. Since $\lambda^{-m}\mathcal{L}^m f_1 \le \lambda^{-m}\mathcal{L}^m g \le \lambda^{-m}\mathcal{L}^m f_2$,

$$\|\lambda^{-m}\mathcal{L}^m g - \nu(g)h\| \le \varepsilon(1 + \|h\|)$$

for large m. $\square$

The proof of 1.7 is finished.

C. Construction of Gibbs measures

We continue to assume that $\phi \in \mathscr{F}_A \cap \mathscr{C}(\Sigma_A^+)$ and ν, h, λ are as in Ruelle's Perron-Frobenius Theorem. Then $\mu = h\nu$ is a probability measure on Σ_A^+; $\mu(f) = \nu(hf) = \int f(\underline{x})h(\underline{x})\, d\nu(\underline{x})$.

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

Proof. We need to show that $\mu(f) = \mu(f \circ \sigma)$ for $f \in \mathscr{C}(\Sigma_A^+)$. Notice that

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

Using this

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

Because μ is σ-invariant on Σ_A^+ there is a natural way to make μ into a measure on Σ_A. For $f \in \mathscr{C}(\Sigma_A)$ define $f^* \in \mathscr{C}(\Sigma_A^+)$ by

$$f^*(\{x_i\}_{i=0}^\infty) = \min\{f(\underline{y}) : \underline{y} \in \Sigma_A,\ y_i = x_i \text{ for all } i \geq 0\}\,.$$

Notice that for $m, n \geq 0$ one has

$$\|(f\circ\sigma^n)^*\circ\sigma^m - (f\circ\sigma^{n+m})^*\| \leq \mathrm{var}_n f\,.$$

Hence

$$|\mu((f\circ\sigma^n)^*) - \mu((f\circ\sigma^{n+m})^*)| = |\mu((f\circ\sigma^n)^*\circ\sigma^m) - \mu((f\circ\sigma^{n+m})^*)| \leq \mathrm{var}_n f$$

which approaches 0, as $n \to \infty$, since f is continuous.
Hence $\tilde{\mu}(f) = \lim_{n\to\infty} \mu((f\circ\sigma^n)^*)$ exists by the Cauchy criterion. It is straightforward to check that $\tilde{\mu} \in \mathscr{C}(\Sigma_A)^*$. By the Riesz Representation Theorem we see that $\tilde{\mu}$ defines a probability measures on Σ_A, which we will denote by μ despite the possible ambiguity. Note that

$$\tilde{\mu}(f\circ\sigma) = \lim_{n\to\infty} \mu((f\circ\sigma^{n+1})^*) = \tilde{\mu}$$

proving that $\tilde{\mu}$ is σ-invariant. Also $\tilde{\mu}(f) = \mu(f)$ for $f \in \mathscr{C}(\Sigma_A)^*$.

Recall that μ is *ergodic* if $\mu(E) = 0$ or 1 whenever E is a Borel set with $\sigma^{-1}E = E$. One calls μ *mixing* if

$$\lim_{n\to\infty} \mu(E \cap \sigma^{-n}F) = \mu(E)\mu(F),$$

for all Borel sets E and F. It is clear that mixing implies ergodicity and a standard argument shows that the mixing condition only need be checked for E and F in a basis for the topology.

1.14. Proposition. *μ is mixing for $\sigma : \Sigma_A \to \Sigma_A$.*

Proof. Writing $S_m\phi(\underline{x}) = \sum_{k=0}^{m-1} \phi(\sigma^k \underline{x})$ one checks inductively that for f, $g \in \mathscr{C}(\Sigma_A^+)$ one has

$$(\mathcal{L}^m f)(\underline{x}) = \sum_{\underline{y}\in\sigma^{-m}\underline{x}} e^{S_m\phi(\underline{y})} f(\underline{y})\,.$$

Then

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

Let

$$\begin{aligned}E &= \{\underline{y}\in\Sigma_A : y_i = a_i, r\le i\le s\}\,,\\ F &= \{\underline{y}\in\Sigma_A : y_i = b_i, u\le i\le v\}\,.\end{aligned}$$

In checking the mixing condition for E and F we may assume $r = u = 0$ because μ is σ-invariant. We want to calculate

$$\begin{aligned}\mu(E\cap\sigma^{-n}F) &= \mu(\chi_E\cdot\chi_{\sigma^{-n}F})\\ &= \mu(\chi_E\cdot(\chi_F\circ\sigma^n))\\ &= \nu(h\chi_E\cdot(\chi_F\circ\sigma^n))\\ &= \lambda^{-n}\mathcal{L}^{*n}\nu(h\chi_E\cdot(\chi_F\circ\sigma^n))\\ &= \nu(\lambda^{-n}\mathcal{L}^n(h\chi_E\cdot(\chi_F\circ\sigma^n)))\\ &= \nu(\lambda^{-n}\mathcal{L}^n(h\chi_E)\cdot\chi_F)\,.\end{aligned}$$

Now

$$\begin{aligned}|\mu(E\cap\sigma^{-n}F) - \mu(E)\mu(F)| &= |\mu(E\cap\sigma^{-n}F) - \nu(h\chi_E)\nu(h\chi_F)|\\ &= |\nu\left((\lambda^{-n}\mathcal{L}^n(h\chi_E) - \nu(h\chi_E)h)\chi_F\right)|\\ &\le \|\lambda^{-n}\mathcal{L}^n(h\chi_E) - \nu(h\chi_E)h\|\ \nu(F)\,.\end{aligned}$$

Because $\chi_E \in \mathscr{C}_s$ Lemma 1.12 gives, for $n \ge s$,

$$\|\lambda^{-n}\mathcal{L}^n(h\chi_E) - \nu(h\chi_E)h\| \leq A\mu(E)\beta^{n-s}$$

where $\beta \in (0,1)$. One then has

$$|\mu(E \cap \sigma^{-n}F) - \mu(E)\mu(F)| \leq A'\mu(E)\mu(F)\beta^{n-s}$$

for $n \geq s$ where $A' = A\ (\inf h)^{-1}$. Thus $\mu(E \cap \sigma^{-n}F) \to \mu(E)\mu(F)$. □

1.15. Lemma. *Let $a = \sum_{k=0}^{\infty} \mathrm{var}_k\phi < \infty$. If $\underline{x}, \underline{y} \in \Sigma_A$ with $x_i = y_i$ for $i \in [0,m)$, then*

$$|S_m\phi(\underline{x}) - S_m\phi(\underline{y})| \leq a\,.$$

Proof. Define $\underline{y}'$ by

$$y'_i = \begin{cases} y_i & \text{for} \quad i \geq 0 \\ x_i & \text{for} \quad i \leq 0\,. \end{cases}$$

Since $\phi \in \mathscr{C}(\Sigma_A^+)$, $\phi(\sigma^k\underline{y}') = \phi(\sigma^k\underline{y})$ for $k \geq 0$. Hence

$$\begin{aligned} |S_m\phi(\underline{x}) - S_m\phi(\underline{y})| &\leq \sum_{k=0}^{m-1} |\phi(\sigma^k\underline{x}) - \phi(\sigma^k\underline{y}')| \\ &\leq \sum_{k=0}^{m-1} \mathrm{var}_{m-1-k}\phi \\ &\leq a\,. \qquad \square \end{aligned}$$

We now complete the proof of 1.4.

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

Proof. Let $E = \{\underline{y} \in \Sigma_A : y_i = x_i \ \text{ for } \ i \in [0,m)\}$. For any $\underline{z} \in \Sigma_A^+$ there is at most one $\underline{y}' \in \sigma^{-m}\underline{z}$ with $\underline{y}' \in E$. Thus, using 1.15,

$$\begin{aligned} \mathcal{L}^m(h\chi_E)(\underline{z}) &= \sum_{\underline{y} \in \sigma^{-m}\underline{z}} e^{S_m\phi(\underline{y})} h(\underline{y})\chi_E(\underline{y}) \\ &\leq e^{S_m\phi(\underline{x})} e^a \ \|h\| \end{aligned}$$

and so

$$\begin{aligned} \mu(E) &= \nu(h\chi_E) \\ &= \lambda^{-m}\nu(\mathcal{L}^m(h\chi_E)) \\ &\leq \lambda^{-m} e^{S_m\phi(\underline{x})} e^a \ \|h\|\,. \end{aligned}$$

Thus take $c_2 = e^a\|h\|$. On the other hand, for any $\underline{z} \in \Sigma_A^+$ there is at least one $\underline{y}' \in \sigma^{-m-M}\underline{z}$ with $\underline{y}' \in E$. Then

$$\begin{aligned} \mathcal{L}^{m+M}(h\chi_E)(\underline{z}) &\geq e^{S_{m+M}\phi(\underline{y}')} h(\underline{y}') \\ &\geq e^{-M\|\phi\|-a}\ (\inf h)\ e^{S_m\phi(\underline{x})} \end{aligned}$$

and

$$\mu(E) = \lambda^{-m-M}\nu(\mathcal{L}^{m+M}(h\chi_E)) \geq c_1\lambda^{-m}e^{S_m\phi(\underline{x})}$$

where $c_1 = \lambda^{-M}e^{-M\|\phi\|-a}$. We have verified the desired inequalities on measures of cylinder sets given in 1.4 with $P = \log\lambda$.

Suppose now that μ' is any other measure satisfying the inequalities in Theorem 1.4 with constants c_1', c_2', P'. For $\underline{x} \in \Sigma_A$ let $E_m(\underline{x}) = \{\underline{y} \in \Sigma_A \ : \ y_i = x_i \ \text{ for all } \ i \in [0,m)\}$. Let T_m be a finite subset of Σ_A so that $\Sigma_A = \bigcup_{\underline{x}\in T_m} E_m(\underline{x})$ disjointly. Then

$$\begin{aligned} c_1'e^{-P'm}\sum_{\underline{x}\in T_m} e^{S_m\phi(\underline{x})} &\leq \sum_{\underline{x}\in T_m}\mu'(E_m(\underline{x})) \\ &= 1 \\ &\leq c_2'e^{-P'm}\sum_{\underline{x}\in T_m} e^{S_m\phi(\underline{x})}\,. \end{aligned}$$

From this one sees that $P' = \lim_{m\to\infty}\frac{1}{m}\log\left(\sum_{\underline{x}\in T_m} e^{S_m\phi(\underline{x})}\right)$. One can apply the same reasoning to μ; hence $P' = P$ as they equal the same limit.

The estimates on $\mu'(E_m(\underline{x}))$ and $\mu(E_m(\underline{x}))$ give us $\mu'(E_m(\underline{x})) \leq d\mu(E_m(\underline{x}))$ where $d = c_2'c_1^{-1}$. Taking limits this extends to $\mu'(E) \leq d\mu(E_m)$ for all Borel sets E. In particular $\mu'(E) = 0$ when $\mu(E) = 0$. By the Radon-Nikodym Theorem $\mu' = f\mu$ for some μ-integrable f. Applying σ one has

$$\begin{aligned} \mu' &= \sigma^*\mu' \\ &= (f\circ\sigma)\ \sigma^*\mu \\ &= (f\circ\sigma)\mu\,. \end{aligned}$$

As the Radon-Nikodym derivative is unique up to μ-equivalence, $f\circ\sigma \overset{\text{a.e.}}{=} f$. Because μ is ergodic this gives f equivalent to some constant c.

$1 = \mu'(\Sigma_A) = \int c\ d\mu = c$ and $\mu = \mu'$. $\square$

D. Variational principle

We will describe Gibbs measures as those maximizing a certain quantity, in a way analogous to Lemma 1.1. If $\mathcal{C} = \{C_1,\dots,C_k\}$ is a partition of a measure space $(X,\mathscr{B},\mu)$ (*i.e.*, the C_i's are pairwise disjoint and $X = \bigcup_{i=1}^k C_i$), one defines the entropy

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

If $\mathcal{D}$ is another (finite) partition,

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

1.17. Lemma. $H_\mu(\mathcal{C} \vee \mathcal{D}) \le H_\mu(\mathcal{C}) + H_\mu(\mathcal{D})$.

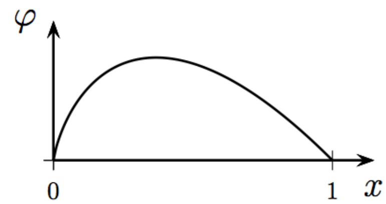
Proof.

$$\begin{aligned} H_\mu(\mathcal{C} \vee \mathcal{D}) - H_\mu(\mathcal{C}) &= \sum_{i,j} (-\mu(C_i \cap D_j) \log \mu(C_i \cap D_j)) - \sum_i (-\mu(C_i) \log \mu(C_i)) \\ &= \sum_{i,j} -\mu(C_i \cap D_j) \log \frac{\mu(C_i \cap D_j)}{\mu(C_i)} \\ &= \sum_j \sum_i \mu(C_i) \left(\frac{-\mu(C_i \cap D_j)}{\mu(C_i)} \log \frac{\mu(C_i \cap D_j)}{\mu(C_i)} \right). \end{aligned}$$

The function $\varphi(x) = -x \log x$ ($\varphi(0) = 0$) is concave on $[0,1]$ as $\varphi''(x) < 0$ for $x \in (0,1)$. From this it follows that

$$\varphi(a_1 x_1 + a_2 x_2) \ge a_1 \varphi(x_1) + a_2 \varphi(x_2)$$

when $x_1, x_2 \in [0,1]$, $a_1 + a_2 = 1$. Inductively one sees that $\varphi(\sum_{i=1}^n a_i x_i) \ge \sum_{i=1}^n a_i \varphi(x_i)$ when $\sum_{i=1}^n a_i = 1$ and $a_i \ge 0$.



Applying this to $a_i = \mu(C_i)$ and $x_i = \frac{\mu(C_i \cap D_j)}{\mu(C_i)}$ we get

$$\sum_i \mu(C_i)\ \varphi\left(\frac{\mu(C_i \cap D_j)}{\mu(C_i)}\right) \le \varphi\left(\sum_i \mu(C_i \cap D_j)\right) = \varphi(\mu(D_j)).$$

So

$$H_\mu(\mathcal{C} \vee \mathcal{D}) - H_\mu(\mathcal{C}) \le \sum_j \varphi(\mu(D_j)) = H_\mu(\mathcal{D}). \qquad \square$$

1.18. Lemma. *Suppose* $\{a_m\}_{m=1}^\infty$ *is a sequence satisfying* $\inf \frac{a_m}{m} > -\infty$, $a_{m+n} \le a_m + a_n$ *for all* m, n. *Then* $\lim_{m\to\infty} \frac{a_m}{m}$ *exists and equals* $\inf_m \frac{a_m}{m}$.

Proof. Fix $m > 0$. For $j > 0$, write $j = km + n$ with $0 \le n < m$. Then

$$\frac{a_j}{j} = \frac{a_{km+n}}{km+n} \le \frac{a_{km}}{km} + \frac{a_n}{km} \le \frac{ka_m}{km} + \frac{a_n}{km}.$$

Letting $j \to \infty$, $k \to \infty$ one gets

$$\limsup_j \frac{a_j}{j} \le \frac{a_m}{m}.$$

Thus $\limsup_j \frac{a_j}{j} \le \inf_m \frac{a_m}{m}$. As $\liminf_j \frac{a_j}{j} \ge \inf_m \frac{a_m}{m}$, one gets that $\lim_j \frac{a_j}{j}$ exists and equals $\inf_m \frac{a_m}{m}$. $\square$

1.19. Lemma. *If $\mathcal{D}$ is a (finite) partition of $(X, \mathscr{B}, \mu)$ and T an automorphism of $(X, \mathscr{B}, \mu)$, then*

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

exists.

Proof. Let $a_m = H_\mu(\mathcal{D} \vee T^{-1}\mathcal{D} \vee \cdots \vee T^{-m+1}\mathcal{D})$. Then

$$a_{m+n} \le H_\mu(\mathcal{D}\vee T^{-1}\mathcal{D}\vee\cdots\vee T^{-m+1}\mathcal{D}) + H_\mu(T^{-m}\mathcal{D}\vee\cdots\vee T^{-m-n+1}\mathcal{D}) \le a_m + a_n$$

since

$$T^{-m}\mathcal{D} \vee \cdots \vee T^{-m-n+1}\mathcal{D} = T^{-m}(\mathcal{D} \vee \cdots \vee T^{-n+1}\mathcal{D})$$

and μ is T-invariant. $\square$

Definition. *Let $\mu \in \mathscr{M}_\sigma(\Sigma_A)$ and $\mathcal{U} = \{U_1, \ldots, U_n\}$ where $U_i = \{\underline{x} \in \Sigma_A : x_0 = i\}$. Then $s(\mu) = h_\mu(\sigma, \mathcal{U})$ is called the* entropy *of μ.*

Suppose now that $\phi \in \mathscr{C}(\Sigma_A)$ and that $a_0 a_1 \cdots a_{m-1}$ are integers between 1 and n satisfying $A_{a_k a_{k+1}} = 1$. Write

$$\sup_{a_0a_1\cdots a_{m-1}} S_m\phi = \sup\left\{\sum_{k=0}^{m-1} \phi(\sigma^k \underline{x}) : \underline{x} \in \Sigma_A,\ x_i = a_i \text{ for all } 0 \le i < m\right\}$$

and

$$Z_m(\phi) = \sum_{a_0a_1\cdots a_{m-1}} \exp\left(\sup_{a_0a_1\cdots a_{m-1}} S_m\phi\right).$$

1.20. Lemma. *For $\phi \in \mathscr{C}(\Sigma_A)$, $P(\phi) = \lim_{m\to\infty} \frac{1}{m} \log Z_m(\phi)$ exists (called the* pressure *of ϕ).*

Proof. Notice that

$$\sup_{a_0a_1\cdots a_{m+n-1}} S_{m+n}\phi \le \sup_{a_0a_1\cdots a_{m-1}} S_m\phi + \sup_{a_m\cdots a_{m+n-1}} S_n\phi\,.$$

From this one gets $Z_{m+n}(\phi) \le Z_m(\phi)Z_n(\phi)$; the terms in $Z_{m+n}(\phi)$ are bounded by terms in $Z_m(\phi)Z_n(\phi)$ and $Z_m(\phi)Z_n(\phi)$ may have more terms, all positive. Apply Lemma 1.18 to $a_m = \log Z_m(\phi)$ (notice $a_m \ge -m\|\phi\|$). $\square$

1.21. Proposition. *Suppose $\phi \in \mathscr{C}(\Sigma_A)$ and $\mu \in \mathscr{M}_\sigma(\Sigma_A)$. Then*

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

Proof. As $\int \phi \circ \sigma^k d\mu = \int \phi d\mu$, $\frac{1}{m}\int S_m\phi d\mu = \int \phi d\mu$ where $S_m\phi(\underline{x}) = \sum_{k=0}^{m-1} \phi(\sigma^k \underline{x})$. Hence

$$s(\mu) + \int \phi \; d\mu \le \lim_{m\to\infty} \frac{1}{m}\left(H_\mu(\mathcal{U} \vee \cdots \vee \sigma^{-m+1}\mathcal{U}) + \int S_m\phi \; d\mu\right).$$

Now $\mathcal{U} \vee \cdots \vee \sigma^{-m+1}\mathcal{U}$ partitions points $\underline{x} \in \Sigma_A$ according to $x_0 x_1 \cdots x_{m-1}$. Thus

$$\begin{aligned} &H_\mu(\mathcal{U} \vee \cdots \vee \sigma^{-m+1}\mathcal{U}) + \int S_m\phi \; d\mu \\ &\le \sum_{a_0\cdots a_{m-1}} \mu(a_0\cdots a_{m-1})(-\log \mu(a_0 \cdots a_{m-1})) \; + \sup_{a_0 a_1 \cdots a_{m-1}} S_m\phi \\ &\le \log Z_m(\phi) \text{ by Lemma 1.1}\,. \end{aligned}$$

Now let $m \to \infty$. $\square$

1.22. Theorem. *Let $\phi \in \mathscr{F}_A$, Σ_A topologically mixing and μ_ϕ the Gibbs measure of ϕ. Then μ_ϕ is the unique $\mu \in \mathscr{M}_\sigma(\Sigma_A)$ for which*

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

Proof. Given $a_0 \cdots a_{m-1}$, pick $\underline{x}$ with $x_i = a_i$ $(i = 0, \ldots, m-1)$ and

$$S_m\phi(\underline{x}) = \sup_{a_0 a_1 \cdots a_{m-1}} S_m\phi\,.$$

Now, as $\mu = \mu_\phi$ is the Gibbs measure,

$$\frac{\mu\{\underline{y} \in \Sigma_A : y_i = a_i \;\; \forall\, 0 \le i < m\}}{\exp(-Pm + S_m\phi(\underline{x}))} \in [c_1, c_2]\,.$$

Summing the measure of these sets over all possible $a_0 \cdots a_{m-1}$'s gives 1; so

$$c_1 \exp(-Pm) Z_m(\phi) \le 1 \le c_2 \exp(-Pm) Z_m(\phi)$$

or

$$\frac{Z_m(\phi)}{\exp(Pm)} \in [c_2^{-1}, c_1^{-1}]\,.$$

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

If $y_i = x_i$ for all $i = 0, \ldots, m-1$, then

$$\begin{aligned} &|S_m\phi(\underline{y}) - S_m\phi(\underline{x})| \le \sum_{k=0}^{m-1} |\phi(\sigma^k \underline{y}) - \phi(\sigma^k \underline{y})| \\ &\le \mathrm{var}_0\phi + \mathrm{var}_1\phi + \cdots + \mathrm{var}_{[\frac{m}{2}]}\phi + \mathrm{var}_{m-[\frac{m}{2}]}\phi + \cdots + \mathrm{var}_0\phi \\ &\le 2c \sum_{k=0}^{[\frac{m}{2}]} \alpha^k \le \frac{2c}{1-\alpha} = d\,. \end{aligned}$$

Hence, if $B = \{\underline{y} \in \Sigma_A : y_i = a_i \text{ for all } i = 0, \ldots, m-1\}$, then for $\underline{x} \in B$,

$$\begin{aligned}
&-\mu(B)\log\mu(B) + \int_B S_m\phi \, d\mu \geq -\mu(B)\ [\ \log\mu(B) - S_m\phi(\underline{x}) + d\ ]\\
&\geq -\mu(B)\ [\ \log\left(c_2 e^{-Pm+S_m\phi(\underline{x})}\right) - S_m\phi(\underline{x}) + d\ ]\\
&\geq \mu(B)(Pm - \log c_2 - d)\,.
\end{aligned}$$

Since

$$\begin{aligned}
H_\mu(\mathcal{U} \vee \cdots \vee \sigma^{-m+1}\mathcal{U}) + \int S_m\phi \, d\mu \\
&= \sum_B \big(-\mu(B)\log\mu(B) + \int_B S_m\phi \, d\mu\big)\\
&\geq \sum_B \mu(B)(Pm - \log c_2 - d) = Pm - \log c_2 - d,
\end{aligned}$$

we get

$$\begin{aligned}
&s(\mu) + \int \phi \, d\mu = \lim_{m\to\infty} \frac{1}{m}\left(H_\mu(\mathcal{U} \vee \cdots \vee \sigma^{-m+1}\mathcal{U}) + \int S_m\phi \, d\mu\right)\\
&\geq \lim_{m\to\infty} \frac{1}{m}(Pm - \log c_1 - d) = P = P(\phi)\,.
\end{aligned}$$

The reverse inequality was in Proposition 1.21. So

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

To prove uniqueness we will need a couple of lemmas.

1.23. Lemma. *Let X be a compact metric space, $\mu \in \mathscr{M}(X)$, and $\mathcal{D} = \{D_1, \ldots, D_n\}$ a Borel partition of X. Suppose $\{\mathcal{C}_m\}_{m=1}^\infty$ is a sequence of partitions so that* $\operatorname{diam}(\mathcal{C}_m) = \max_{C\in\mathcal{C}_m} \operatorname{diam}(C) \to 0$ *as $m \to \infty$. Then there are partitions $\{E_1^m, \ldots, E_n^m\}$ so that*

1. *each E_i^m is a union of members of $\mathcal{C}_m$,*
2. $\lim_{m\to\infty} \mu(E_i^m \Delta D_i) = 0$ *for each i.*

Proof. Pick compacts $K_1, \ldots, K_n$ with $K_i \subset D_i$ and $\mu(D_i \backslash K_i) < \varepsilon$. Let $\delta = \inf_{i\neq j} d(K_i, K_j)$ and consider m with $\operatorname{diam}(\mathcal{C}_m) \leq \frac{\delta}{2}$. Divide the elements $C \in \mathcal{C}_m$ into groups whose unions are $E_1^m, \ldots, E_n^m$ so that

$$C \subset E_i^m \quad \text{if} \quad C \cap K_i \neq \emptyset\,.$$

As $\operatorname{diam}(\mathcal{C}_m) \leq \frac{\delta}{2}$ any $C \in \mathcal{C}_m$ can intersect at most one K_i. Put a C hitting no K_i in any E_i^m you like. Then $E_i^m \supset K_i$ and

$$\mu(E_i^m \Delta D_i) = \mu(D_i \backslash E_i^m) + \mu(E_i^m \backslash D_i) \leq \varepsilon + \mu\left(X \backslash \bigcup_{i=1}^n K_i\right) \leq (n+1)\,\varepsilon\,. \qquad \square$$

1.24. Lemma. *Suppose* $0 \le p_1, \dots, p_m \le 1$, $s = p_1 + \cdots + p_m \le 1$ *and* $a_1, \dots, a_m \in \mathbb{R}$. *Then*

$$\sum_{i=1}^{m} p_i(a_i - \log a_i) \le s\left(\log \sum_{i=1}^{m} e^{a_i} - \log s\right).$$

Proof. This generalizes 1.1. One shows by calculus that the left side is maximized at $p_i = \frac{se^{a_i}}{\sum_j e^{a_j}}$. $\square$

Proof of Theorem 1.22 (continued). Let $\nu \in \mathscr{M}_\sigma(\Sigma_A)$ satisfy $s(\nu) + \int \phi \, d\nu = P$. First suppose ν is singular with respect to μ. Then there is a Borel set B with $\sigma(B) = B$, $\mu(B) = 0$ and $\nu(B) = 1$. Let $\mathcal{C}_m = \sigma^{-\left[\frac{m}{2}\right]+1}\mathcal{U} \vee \cdots \vee \mathcal{U} \vee \cdots \vee \sigma^{\left[\frac{m}{2}\right]}\mathcal{U}$. Then $\operatorname{diam}(\mathcal{C}_m) \to 0$ (use d_β metric). Applying 1.23 to $\{B, X \backslash B\}$ one finds sets E^m which are unions of elements of $\mathcal{C}_m$ and satisfy $(\mu + \nu)(B \Delta E^m) \to 0$. As $\mu + \nu$ is σ-invariant and $\sigma^{-m+\left[\frac{m}{2}\right]}B = B$ one has $(\mu + \nu)(B \Delta F^m) \to 0$ where $F^m = \sigma^{-m+\left[\frac{m}{2}\right]}E^m$ is a union of members of $\mathcal{U} \vee \cdots \vee \sigma^{-m+1}\mathcal{U}$. Since $s(\nu) = \inf \frac{1}{m} H_\nu(\mathcal{U} \vee \cdots \vee \sigma^{-m+1}\mathcal{U})$ one has

$$P = P(\phi) = s(\nu) + \int \phi \, d\nu \le \frac{1}{m}\left(H_\nu(\mathcal{U} \vee \cdots \vee \sigma^{-m+1}\mathcal{U}) + \int S_m\phi \, d\nu\right)$$

or

$$mP \le \sum_{B \in \mathcal{U} \vee \cdots \vee \sigma^{-m+1}\mathcal{U}} \left[-\nu(B)\log\nu(B) + \int_B S_m\phi \, d\nu\right].$$

Picking $\underline{x}_B \in B$ one has $S_m\phi \le S_m\phi(\underline{x}_B) + d$ on B and so

$$\begin{aligned} mP &\le d + \sum_B \nu(B)(S_m\phi(\underline{x}_B) - \log\nu(B)) \\ &\le d + \sum_{B \subset F^m} \nu(B)(S_m\phi(\underline{x}_B) - \log\nu(B)) \\ &\quad + \sum_{B \subset X \backslash F^m} \nu(B)(S_m\phi(\underline{x}_B) - \log\nu(B)). \end{aligned}$$

Applying 1.24

$$\begin{aligned} mP - d &\le \nu(F^m)\,\log \sum_{B \subset F^m} \exp(S_m\phi(\underline{x}_B)) \\ &\quad + \quad \nu(X \backslash F^m)\;\log \sum_{B \subset X \backslash F^m} \exp(S_m\phi(\underline{x}_B)) + 2K^*. \end{aligned}$$

where $K^* = \sup_{0 \le s \le 1}(-s\log s)$. Rearranging terms:

$$\begin{aligned}
-2K^* - d \ \le\ & \nu(F^m)\ \log \sum_{B\subset F^m} \exp\big(S_m\phi(\underline{x}_B) - mP\big) \\
+\ \ & \nu(X\backslash F^m)\ \log \sum_{B\subset X\backslash F^m} \exp\big(S_m\phi(\underline{x}_B) - mP\big) \\
\le\ \ & \nu(F^m)\ \log \sum_{B\subset F^m} c_2^{-1}\mu(B)\ +\ \nu(X\backslash F^m)\ \log \sum_{B\subset X\backslash F^m} c_2^{-1}\mu(B) \\
\le\ \ & \log c_2^{-1} + \nu(F^m)\log\mu(F^m) + \nu(X\backslash F^m)\log\mu(X\backslash F^m)\,.
\end{aligned}$$

Letting $m\to\infty$, $\nu(F^m)\to 1$, $\mu(F^m)\to 0$ and the above inequality is contradictory.

In general, for $\nu'\in\mathscr{M}_\sigma(\Sigma_A)$, write $\nu' = \beta\nu + (1-\beta)\mu'$ where $\beta\in(0,1)$, $\nu\in\mathscr{M}_\sigma(\Sigma_A)$ is singular w.r.t. μ and $\mu'\in\mathscr{M}_\sigma(\Sigma_A)$ is absolutely continuous w.r.t. μ. As ν and μ' are supported on disjoints sets

$$P_{\nu'}(\phi) = \beta P_\nu(\phi) + (1-\beta)P_{\mu'}(\phi),$$

where $P_\nu(\phi) = s(\nu) + \int \phi\ d\nu$. Suppose $P_{\nu'}(\phi) = P$. Since $P_\nu(\phi)\le P$ and $P_{\mu'}(\phi)\le P$ (Prop. 1.21), we have $P_\nu(\phi) = P$ or $\beta = 0$. We just saw that $P_\nu(\phi)\ne P$. Thus $\nu' = \mu'$ and write $\nu' = \frac{d\nu'}{d\mu}\mu$. Then the $\frac{d\nu'}{d\mu}$ is σ-invariant up to equivalence as ν', μ are both invariant and the Radon-Nikodym derivative is unique up to equivalence. So $\frac{d\nu'}{d\mu}$ is constant and $\nu' = \mu$. □

E. Further properties

In this section we look at more examples of the good behavior of Gibbs measures. Throughout we assume $\mu = \mu_\phi$ with $\phi\in\mathscr{F}_A$ and $\sigma|_{\Sigma_A}$ topologically mixing.

Two partitions $\mathcal{P}$ and $\mathcal{Q}$ are called *ε-independent* if

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

Let $\mathcal{U} = \{U_1,\dots,U_n\}$ be the partition of Σ_A with

$$U_j = \{\underline{x}\in\Sigma_A : x_0 = j\}\,.$$

The partition $\mathcal{U}$ is called *weak-Bernoulli* (for σ and μ) if for every $\varepsilon > 0$ there is an $N(\varepsilon)$ so that

$$\mathcal{P} = \mathcal{U}\vee\sigma^{-1}\mathcal{U}\vee\cdots\vee\sigma^{-s}\mathcal{U} \quad\text{and}\quad \mathcal{Q} = \sigma^{-t}\mathcal{U}\vee\cdots\vee\sigma^{-t-r}\mathcal{U}$$

are ε-independent for all $s\ge 0$, $r\ge 0$, $t\ge s+N(\varepsilon)$. A well-known theorem of Friedman and Ornstein [4] states that if $\mathcal{U}$ is weak-Bernoulli, then (σ,μ) is conjugate to a Bernoulli shift.

1.25. Theorem. *$\mathcal{U}$ is weak-Bernoulli for the Gibbs measure $\mu = \mu_\phi$.*

Proof. We may assume that $\phi \in \mathscr{C}(\Sigma_A^+)$ as before. For $P \in \mathcal{P}$ we have $\chi_P \in \mathscr{C}_r$. As in the proof of 1.14, for $Q \in \mathcal{Q}$

$$|\mu(P \cap Q) - \mu(P)\mu(Q)| \le A'\mu(P)\mu(Q)\beta^{t-s}$$

where $\beta \in (0,1)$. Summing over P, Q

$$\sum_{P,Q} |\mu(P \cap Q) - \mu(P)\mu(Q)| \le A'\beta^{t-s} \le \varepsilon$$

when $t - s$ is large. $\square$

Because μ is mixing, $\mu(f \cdot (g \circ \sigma^n)) \to \mu(f)\mu(g)$ as $n \to \infty$ for f, g continuous (in fact L^2) functions. The above proof used that this convergence was exponentially fast for characteristic functions of cylinder sets. This exponential convergence will now be carried over to functions in $\mathscr{F}_A$. For $\alpha \in (0,1)$ let $\mathscr{H}_\alpha$ be the family of $f \in \mathscr{C}(\Sigma_A)$ with $\mathrm{var}_k f \le c\alpha^k$ for some c. $\mathscr{H}_\alpha$ is a Banach space under the norm

$$\|f\|_\alpha = \|f\| + \sup_{k \ge 0}(\alpha^{-k}\mathrm{var}_k f)\,.$$

1.26. Exponential Cluster Property. *For fixed $\alpha \in (0,1)$ there are constants D and $\gamma \in (0,1)$ so that*

$$|\mu(f \cdot (g \circ \sigma^n)) - \mu(f)\mu(g)| \le D\|f\|_\alpha \|g\|_\alpha \gamma^n$$

for all $f, g \in \mathscr{H}_\alpha$, $n \ge 0$.

Proof. For $k \ge 0$ and $\underline{x} \in \Sigma_A$, let

$$E_k(\underline{x}) = \{\underline{y} \in \Sigma_a : y_i = x_i \ \text{ for all } \ |i| \le k\}\,.$$

Define $f_k(\underline{x}) = \mu(E_k(\underline{x}))^{-1} \int_{E_k(\underline{x})} f \, d\mu$. Then $\mu(f_k) = \mu(f)$ and $\|f - f_k\| \le \|f\|_\alpha \alpha^k$. Hence

$$\begin{aligned} &|\mu(f \cdot (g \circ \sigma^n)) - \mu(f)\mu(g)| \le \\ &|\mu(f_k \cdot (g_k \circ \sigma^n)) - \mu(f_k)\mu(g_k)| + |\mu((f - f_k)\cdot(g \circ \sigma^n))| + |\mu(f_k \cdot ((g - g_k) \circ \sigma^n))| \\ &\le |\mu(f_k \cdot (g_k \circ \sigma^n)) - \mu(f_k)\mu(g_k)| + 2\alpha^k \|f\|_\alpha \|g\|_\alpha. \end{aligned}$$

Now f_k is measurable with respect to the partition $\mathcal{P} = \{E_k(\underline{x})\}_{\underline{x}}$; *i.e.*, $f_k = \sum_{P \in \mathcal{P}} a_P \chi_P$. Also $g_k = \sum_{P \in \mathcal{P}} b_P \chi_P$. Hence

$$\begin{aligned} |\mu(f_k \cdot (g_k \circ \sigma^n)) - \mu(f_k)\mu(g_k)| &\le \sum_{P,Q \in \mathcal{P}} |a_P b_P| \; |\mu(P \cap \sigma^{-n}Q) - \mu(P)\mu(Q)| \\ &\le \|f\|\|g\| A'\beta^{n-2k} \\ &\le \|f\|_\alpha \|g\|_\alpha A'\beta^{n-2k}\,. \end{aligned}$$

Letting $k = [n/3]$ we get the result with $\gamma = \max(\alpha^{1/3}, \beta^{1/3})$. $\square$

1.27. Central Limit Theorem. *For $\psi \in \mathscr{F}_A$ there is a $\xi = \xi(\psi) \in [0,\infty)$ so that*

$$\mu\left\{\underline{x} \in \Sigma_A : \frac{1}{\sqrt{n}}(S_n\psi(\underline{x}) - n\mu(\psi)) < r\right\} \overset{n\to\infty}{\longrightarrow} \frac{1}{\xi\sqrt{2\pi}}\int_{-\infty}^{r} e^{-x^2/2\xi^2}\, dx\,.$$

For $\xi = 0$, convergence is asserted for $r \neq 0$ and the expression on the right is taken to be 0 for $r < 0$ and 1 for $r > 0$.

Remark. *We omit the proof, referring the reader to M. Ratner [13]. It is interesting to know when $\xi(\psi) = 0$. This happens (see [13]) precisely when*

$$\psi - \mu(\psi) = u \circ \sigma - u$$

has a solution $u \in L^2(\mu)$. It is interesting that in case such u can be found one can find $u \in \mathscr{F}_A$ and so ψ is homologous to a constant. The reasoning for this is very roundabout and it would be good to find a nice direct proof.

1.28. Theorem. *Let Σ_A be topologically mixing and $\phi, \psi \in \mathscr{F}_A$. The following are equivalent:*

(i) $\mu_\phi = \mu_\psi$.
(ii) *There is a constant K so that $S_m\phi(\underline{x}) - S_m\psi(\underline{x}) = mK$ whenever $\sigma^m\underline{x} = \underline{x}$.*
(iii) *There are a constant K and a $u \in \mathscr{F}_A$ so that*

$$\phi(\underline{x}) = \psi(\underline{x}) + K + u(\sigma\underline{x}) - u(\underline{x}) \text{ for all } \underline{x} \in \Sigma_A.$$

(iv) *There are constants K and L so that $|S_m\phi(\underline{x}) - S_m\psi(\underline{x}) - mK| \leq L$ for all $\underline{x}$ and all $m > 0$.*

If these conditions hold, then $K = P(\phi) - P(\psi)$.

Proof. (iii) $\Rightarrow$ (iv) is obvious and (iv) $\Rightarrow$ (i) is just like Lemma 1.5. Assume $\mu_\phi = \mu_\psi$ and $\sigma^m\underline{x} = \underline{x}$. From the definition of a Gibbs measure and $\mu_\phi = \mu_\psi$ one sees that

$$\frac{\exp(-P(\phi)j + S_j\phi(\underline{x}))}{\exp(-P(\psi)j + S_j\psi(\underline{x}))} \in [d_1, d_2],$$

where $d_1 > 0$, $d_2 > 0$ are independent of $\underline{x}$ and j. This is equivalent to

$$|S_j\phi(\underline{x}) - S_j\psi(\underline{x}) - j(P(\phi) - P(\psi))| \leq M$$

for some M independent of $j, \underline{x}$. If $\sigma^m\underline{x} = \underline{x}$, letting $j = km$, $S_j\phi(\underline{x}) = kS_m\phi(\underline{x})$ and

$$M > k\,|S_m\phi(\underline{x}) - S_m\psi(\underline{x}) - m(P(\phi) - P(\psi))|\,.$$

Letting $k \to \infty$ we get (ii) with $K = P(\phi) - P(\psi)$.

Now assume (ii). In proving (iii) we will need the following standard lemma.

1.29. Lemma. *If $T : X \to X$ is a topologically transitive continuous map of a compact metric space, then there is a point $x \in X$ so that*

$$\text{if } U \neq \emptyset \text{ is open, } N > 0, \text{ then } T^n x \in U \text{ for some } n \geq N.$$

Proof. As X is 2nd countable, let $U_1, U_2, \dots$ be a basis for the topology. By transitivity, the open set

$$V_{i,N} = \bigcup_{n \geq N} T^{-n} U_i$$

is dense in X. By Baire Category Theorem there is an $x \in \bigcap_{i,N} V_{i,N}$. $\square$

Continuing the proof of (iii) from (ii), let $\underline{x}$ be as in the lemma for $X = \Sigma_A$, $T = \sigma$ (topological mixing is stronger than transitivity). Let $\eta = \phi - \psi - K \in \mathscr{F}_A$. Let $\Gamma = \{\sigma^k \underline{x} : k \geq 0\}$ and define $u : \Gamma \to \mathbb{R}$ by

$$u(\sigma^k \underline{x}) = \sum_{j=0}^{k-1} \eta(\sigma^j \underline{x}) \, .$$

As Γ is dense in Σ_A, Γ must be infinite (except in the trivial case of Σ_A =one point) and $\underline{x}$ is not periodic. Thus $\sigma^k \underline{x} \neq \sigma^m \underline{x}$ for $m \neq k$ and u is well defined on Γ. We will estimate $\text{var}_r(u|\Gamma)$. Suppose $\underline{y} = \sigma^k \underline{x}$, and $\underline{z} = \sigma^m \underline{x}$ $(m > k)$ agree in places $-r$ to r. Then $x_{k+s} = x_{m+s}$ for all $|s| \leq r$. Define $\underline{w} \in \Sigma_A$ by

$$w_i = x_t \quad \text{for} \quad i \equiv t \pmod{m-k} \, , \; k \leq t \leq m \, .$$

Then $\sigma^{m-k} \underline{w} = \underline{w}$ and $\underline{w}, \underline{x}$ agree in places $k - r$ to $m + r$; hence $\sigma^j \underline{x}$, $\sigma^j \underline{w}$ agree in places $k - r - j$ through $m + r - j$. Now

$$u(\underline{w}) - u(\underline{y}) = \sum_{j=k}^{m-1} \eta(\sigma^j \underline{x}) \, .$$

Since (ii) gives

$$\sum_{j=k}^{m-1} \eta(\sigma^j \underline{w}) = 0 \, ,$$

$$\begin{aligned} |u(\underline{z}) - u(\underline{y})| &\leq \sum_{j=k}^{m-1} |\eta(\sigma^j \underline{x}) - \eta(\sigma^j \underline{w})| \\ &\leq \text{var}_r \eta + \text{var}_{r+1} \eta + \cdots + \text{var}_{r+1} \eta + \text{var}_r \eta \\ &\leq 2 \sum_{s=r}^{\infty} \text{var}_s \eta. \end{aligned}$$

Since $\eta \in \mathscr{F}_A$, $\text{var}_s \eta \leq c\alpha^s$ for some $\alpha \in (0,1)$ and

$$\mathrm{var}_r(u|_\Gamma) \le 2c \sum_{s=r}^{\infty} \alpha^s = \frac{2c}{1-\alpha}\ \alpha^r \ .$$

So u is uniformly continuous on Γ and therefore extends uniquely to a continuous $u : \Sigma_A = \overline{\Gamma} \to \mathbb{R}$. Because $\mathrm{var}_r u = \mathrm{var}_r(u|_\Gamma)$, $u \in \mathscr{F}_A$. For $\underline{z} \in \Gamma$,

$$u(\sigma\underline{z}) - u(\underline{z}) = \eta(\underline{z})$$

and this equation extends to Σ_A by continuity. $\square$

References

For discussions of Gibbs' states and statistical mechanics we refer to Ruelle's book [9] and Lanford [6]. The definition of Gibbs state in statistical mechanics does not coincide with what we gave in Section 1. There Gibbs states are defined for more general ϕ and then our theorem corresponds to one in statistical mechanics about uniqueness of Gibbs states [3, 10, 6]. In those papers one deal with Σ_n instead of slightly more general Σ_A. Proofs of Theorem 1.4 for Σ_A appear in [2, 11, 12].

The variational characterization (Theorem 1.22) of the Gibbs measure μ_ϕ is due to Lanford-Ruelle [7] in the Σ_n case. For Σ_A one can find it in [2] or [11]. We have followed the proof in [2], which in turn was based on Adler and Weiss' proof [1] of a theorem of Parry (Theorem 1.22 with $\phi = 0$).

The weak Bernoulli condition (Theorem 1.25) was verified for the Σ_n case first by Gallavotti [5]. It was extended to Σ_A independently by each of Bunimovich, Ratner, Ruelle and this author. Theorem 1.28 is taken from Livšic [8] and Sinai [12]. Theorem 1.26 is from [11].

Ruelle gave two proofs of his Perron-Frobenius Theorem, for different circumstances [10, 11]. We have followed parts of each.

1. R. L. Adler, B. Weiss: *Similarity of automorphisms of the torus*, Memoirs of the American Mathematical Society **98**, American Mathematical Society, Providence, R.I. 1970.
2. R. Bowen: Some systems with unique equilibrium states, *Math. Systems Theory* **8** (1974/75), no. 3, 193–202.
3. R. L. Dobrushin: The problem of uniqueness of a Gibbsian random field and the problem of phase transitions, *Func. Anal. and its Appl.* **2** (1968) no. 4, 44–57.
4. N. A. Friedman, D. S. Ornstein: On isomorphism of weak Bernoulli transformations, *Advances in Math.* **5** (1970) 365–394.
5. G. Gallavotti: Ising model and Bernoulli schemes in one dimension, *Comm. Math. Phys.* **32** (1973), 183–190.
6. O. E. Lanford: *Entropy and equilibrium states in classical statistical mechanics.* In "statistical mechanics and mathematical problems" (ed. A. Lenard), Lecture Notes in Physics **20**, pp. 1–113. Springer, Berlin, 1973.
7. O. E. Lanford, D. Ruelle: Observables at infinity and states with short range correlations in statistical mechanics, *Comm. Math. Phys.* **13** (1969) 194–215.

8. A. N. Livšic: Certain properties of the homology of Y-systems, *Mat. Zametki* **10** (1971), 555–564. English translation: *Math. Notes Acad. Sci. USSR* **10** (1971), 758–763.
9. D. Ruelle: *Statistical mechanics: Rigorous results*, W. A. Benjamin, Inc., New York-Amsterdam 1969. ([1])
10. D. Ruelle: Statistical mechanics of a one-dimensional lattice gas, *Comm. Math. Phys.* **9** (1968) 267–278.
11. D. Ruelle: A measure associated with axiom-A attractors, *Amer. J. Math.* **98** (1976), no. 3, 619–654.
12. Ya. Sinai: Gibbs measures in ergodic theory, *Uspehi Mat. Nauk* **27** (1972), no. 4, 21–64. English translation: *Russian Math. Surveys* **27** (1972), no. 4, 21–69.
13. M. Ratner: The central limit theorem for geodesic flows on n-dimensional manifolds of negative curvature, *Israel J. Math.* **16** (1973), 181–197.

[1] Reprint: World Scientific Publishing Co., Inc., River Edge, NJ; Imperial College Press, London, 1999 (note of the editor).

2

GENERAL THERMODYNAMIC FORMALISM

A. Entropy

In Section D of Chapter 1, we defined the number $h_\mu(T, \mathcal{D})$ when T is an endomorphism of a probability space and $\mathcal{D}$ a finite measurable partition. We now define the *entropy* of μ w.r.t. T by

$$h_\mu(T) = \sup_{\mathcal{D}} h_\mu(T, \mathcal{D}),$$

where $\mathcal{D}$ ranges over all finite partitions. We will now turn to some computational lemmas.

We define

$$\begin{aligned} H_\mu(\mathcal{C}|\mathcal{D}) &= H_\mu(\mathcal{C} \vee \mathcal{D}) - H_\mu(\mathcal{D}) \\ &= -\sum_i \mu(D_i) \sum_j \frac{\mu(C_j \cap D_i)}{\mu(D_i)} \log\left(\frac{\mu(C_j \cap D_i)}{\mu(D_i)}\right) \\ &\geq 0\,. \end{aligned}$$

Lemma 1.17 says that $H_\mu(\mathcal{C}|\mathcal{D}) \leq H_\mu(\mathcal{C})$. We write $\mathcal{C} \subset \mathcal{D}$ if each set in $\mathcal{C}$ is a union of sets in $\mathcal{D}$.

2.1. Lemma.

(a) $H_\mu(\mathcal{C}|\mathcal{D}) \leq H_\mu(\mathcal{C}|\mathcal{E})$ *if* $\mathcal{D} \supset \mathcal{E}$.
(b) $H_\mu(\mathcal{C}|\mathcal{D}) = 0$ *if* $\mathcal{D} \supset \mathcal{C}$.
(c) $H_\mu(\mathcal{C} \vee \mathcal{D}|\mathcal{E}) \leq H_\mu(\mathcal{C}|\mathcal{E}) + H_\mu(\mathcal{D}|\mathcal{E})$.
(d) $H_\mu(\mathcal{C}) \leq H_\mu(\mathcal{D}) + H_\mu(\mathcal{C}|\mathcal{D})$.

Proof. Letting $\varphi(x) = -x \log x$, $H_\mu(\mathcal{C}|\mathcal{D}) = \sum_j \sum_i \mu(D_i)\ \varphi\left(\frac{\mu(C_j \cap D_i)}{\mu(D_i)}\right)$. Since $\mathcal{E} \subset \mathcal{D}$, one can rewrite this as

$$H_\mu(\mathcal{C}|\mathcal{D}) = \sum_j \sum_{E \in \mathcal{E}} \mu(E) \sum_{D_i \subset E} \frac{\mu(D_i)}{\mu(E)}\ \varphi\left(\frac{\mu(C_j \cap D_i)}{\mu(D_i)}\right).$$

By the concavity of φ (see the proof of Lemma 1.17) one has $\varphi(\sum a_i x_i) \geq \sum a_i \varphi(x_i)$ where

$$a_i = \frac{\mu(D_i)}{\mu(E)}\ ,\ x_i = \frac{\mu(C_j \cap D_i)}{\mu(D_i)}\,.$$

Hence

$$H_\mu(\mathcal{C}|\mathcal{D}) \leq \sum_j \sum_{E\in\mathcal{E}} \mu(E)\ \varphi\left(\frac{\mu(C_j\cap E)}{\mu(E)}\right) = H_\mu(\mathcal{C}|\mathcal{E})\,.$$

To see (b) one notes that $\mathcal{C}\vee\mathcal{D} = \mathcal{D}$ when $\mathcal{D}\supset\mathcal{C}$. For (c) one writes

$$\begin{aligned} H_\mu(\mathcal{C}\vee\mathcal{D}|\mathcal{E}) &= H_\mu(\mathcal{C}\vee\mathcal{D}\vee\mathcal{E}) - H_\mu(\mathcal{D}\vee\mathcal{E}) + H_\mu(\mathcal{D}\vee\mathcal{E}) - H_\mu(\mathcal{E}) \\ &= H_\mu(\mathcal{C}|\mathcal{D}\vee\mathcal{E}) + H_\mu(\mathcal{D}|\mathcal{E}) \\ &\leq H_\mu(\mathcal{C}|\mathcal{E}) + H_\mu(\mathcal{D}|\mathcal{E}) \end{aligned}$$

by (a). Finally

$$\begin{aligned} H_\mu(\mathcal{C}) &= H_\mu(\mathcal{C}\vee\mathcal{D}) - H_\mu(\mathcal{D}|\mathcal{C}) \\ &\leq H_\mu(\mathcal{C}\vee\mathcal{D}) = H_\mu(\mathcal{D}) + H_\mu(\mathcal{C}|\mathcal{D})\,. \quad \square \end{aligned}$$

2.2. Lemma. *Let T be an endomorphism of a probability space $(X, \mathscr{B}, \mu)$, $\mathcal{C}$ and $\mathcal{D}$ finite partitions. Then*

(a) $H_\mu(T^{-k}\mathcal{C}|T^{-k}\mathcal{D}) = H_\mu(\mathcal{C}|\mathcal{D})$ *for* $k \geq 0$,
(b) $h_\mu(T,\mathcal{C}) \leq h_\mu(T,\mathcal{D}) + H_\mu(\mathcal{C}|\mathcal{D})$,
(c) $h_\mu(T,\mathcal{C}\vee\cdots\vee T^{-n}\mathcal{C}) = h_\mu(T,\mathcal{C})$.

Proof. As μ is T-invariant,

$$\begin{aligned} H_\mu(T^{-k}\mathcal{C}|T^{-k}\mathcal{D}) &= H_\mu(T^{-k}\mathcal{C}\vee T^{-k}\mathcal{D}) - H_\mu(T^{-k}\mathcal{D}) \\ &= H_\mu(\mathcal{C}\vee\mathcal{D}) - H_\mu(\mathcal{D}) = H_\mu(\mathcal{C}|\mathcal{D})\,. \end{aligned}$$

Using Lemma 2.1

$$\begin{aligned} H_\mu(\mathcal{C}\vee\cdots\vee T^{-m+1}\mathcal{C}) &\leq H_\mu(\mathcal{D}\vee\cdots\vee T^{-m+1}\mathcal{D}) \\ &\quad + H_\mu(\mathcal{C}\vee\cdots\vee T^{-m+1}\mathcal{C}|\mathcal{D}\vee\cdots\vee T^{-m+1}\mathcal{D}) \\ &\leq H_\mu(\mathcal{D}\vee\cdots\vee T^{-m+1}\mathcal{D}) \\ &\quad + \sum_{k=0}^{m-1} H_\mu(T^{-k}\mathcal{C}|\mathcal{D}\vee\cdots\vee T^{-m+1}\mathcal{D}) \\ &\leq H_\mu(\mathcal{D}\vee\cdots\vee T^{-m+1}\mathcal{D}) + \sum_{k=0}^{m-1} H_\mu(T^{-k}\mathcal{C}|T^{-k}\mathcal{D}) \\ &= H_\mu(\mathcal{D}\vee\cdots\vee T^{-m+1}\mathcal{D}) + mH_\mu(\mathcal{C}|\mathcal{D})\,. \end{aligned}$$

Dividing by m and letting $m\to\infty$,

$$h_\mu(T,\mathcal{C}) \le h_\mu(T,\mathcal{D}) + H_\mu(\mathcal{C}|\mathcal{D})\,.$$

Set $\mathcal{D} = \mathcal{C} \vee \cdots \vee T^{-n}\mathcal{C}$. Then

$$\frac{1}{m} H_\mu(\mathcal{D} \vee \cdots \vee T^{-m+1}\mathcal{D}) = \frac{1}{m} H_\mu(\mathcal{C} \vee \cdots \vee T^{-m-n+1}\mathcal{C})\,.$$

Letting $m \to \infty$, (as $\frac{m}{m+n} \to 1$) we get

$$h_\mu(T,\mathcal{D}) = h_\mu(T,\mathcal{C})\,. \quad \square$$

2.3. Lemma. *Let X be a compact metric space, $\mu \in \mathscr{M}(X)$, $\varepsilon > 0$ and $\mathcal{C}$ a finite Borel partition. There is a $\delta > 0$ so that $H_\mu(\mathcal{C}|\mathcal{D}) < \varepsilon$ whenever $\mathcal{D}$ is a partition with* $\mathrm{diam}(\mathcal{D}) < \delta$.

Proof. Let $\mathcal{C} = \{C_1, \ldots, C_n\}$. In Lemma 1.23 we showed that, for any $\alpha > 0$, one could find $\delta > 0$ such that whenever $\mathcal{D}$ satisfies $\mathrm{diam}(\mathcal{D}) < \delta$ there is a $\mathcal{E} = \{E_1, \ldots, E_n\} \subset \mathcal{D}$ with

$$\mu(E_i \Delta C_i) < \alpha\,.$$

The expression

$$H_\mu(\mathcal{C}|\mathcal{E}) = \sum_{i,j} \mu(E_j)\ \varphi\left(\frac{\mu(C_j \cap E_i)}{\mu(E_i)}\right)$$

depends continuously upon the numbers

$$\mu(C_j \cap E_i) \quad \text{and} \quad \mu(E_i) = \sum_j \mu(C_j \cap E_i)$$

and vanishes when $\mu(C_j \cap E_i) = \delta_{ij}\ \mu(E_i)$. Hence, for α small, $H_\mu(\mathcal{C}|\mathcal{E}) < \varepsilon$. Then $H_\mu(\mathcal{C}|\mathcal{D}) \le H_\mu(\mathcal{C}|\mathcal{E}) < \varepsilon$ by 2.1 (a). $\square$

2.4. Proposition. *Suppose $T : X \to X$ is a continuous map of a compact metric space, $\mu \in \mathscr{M}_T(X)$ and that $\mathcal{D}_n$ is a sequence of partitions with* $\mathrm{diam}(\mathcal{D}_n) \to 0$. *Then*

$$h_\mu(T) = \lim_{n\to\infty} h_\mu(T,\mathcal{D}_n)\,.$$

Proof. Of course $h_\mu(T) \ge \limsup_n h_\mu(T,\mathcal{D}_n)$. Consider any partition $\mathcal{C}$. By Lemmas 2.2 (b) and 2.3

$$h_\mu(T,\mathcal{C}) \le \liminf_n h_\mu(T,\mathcal{D}_n)\,.$$

Varying $\mathcal{C}$, $h_\mu(T) \le \liminf_n h_\mu(T,\mathcal{D}_n)$. $\square$

A homeomorphism $T : X \to X$ is called *expansive* if there exists $\varepsilon > 0$ so that

$$d(T^k x, T^k y) \le \varepsilon \quad \text{for all } k \in \mathbb{Z} \quad \Rightarrow \quad x = y\,.$$

2.5. Proposition. *Suppose $T : X \to X$ is a homeomorphism with expansive constant ε. Then $h_\mu(T) = h_\mu(T, \mathcal{D})$ whenever $\mu \in \mathscr{M}_T(X)$, and* $\operatorname{diam}(\mathcal{D}) \le \varepsilon$.

Proof. Let $\mathcal{D}_n = T^n\mathcal{D} \vee \cdots \vee \mathcal{D} \vee \cdots \vee T^{-n}\mathcal{D}$. Then $\operatorname{diam}(\mathcal{D}_n) \to 0$ using expansiveness. Hence $h_\mu(T) = \lim_n h_\mu(T, \mathcal{D}_n)$. But $h_\mu(T, \mathcal{D}_n) = h_\mu(T, \mathcal{D})$ by Lemma 2.2 (c). □

Consider the case of $\sigma : \Sigma_A \to \Sigma_A$ and standard partition $\mathcal{U} = \{U_1, \ldots, U_n\}$ where $U_i = \{\underline{x} \in \Sigma_A : x_0 = i\}$. Then σ is expansive and 2.5 gives that $h_\mu(\sigma) = h_\mu(\sigma, \mathcal{U})$ for $\mu \in \mathscr{M}_\sigma(\Sigma_A)$. Now $h_\mu(\sigma, \mathcal{U})$ is what we denoted by $s(\mu)$ in Chapter 1. That $s(\mu) = h_\mu(\sigma)$ implies that the number $s(\mu)$ does not depend on the homeomorphism σ and partition $\mathcal{U}$, but only on σ as an automorphism of the probability space $(\Sigma_A, \mathscr{B}, \mu)$ (because of the definition of $h_\mu(\sigma)$).

2.6. Lemma. $h_\mu(T^n) = nh_\mu(T)$ *for* $n > 0$.

Proof. Let $\mathcal{C}$ be a partition and $\mathcal{E} = \mathcal{C} \vee \cdots \vee T^{-n+1}\mathcal{C}$. Then

$$\begin{aligned} nh_\mu(T, \mathcal{C}) &= \lim_{m\to\infty} \frac{n}{nm} H_\mu(\mathcal{C} \vee \cdots \vee T^{-nm+1}\mathcal{C}) \\ &= \lim_{m\to\infty} \frac{1}{m} H_\mu(\mathcal{E} \vee T^{-n}\mathcal{E} \vee \cdots \vee T^{(-m+1)n}\mathcal{E}) \\ &= h_\mu(T^n, \mathcal{E}) \le h_\mu(T^n) = nh_\mu(T)\,. \end{aligned}$$

Varying $\mathcal{C}$, $nh_\mu(T) \le h_\mu(T^n)$. On the other hand

$$h_\mu(T^n, \mathcal{C}) \le h_\mu(T^n, \mathcal{E})$$

by 2.2 (b) and 2.1 (b). Hence

$$h_\mu(T^n) = \sup_{\mathcal{C}} h_\mu(T^n, \mathcal{C}) \le n \sup_{\mathcal{C}} h_\mu(T, \mathcal{C}) = nh_\mu(T)\,. \quad \square$$

B. Pressure

Throughout this section $T : X \to X$ will be a fixed continuous map on the compact metric space X. We will define the pressure $P(\phi)$ of $\phi \in \mathscr{C}(X)$ in a way which generalizes Section D in Chapter 1.

Let $\mathcal{U}$ be a finite open cover of X, $W_m(\mathcal{U})$ the set of all m-strings

$$\underline{\mathrm{U}} = U_{i_0} U_{i_1} \cdots U_{i_{m-1}}$$

of members of $\mathcal{U}$. One writes $m = m(\underline{\mathrm{U}})$,

$$X(\underline{\mathrm{U}}) = \{x \in X : T^k x \in U_{i_k} \text{ for } k = 0, \ldots, m-1\}$$

$$S_m\phi(\underline{\mathrm{U}}) = \sup\left\{\sum_{k=0}^{m-1}\phi(T^k x) : x \in X(\underline{\mathrm{U}})\right\}.$$

In case $X(\underline{\mathrm{U}}) = \emptyset$, we let $S_m\phi(\underline{\mathrm{U}}) = -\infty$. We say that $\Gamma \subset W_m(\mathcal{U})$ *covers* X if $X = \bigcup_{\underline{\mathrm{U}}\in\Gamma} X(\underline{\mathrm{U}})$. Finally one defines

$$Z_m(\phi,\mathcal{U}) = \inf_{\Gamma}\sum_{\underline{\mathrm{U}}\in\Gamma}\exp(S_m\phi(\underline{\mathrm{U}})),$$

where Γ runs over all subsets of $W_m(\mathcal{U})$ covering X.

2.7. Lemma. *The limit*

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

exists and is finite.

Proof. If $\Gamma_m \subset W_m(\mathcal{U})$ and $\Gamma_n \subset W_n(\mathcal{U})$ each cover X, then

$$\Gamma_m\Gamma_n = \{\underline{\mathrm{UV}} : \underline{\mathrm{U}} \in \Gamma_m,\ \underline{\mathrm{V}} \in \Gamma_n\} \subset W_{m+n}(\mathcal{U})$$

covers X. One sees that

$$S_{m+n}\phi(\underline{\mathrm{UV}}) \le S_m\phi(\underline{\mathrm{U}}) + S_n\phi(\underline{\mathrm{V}})$$

and so

$$\sum_{\underline{\mathrm{UV}}\in\Gamma_m\Gamma_n}\exp(S_{m+n}\phi(\underline{\mathrm{UV}})) \le \sum_{\underline{\mathrm{U}}\in\Gamma_m}\exp(S_m\phi(\underline{\mathrm{U}}))\ \sum_{\underline{\mathrm{V}}\in\Gamma_n}\exp(S_n\phi(\underline{\mathrm{V}})).$$

Thus

$$Z_{m+n}(\phi,\mathcal{U}) \le Z_m(\phi,\mathcal{U})\ Z_n(\phi,\mathcal{U})$$

and $Z_m(\phi,\mathcal{U}) \ge e^{-m\|\phi\|}$. Hence $a_m = \log Z_m(\phi,\mathcal{U})$ satisfies the hypotheses of Lemma 1.18. $\square$

2.8. Proposition. *The limit*

$$P(\phi) = \lim_{\mathrm{diam}(\mathcal{U})\to 0} P(\phi,\mathcal{U})$$

exists (but may be $+\infty$).

Proof. Suppose $\mathcal{V}$ is an open cover refining $\mathcal{U}$, *i.e.*, every $V \in \mathcal{V}$ lies in some $U(V) \in \mathcal{U}$. For $\underline{\mathrm{V}} \in W_m(\mathcal{V})$ let $U(\underline{\mathrm{V}}) = U(V_{i_0})\cdots U(V_{i_{m-1}})$. If $\Gamma_m \subset W_m(\mathcal{V})$ covers X, then $U(\Gamma_m) = \{U(\underline{\mathrm{V}}) : \underline{\mathrm{V}} \in \Gamma_m\} \subset W_m(\mathcal{U})$ covers X.
Let $\gamma = \gamma(\phi,\mathcal{U}) = \sup\{|\phi(x) - \phi(y)| : x, y \in U \text{ for some } U \in \mathcal{U}\}$.
Then $S_m\phi(U(\underline{\mathrm{V}})) \le S_m\phi(\underline{\mathrm{V}}) + m\gamma$ and so $Z_m(\phi,\mathcal{U}) \le e^{m\gamma}Z_m(\phi,\mathcal{V})$, which gives

$$P(\phi, \mathcal{U}) \leq P(\phi, \mathcal{V}) + \gamma\,.$$

Now for any $\mathcal{U}$, all $\mathcal{V}$ with small diameter refine $\mathcal{U}$ and so

$$P(\phi, \mathcal{U}) - \gamma(\phi, \mathcal{U}) \leq \liminf_{\operatorname{diam}(\mathcal{V}) \to 0} P(\phi, \mathcal{V})\,.$$

Letting $\operatorname{diam}(\mathcal{U}) \to 0$, $\gamma(\phi, \mathcal{U}) \to 0$ and

$$\limsup_{\operatorname{diam}(\mathcal{U}) \to 0} P(\phi, \mathcal{U}) \leq \liminf_{\operatorname{diam}(\mathcal{V}) \to 0} P(\phi, \mathcal{V})\,.$$

We are done. $\square$

In cases where confusion may arise we write the topological pressure $P(\phi)$ as $P_T(\phi)$.

2.9. Lemma. *Let $S_n\phi(x) = \sum_{k=0}^{n-1} \phi(T^k x)$. Then*

$$P_{T^n}(S_n\phi) = nP_T(\phi) \text{ for } n > 0\,.$$

Proof. Let $\mathcal{V} = \mathcal{U} \vee \cdots \vee T^{-n+1}\mathcal{U}$. Then $W_m(\mathcal{V})$ and $W_{mn}(\mathcal{U})$ are in one-to-one correspondence; for $\underline{\mathrm{U}} = U_{i_0}U_{i_1}\cdots U_{i_{mn-1}}$ let $\underline{\mathrm{V}} = V_{i_0}\cdots V_{i_{m-1}}$ where $V_{i_k} = U_{i_{kn}} \cap T^{-1}U_{i_{kn+1}} \cap \cdots \cap T^{-n+1}U_{i_{kn+n-1}}$. One sees that $X(\underline{\mathrm{U}}) = X(\underline{\mathrm{V}})$ and $S^T_{mn}\phi(\underline{\mathrm{U}}) = S^{T^n}_m(S_n\phi)(\underline{\mathrm{V}})$. Thus one gets

$$Z^T_{mn}(\phi, \mathcal{U}) = Z^{T^n}_m(S_n\phi, \mathcal{V}) \quad \text{and} \quad nP_T(\phi, \mathcal{U}) = P_{T^n}(S_n\phi, \mathcal{V})\,.$$

As $\operatorname{diam}(\mathcal{U}) \to 0$, $\operatorname{diam}(\mathcal{V}) \to 0$ and so $nP_T(\phi) = P_{T^n}(S_n\phi)$. $\square$

We now come to our first interesting result about the pressure $P(\phi)$.

2.10. Theorem. *Let $T : X \to X$ be a continuous map on a compact metric space and $\phi \in \mathscr{C}(X)$. Then*

$$h_\mu(T) + \int \phi \, d\mu \leq P_T(\phi),$$

for any $\mu \in \mathscr{M}_T(X)$.

We will first need a couple of lemmas.

2.11. Lemma. *Suppose $\mathcal{D}$ is a Borel partition of X such that each $x \in X$ is in the closures of at most M members of $\mathcal{D}$. Then*

$$h_\mu(T, \mathcal{D}) + \int \phi \, d\mu \leq P_T(\phi) + \log M\,.$$

Proof. Let $\mathcal{U}$ be a finite open cover of X each member of which intersects at most M members of $\mathcal{D}$. Let $\Gamma_m \subset W_m(\mathcal{U})$ cover X. For each $B \in \mathcal{D}_m = \mathcal{D} \vee \cdots \vee T^{-m+1}\mathcal{D}$ pick $x_B \in B$ with $\int_B S_m\phi \, d\mu \leq \mu(B)\, S_m\phi(x_B)$. Now

$$\begin{aligned} h_\mu(T, \mathcal{D}) + \int \phi \, d\mu &\leq \frac{1}{m}\left(H_\mu(\mathcal{D}_m) + \int S_m\phi \, d\mu\right) \\ &\leq \frac{1}{m}\sum_B \mu(B)(-\log \mu(B) + S_m\phi(x_B)) \\ &\leq \frac{1}{m}\log\sum_B \exp(S_m\phi(x_B)) \end{aligned}$$

by Lemma 1.1. For each x_B pick $\underline{\mathrm{U}}_B \in \Gamma_m$ with $x_B \in X(\underline{\mathrm{U}}_B)$. This map $B \to \underline{\mathrm{U}}_B$ is at most M^m to one. As $S_m\phi(x_B) \leq S_m\phi(\underline{\mathrm{U}}_B)$, one has

$$\begin{aligned} h_\mu(T, \mathcal{D}) + \int \phi \, d\mu &\leq \frac{1}{m}\log\sum_{\underline{\mathrm{U}}\in\Gamma_m} M^m \ \exp(S_m\phi(\underline{\mathrm{U}})) \\ &\leq \log M + \frac{1}{m}\log Z_m(\phi, \mathcal{U})\,. \end{aligned}$$

Letting $m \to \infty$ and then $\operatorname{diam}(\mathcal{U}) \to 0$, we obtain the desired inequality. $\square$

2.12. Lemma. *Let $\mathcal{A}$ be a finite open cover of X. For each $n > 0$ there is a Borel partition $\mathcal{D}_n$ of X so that*

(a) *D lies inside some member of $T^{-k}\mathcal{A}$ for each $D \in \mathcal{D}_n$ and $k = 0, \ldots, n-1$,*
(b) *at most $n|\mathcal{A}|$ sets in $\mathcal{D}_n$ can have a point in all their closures.*

Proof. Let $\mathcal{A} = \{A_1, \ldots, A_m\}$ and $g_1, \ldots, g_m$ be a partition of unity subordinate to A. Then $G = (g_1, \ldots, g_m) : X \to s_{m-1} \subset \mathbb{R}^m$ where s_{m-1} is an $m-1$ dimensional simplex. Now $\mathcal{U} = \{U_1, \ldots, U_m\}$ is an open cover where $U_i = \{\underline{x} \in s_{m-1} : x_i > 0\}$ and $G^{-1}U_i \subset A_i$. Since $(s_{m-1})^n$ is $nm - n$ dimensional, there is a Borel partition $\mathcal{D}_n^*$ of $(s_{m-1})^n$ so that

(a') each member of $\mathcal{D}_n^*$ lies in some $U_{i_1} \times \cdots \times U_{i_n}$, and
(b') at most nm members of $\mathcal{D}_n^*$ can have a common point in all their closures.

Then $\mathcal{D}_n = L^{-1}\mathcal{D}_n^*$ works where

$$L = (G, G\circ T, \ldots, G\circ T^{n-1}) : X \to (s_{m-1})^n\,. \quad \square$$

Proof of 2.10. Let $\mathcal{C}$ be a Borel partition and $\varepsilon > 0$. By Lemma 2.3 find an open cover $\mathcal{A}$ so that $H_\mu(\mathcal{C}|\mathcal{D}) < \varepsilon$ whenever $\mathcal{D}$ is a partition every member of which is contained in some member of $\mathcal{A}$. Fix $n > 0$, let $\mathcal{E} = \mathcal{C} \vee \cdots \vee T^{-n+1}\mathcal{C}$ and $\mathcal{D}_n$ as in Lemma 2.12. Then (see the proof of 2.6)

$$
\begin{aligned}
h_\mu(T, \mathcal{C}) + \int \phi \, d\mu &\le \frac{1}{n}\left(h_\mu(T^n, \mathcal{E}) + \int S_n\phi \, d\mu\right) \\
&\le \frac{1}{n}\left(h_\mu(T^n, \mathcal{D}_n) + \int S_n\phi \, d\mu\right) + \frac{1}{n} H_\mu(\mathcal{E}|\mathcal{D}_n) \\
&\le \frac{1}{n}\left(P_{T^n}(S_n\phi) + \log(n|\mathcal{A}|)\right) + \frac{1}{n} H_\mu(\mathcal{E}|\mathcal{D}_n)
\end{aligned}
$$

by Lemma 2.11. Now

$$
H_\mu(\mathcal{E}|\mathcal{D}_n) \le \sum_{k=0}^{n-1} H_\mu(T^{-k}\mathcal{C}|\mathcal{D}_n)\,.
$$

Since $\mathcal{D}_n$ refines $T^{-k}\mathcal{A}$ for each k, one has $H_\mu(T^{-k}\mathcal{C}|\mathcal{D}_n) < \varepsilon$ (since μ is T-invariant, $T^{-k}\mathcal{A}$ bears the same relation to $T^{-k}\mathcal{C}$ as $\mathcal{A}$ to $\mathcal{C}$). Hence, using 2.9,

$$
h_\mu(T, \mathcal{C}) + \int \phi \, d\mu \le P_T(\phi) + \frac{1}{n}\log(n|A|) + \varepsilon\,.
$$

Now let $n \to \infty$ and then $\varepsilon \to 0$. $\square$

2.13. Proposition. *Let $T_1 : X_1 \to X_1$, $T_2 : X_2 \to X_2$ be continuous maps on compact metric spaces, $\pi : X_1 \to X_2$ continuous and onto satisfying $\pi \circ T_1 = T_2 \circ \pi$. Then*

$$
P_{T_2}(\phi) \le P_{T_1}(\phi \circ \pi)
$$

for $\phi \in \mathscr{C}(X_2)$.

Proof. For $\mathcal{U}$ an open cover of X_2 one sees that

$$
P_{T_2}(\phi, \mathcal{U}) = P_{T_1}(\phi \circ \pi, \pi^{-1}\mathcal{U})\,.
$$

As in the proof of 2.8

$$
P_{T_1}(\phi \circ \pi, \pi^{-1}\mathcal{U}) \le P_{T_1}(\phi \circ \pi) + \gamma(\phi \circ \pi, \pi^{-1}\mathcal{U})\,.
$$

But $\gamma(\phi \circ \pi, \pi^{-1}\mathcal{U}) = \gamma(\phi, \mathcal{U}) \to 0$ as $\operatorname{diam}(\mathcal{U}) \to 0$. Hence, letting $\operatorname{diam}(\mathcal{U}) \to 0$ we get $P_{T_2}(\phi) \le P_{T_1}(\phi \circ \pi)$. $\square$

C. Variational principle

Let $\mathcal{U}$ be a finite open cover of X. We say that $\varGamma \subset W^*(\mathcal{U}) = \bigcup_{m>0} W_m(\mathcal{U})$ *covers* $K \subset X$ if $K \subset \bigcup_{\underline{\mathrm{U}} \in \varGamma} X(\underline{\mathrm{U}})$. For $\lambda > 0$ and $\varGamma \subset W^*(\mathcal{U})$ define

$$
Z(\varGamma, \lambda) = \sum_{\underline{\mathrm{U}} \in \varGamma} \lambda^{m(\underline{\mathrm{U}})} \ \exp(S_{m(\underline{\mathrm{U}})}\phi(\underline{\mathrm{U}})).
$$

2.14. Lemma. *Let $P = P(\phi, \mathcal{U})$ and $\lambda > 0$. Suppose that $Z(\Gamma, \lambda) < 1$ for some Γ covering X. Then $\lambda \le e^{-P}$.*

Proof. As X is compact we may take Γ finite and $\Gamma \subset \bigcup_{m=1}^{M} W_m(\mathcal{U})$. Then $Z(\Gamma^n, \lambda) \le Z(\Gamma, \lambda)^n$ where $\Gamma^n = \{\underline{\mathrm{U}}_1\underline{\mathrm{U}}_2 \cdots \underline{\mathrm{U}}_n : \underline{\mathrm{U}}_i \in \Gamma\}$. Letting $\Gamma^* = \bigcup_{n=1}^{\infty} \Gamma^n$, one has

$$Z(\Gamma^*, \lambda) = \sum_{n=1}^{\infty} Z(\Gamma^n, \lambda) < \infty\,.$$

Fix N and consider $x \in X$. Since Γ covers X, one can find $\underline{\mathrm{U}} = \underline{\mathrm{U}}_1\underline{\mathrm{U}}_2 \cdots \underline{\mathrm{U}}_n \in \Gamma^*$ with

(a) $x \in X(\underline{\mathrm{U}})$, and
(b) $N \le m(\underline{\mathrm{U}}) < N + M$.

Let $\underline{\mathrm{U}}^*$ be the first N symbols of $\underline{\mathrm{U}}$. Then

$$S_N\phi(\underline{\mathrm{U}}^*) \le S_{m(\underline{\mathrm{U}})}\phi(\underline{\mathrm{U}}) + M\|\phi\|\,.$$

For Γ^N the set of $\underline{\mathrm{U}}^*$ so obtained,

$$\lambda^N \sum_{\Gamma^N} \exp S_N\phi(\underline{\mathrm{U}}^*) \le \max\left\{1, \lambda^{-M}\right\}\; e^{M\|\phi\|}\; Z(\Gamma^*, \lambda),$$

or $\lambda^N Z_N(\phi, \mathcal{U}) \le$ constant. It follows that $\lambda \le e^{-P}$. $\square$

Let δ_x be the unit-measure concentrated on the point x. Define

$$\begin{aligned} &\delta_{x,n} = n^{-1}(\delta_x + \delta_{Tx} + \cdots + \delta_{T^{n-1}x}) \\ \text{and} \quad &V(x) = \{\mu \in \mathscr{M}(X) : \delta_{x,n_k} \to \mu \ \text{ for some } n_k \to \infty\}\,. \end{aligned}$$

$V(x) \ne \emptyset$ as $\mathscr{M}(X)$ is a compact metric space. Now $T^*\delta_{x,n} = \delta_{Tx,n}$ and for $f \in \mathscr{C}(X)$, $|T^*\delta_{x,n}(f) - \delta_{x,n}(f)| = n^{-1}|f(T^n x) - f(x)| \le 2n^{-1}\|f\|$. This shows $V(x) \subset \mathscr{M}_T(X)$.

Let E be a finite set, $\underline{a} = (a_0, \ldots, a_{k-1}) \in E^k$. One defines the distribution $\mu_{\underline{a}}$ on E by

$$\begin{aligned} &\mu_{\underline{a}}(e) = k^{-1}(\text{number of } j \text{ with } a_j = e) \\ \text{and} \quad &H(\underline{a}) = -\sum_{e \in E} \mu_{\underline{a}}(e) \log \mu_{\underline{a}}(e)\,. \end{aligned}$$

2.15. Lemma. *Let $x \in X$, $\mu \in V(x)$, $\mathcal{U}$ a finite open cover of X and $\varepsilon > 0$. There are m and arbitrarily large N for which one can find $\underline{\mathrm{U}} \in W_N(\mathcal{U})$ satisfying the following*

(a) $x \in X(\underline{\mathrm{U}})$,
(b) $S_N\phi(\underline{\mathrm{U}}) \le N(\int \phi d\mu + \gamma(\mathcal{U}) + \varepsilon)$,

(c) $\underline{\mathrm{U}}$ *contains a subword of length* $km \geq N - m$ *which, when viewed as* $\underline{a} = a_0, \ldots, a_{k-1} \in (\mathcal{U}^m)^k$ *satisfies*

$$\frac{1}{m} H(\underline{a}) \leq h_\mu(T) + \varepsilon.$$

Proof. Let $\mathcal{U} = \{U_1, \ldots, U_q\}$. Recall that

$$\gamma(\mathcal{U}) = \sup\{|\phi(y) - \phi(z)| : y, z \in U_i \text{ for some } i\}\,.$$

Pick a Borel partition $\mathcal{C} = \{C_1, \ldots, C_q\}$ with $\overline{C}_i \subset U_i$. There is an m so that

$$\frac{1}{m} H_\mu(\mathcal{C} \vee \cdots \vee T^{-m+1}\mathcal{C}) \leq h_\mu(T, \mathcal{C}) + \frac{\varepsilon}{2} \leq h_\mu(T) + \frac{\varepsilon}{2}.$$

Let $\delta_{x,n_j} \to \mu$. For $n' > n$ one has

$$\delta_{x,n'} = \frac{n}{n'}\ \delta_{x,n} + \frac{n' - n}{n'}\ \delta_{T^n x, n'-n}\,.$$

If we replaced n_k by the nearest multiple of m, this formula shows that μ would still be the limit. Thus we assume $n_j = mk_j$.

Let $D_1, \ldots, D_t$ be the nonempty members of $\mathcal{C} \vee \cdots \vee T^{-m+1}\mathcal{C}$ and for each D_i find a compact $K_i \subset D_i$ with $\mu(D_i \backslash K_i) < \beta$ ($\beta > 0$ small). Each D_i is contained in some member of $\mathcal{U} \vee \cdots \vee T^{-m+1}\mathcal{U}$ and one can find an open set $V_i \supset K_i$ for which this is still true. Furthermore we may assume $V_i \cap V_j = \emptyset$ for $i \neq j$. Now enlarge each V_i to a Borel set V_i^* still contained in a member of $\mathcal{U} \vee \cdots \vee T^{-m+1}\mathcal{U}$ and so that $\{V_1^*, \ldots, V_t^*\}$ is a Borel partition of X.

Now fix $n_j = mk_j$. Let M_i be the number of $s \in [0, n_j)$ with $T^s x \in V_i^*$ and $M_{i,r}$ the number of such $s \equiv r \pmod m$.
Define

$$p_{i,r} = M_{i,r}/k_j$$

and $p_i = M_i/n_j = \frac{1}{m}(p_{i,0} + \cdots + p_{i,m-1})$. As $\delta_{x,n_j} \to \mu$, one has

$$\liminf_{j\to\infty} p_i \geq \mu(K_i) \geq \mu(D_i) - \beta,$$

and $\limsup_{j\to\infty} p_i \leq \mu(K_i) + t\beta \leq \mu(D_i) + t\beta$. For β small enough and j large enough one has

$$\begin{aligned}\frac{1}{m}\left(-\sum_i p_i \log p_i\right) &\leq \frac{1}{m}\left(-\sum_i \mu(D_i)\log\mu(D_i)\right) + \frac{\varepsilon}{2} \\ &\leq h_\mu(T) + \varepsilon\,.\end{aligned}$$

By the concavity of $\varphi(x) = -x\log x$ (see 1.17)

$$\varphi(p_i) \geq \sum_{r=0}^{m-1} \frac{1}{m}\ \varphi(p_{i,r})$$

and so

$$\sum_i \varphi(p_i) \geq \frac{1}{m} \sum_{r=0}^{m-1} \sum_i \varphi(p_{i,r}) .$$

For some $r \in [0, m)$ one must have $\sum_i \varphi(p_{i,r}) \leq \sum_i \varphi(p_i)$ and so

$$\frac{1}{m} \sum_i \varphi(p_{i,r}) \leq h_\mu(T) + \varepsilon .$$

For $N = n_j + r$ with j large we form $\underline{\mathrm{U}} = U_0 U_1 \cdots U_{N-1} \in \mathfrak{U}^N$ as follows. For $s < r$ pick $U_s \in \mathfrak{U}$ containing $T^s x$. For each V_i^* we choose $U_{0,i} \cap T^{-1} U_{1,i} \cap \cdots \cap T^{-m+1} U_{m-1,i} \supset V_i^*$. For $s > r$ we write $s = r + mp + q$ with $p \geq 0$, $m > q \geq 0$, pick i with $T^{r+mp} x \in V_i^*$ and let $U_s = U_{q,i}$. Letting

$$a_p = U_{0,i} U_{1,i} \cdots U_{m-1,i}$$

we have

$$\underline{\mathrm{U}} = U_0 \cdots U_{r-1} a_0 a_1 \cdots a_{k_j - 1} .$$

Now $\underline{a} = (a_0 a_1 \cdots a_{k_j-1})$ has its distribution $\mu_{\underline{a}}$ on $\mathfrak{U}^m$ given by the probabilities $\{p_{i,r}\}_{i=1}^t$ and some zeros.
So

$$\frac{1}{m} H(\underline{a}) = \frac{1}{m} \sum_i \varphi(p_{i,r}) \leq h_\mu(T) + \varepsilon.$$

We have yet to check (b). Since $\delta_{x,n_j} \to \mu$, for j large we will have $\left|\frac{1}{N}\delta_{x,N}(\phi) - \int \phi \, d\mu\right| < \varepsilon$ or $S_N\phi(x) \leq N(\int \phi \, d\mu + \varepsilon)$. As $x \in X(\underline{\mathrm{U}})$, $S_N \phi(\underline{\mathrm{U}}) \leq S_N \phi(x) + N\gamma(\mathfrak{U})$. $\square$

2.16. Lemma. *Fix a finite set E and $h \geq 0$. Let $R(k,h) = \{\underline{a} \in E^k : H(\underline{a}) \leq h\}$. Then*

$$\limsup_{k \to \infty} \frac{1}{k} \log |R(k,h)| \leq h .$$

Proof. For any distribution ν on E and $\alpha \in (0,1)$ consider

$$R_k(\nu) = \{\underline{a} \in E^k : |\mu_{\underline{a}}(e) - \nu(e)| < \alpha \;\; \forall e \in E\} .$$

Let μ be the Bernoulli measure on $\Sigma = \prod_{i=0}^\infty E$ with the distribution

$$\mu(e) = (1-\alpha)\nu(e) + \alpha/|E| .$$

Each $\underline{a} \in R_k(\nu)$ corresponds to a cylinder set $C_{\underline{a}}$ of Σ. Since each $e \in E$ occurs in $\underline{a}$ at most $k(\nu(e) + \alpha)$ times,

$$\mu(C_{\underline{a}}) \geq \prod_e \mu(e)^{k(\nu(e)+\alpha)} .$$

As the $C_{\underline{a}}$ are disjoint and have total measure 1,

$$1 \geq |R_k(\nu)| \prod_e \mu(e)^{k(\nu(e)+\alpha)},$$

$$\text{or} \quad \frac{1}{k} \log |R_k(\nu)| \leq \sum_e -(\nu(e)+\alpha) \log \mu(e)$$
$$\leq H(\mu) + \sum_e 3\alpha |\log \mu(e)| \,.$$

As $\mu(e) \geq \alpha/|E|$, we get

$$\frac{1}{k} \log |R_k(\nu)| \leq H(\mu) + 3\alpha|E|(\log|E| - \log \alpha) \,.$$

When $\alpha \to 0$, the second term on the right approaches 0 and $H(\mu) \to H(\nu)$ uniformly in ν. Hence, for any $\varepsilon > 0$ one can find α small enough that

$$\frac{1}{k} \log |R_k(\nu)| \leq H(\mu) + \varepsilon,$$

for all k and ν.

Once α is so chosen, let $\mathcal{N}$ be a finite set of distributions on E so that

(a) $H(\nu) \leq h$ for $\nu \in \mathcal{N}$, and
(b) if $H(\nu') \leq h$ then for some $\nu \in \mathcal{N}$ one has

$$|\nu'(e) - \nu(e)| < \alpha \quad \text{for all } \ e.$$

Then $R(k,h) \subset \bigcup_{\nu \in \mathcal{N}} R_k(\nu)$,

$$\frac{1}{k} \log |R(k,h)| \leq \frac{1}{k} \log |\mathcal{N}| + h + \varepsilon$$
$$\text{and} \qquad \limsup_{k \to \infty} \frac{1}{k} \log |R(k,h)| \leq h + \varepsilon.$$

Now let $\varepsilon \to 0$. $\square$

2.17. Variational Principle. *Let $T : X \to X$ be a continuous map on a compact metric space and $\phi \in \mathscr{C}(X)$. Then*

$$P_T(\phi) = \sup_\mu \left(h_\mu(T) + \int \phi \; d\mu \right)$$

where μ runs over $\mathscr{M}_T(X)$.

Proof. Let $\mathfrak{U}$ be a finite cover of X and $\epsilon > 0$. For each $m > 0$ let X_m be the set of points $x \in X$ for which 2.15 holds with this m and some $\mu \in V(x)$. By 2.15 $X = \bigcup_m X_m$ since $V(x) \neq \emptyset$. For $u \in \mathbb{R}$ let $Y_m(u)$ be the set of $x \in X_m$ for which 2.15 holds for some $\mu \in V(x)$ with $\int \phi d\mu \in [u-\varepsilon, u+\varepsilon]$. Set

$$c = \sup_{\mu} \left(h_\mu(T) + \int \phi \, d\mu \right).$$

For $x \in Y_m(u)$ the μ satisfies $h_\mu(T) \le c - u + \varepsilon$.

The $\underline{a} \in (\mathcal{U}^m)^k$ appearing in 2.15 (c) for $x \in Y_m(u)$ lie in $R(k, m(c-u+2\varepsilon), \mathcal{U}^m)$. The number of possibilities for $\underline{\mathrm{U}}$ for any fixed $N = km$ is hence at most

$$b(N) = |E|^m \; |R(k, m(c-u+2\varepsilon), \mathcal{U}^m)| .$$

By 2.16

$$\limsup_{N\to\infty} \frac{1}{N} \log b(N) \le c - u + 2\varepsilon .$$

Let $\Gamma = \Gamma_{m,u}$ be the collection of all $\underline{\mathrm{U}}$ showing up in the present situation for some N greater than a fixed N_0. Then Γ covers $Y_m(u)$ and

$$Z(\Gamma, \lambda) \le \sum_{N=N_0}^{\infty} \lambda^N \; b(N) \; \exp(N(u + 2\varepsilon + \gamma(\mathcal{U}))) .$$

For large enough N_0, $b(N) \le \exp(N(c-u+3\varepsilon))$ and

$$\begin{aligned} Z(\Gamma, \lambda) &\le \sum_{N=N_0}^{\infty} \lambda^N \; \exp(N(c + 5\varepsilon + \gamma(\mathcal{U}))) . \\ &\le \sum_{N=N_0}^{\infty} \beta^N = \frac{\beta^{N_0}}{1-\beta} , \end{aligned}$$

where $\beta = \lambda \exp(c + 5\varepsilon + \gamma(\mathcal{U})) < 1$.

We have seen that for $\lambda < \exp(-(c + 5\varepsilon + \gamma(\mathcal{U})))$ any $Y_m(u)$ can be covered by $\Gamma \subset W^*(\mathcal{U})$ with $Z(\Gamma, \lambda)$ as small as desired. As $X = \bigcup_{m=1}^{\infty} X_m$ and $X_m = Y_m(u_1) \cup \cdots \cup Y_m(u_r)$ where $u_1, \ldots, u_r$ are ε-dense in $[-\|\phi\|, \|\phi\|]$, taking unions of such $\Gamma's$ we obtain a Γ covering X with $Z(\Gamma, \lambda) < 1$. By Lemma 2.14, $\lambda \le e^{-P(\phi, \mathcal{U})}$ or

$$P(\phi, \mathcal{U}) \le c + 5\varepsilon + \gamma(\mathcal{U}) .$$

As ε was arbitrary, $P(\phi, \mathcal{U}) \le c + \gamma(\mathcal{U})$.
Finally

$$\begin{aligned} P(\phi) &\le \lim_{\mathrm{diam}(\mathcal{U})\to 0} P(\phi, \mathcal{U}) \\ &\le \lim_{\mathrm{diam}(\mathcal{U})\to 0} (c + \gamma(\mathcal{U})) = c . \end{aligned}$$

The inequality $c \le P(\phi)$ follows from Theorem 2.10. □

2.18. Corollary. *Suppose $\{X_\alpha\}_{\alpha \in \Lambda}$ is a family of compact subsets of X and $TX_\alpha \subset X_\alpha$ for each α. Then*

$$P_T(\phi) = \sup_{\alpha} P_{T|X_\alpha}(\phi|_{X_\alpha}) .$$

Proof. If $\mu \in \mathscr{M}_T(X_\alpha)$, then $\mu \in \mathscr{M}_T(X)$ and

$$P_T(\phi) \geq h_\mu(T) + \int \phi \, d\mu \, .$$

Hence

$$P_T(\phi) \geq \sup_{\mu \in \mathscr{M}_T(X_\alpha)} \left(h_\mu(T) + \int \phi \, d\mu \right) = P_{T|X_\alpha}(\phi|_{X_\alpha}) \, .$$

If $x \in X_\alpha$, then $V(x) \subset \mathscr{M}_T(X_\alpha)$ and so

$$\begin{aligned} c' &= \sup \left\{ h_\mu(T) + \int \phi \, d\mu \ : \ \mu \in \bigcup_{x \in X} V(x) \right\} \\ &\leq \sup_\alpha P_{T|X_\alpha}(\phi|_{X_\alpha}) \, . \end{aligned}$$

In the proof of 2.17 what was actually used about the number c was $c \geq h_\mu(T) + \int \phi \, d\mu$ for $\mu \in V(x)$. So c' would work just as well there to yield $P_T(\phi) \leq c'$. $\square$

D. Equilibrium states

If $\mu \in \mathscr{M}_T(X)$ satisfies $h_\mu(T) + \int \phi \, d\mu = P_T(\phi)$, then μ is called an *equilibrium state* for ϕ (w.r.t. T). The Gibbs state μ_ϕ of $\phi \in \mathscr{F}_A$ in Chapter 1 was shown to be the unique equilibrium state for such a ϕ.

2.19. Proposition. *Suppose that for some $\varepsilon > 0$ one has $h_\mu(T, \mathcal{D}) = h_\mu(T)$ whenever $\mu \in \mathscr{M}_T(X)$ and* $\operatorname{diam}(\mathcal{D}) < \varepsilon$. *Then every $\phi \in \mathscr{C}(X)$ has an equilibrium state.*

Proof. We show that $\mu \mapsto h_\mu(T)$ is upper semi-continuous on $\mathscr{M}_T(X)$. Then $\mu \mapsto h_\mu(T) + \int \phi d\mu$ will be also, and the proposition follows from 2.17 and the fact that an u.s.c. function on a compact space assumes its supremum.

Fixing $\mu \in \mathscr{M}_T(X)$, $\alpha > 0$, and $\mathcal{D} = \{D_1, \dots, D_n\}$ with $\operatorname{diam}(\mathcal{D}) < \varepsilon$, one has $\frac{1}{m} H_\mu(\mathcal{D} \vee \cdots \vee T^{-m+1}\mathcal{D}) \leq h_\mu(T) + \alpha$ for some m. Let $\beta > 0$ and pick a compact set $K_{i_0, \dots, i_{m-1}} \subset \bigcap_{k=0}^{m-1} T^{-k} D_{i_k}$ with

$$\mu \left(\bigcap_k T^{-k} D_{i_k} \backslash K_{i_0, \dots, i_{m-1}} \right) < \beta \, .$$

Then $D_i \supset L_i = \bigcup_{j=0}^{m-1} \bigcup \{ T^j K_{i_0, \dots, i_{m-1}} \ : \ i_j = i \}$. As the L_i are disjoint compact sets, one can find a partition $\mathcal{D}' = \{D'_1, \dots, D'_n\}$ with $\operatorname{diam}(\mathcal{D}') < \varepsilon$ and $L_i \subset \operatorname{int}(D'_i)$. One then has

$$K_{i_0, \dots, i_{m-1}} \subset \operatorname{int} \left(\bigcap_k T^{-k} D'_{i_k} \right) .$$

If ν is close to μ in the weak topology, one will have

$$\nu\left(\bigcap_k T^{-k} D'_{i_k}\right) \geq \mu(K_{i_0,\ldots,i_{m-1}}) - \beta$$

and $\left|\nu\left(\bigcap_k T^{-k} D'_{i_k}\right) - \mu\left(\bigcap_k T^{-k} D_{i_k}\right)\right| \leq 2\beta n^m$. For β small enough, this implies

$$\begin{aligned} h_\nu(T) = h_\nu(T, \mathcal{D}') &\leq \frac{1}{m} H_\nu(\mathcal{D}' \vee \cdots \vee T^{-m+1}\mathcal{D}') \\ &\leq \frac{1}{m} H_\mu(\mathcal{D} \vee \cdots \vee T^{-m+1}\mathcal{D}) + \alpha \leq h_\mu(T) + 2\alpha\,. \quad \square \end{aligned}$$

2.20. Corollary. *If T is expansive, every $\phi \in \mathscr{C}(X)$ has an equilibrium state.*

Proof. Recall 2.5. $\square$

One notices that the condition in 2.19 has nothing to do with ϕ. Taking $\phi = 0$, one defines the *topological entropy* of T by

$$h(T) = P_T(0)\,.$$

The motivation for this chapter comes from two places: the theory of Gibbs states from statistical mechanics and topological entropy from topological dynamics (see references). Conditions on ϕ become important for the *uniqueness* of equilibrium state and then only after stringent conditions have been placed on the homeomorphism T. The Axiom A diffeomorphisms will be close enough to the subshifts $\sigma : \Sigma_A \to \Sigma_A$ so that one can prove uniqueness statements.

References

The definition of $h_\mu(T)$ is due to Kolmogorov and Sinai (see [2]). For expansive T Ruelle [15] defined $P_T(\phi)$ and proved Theorems 2.10, 2.17 and 2.20. For general T the definition and results are due to Walters [16].

In the transition from Σ_A to a general compact metric space X, most of the work has to do with the more complicated topology of X. Walters' paper is therefore closely related to earlier work on the topological entropy $h(T)$, *i.e.*, the case $\phi = 0$. The definition of $h(T)$ was made by Adler, Konheim and McAndrew [1]. The theorems for this case are due to Goodwyn [10] (Theorem 2.10), Dinaburg [6] (X finite dimensional, 2.17), Goodman [8] (general X, 2.17), and Goodman [9] (2.20). For subshifts these results were proved earlier by Parry [14]. The proofs we have given in these notes are adaptations of [4].

Gurevič [11] gives a T where $\phi = 0$ has no equilibrium states and Misiurewicz [13] gives such a T which is a diffeomorphism. The condition in 2.19 is satisfied by a class of maps which includes all affine maps on Lie groups [3] and Misiurewicz [13] showed that equilibrium states exist under a somewhat weaker condition.

Ruelle [15] showed that for expansive T a Baire set of ϕ have unique equilibrium states. Goodman [9] gives a minimal subshift where $\phi = 0$ has more than one equilibrium state. I believe mathematical physicists know of specific ϕ on Σ_n which do not have unique equilibrium states; in statistical mechanics one looks at $\mathbb{Z}^m$ actions instead of just homeomorphisms and gets nonuniqueness for $m \geq 2$ even with simple ϕ's. Uniqueness was proved in [5] for certain ϕ when T satisfies expansiveness and a very restrictive condition called specification; this result has been carried over to flows by Franco-Sanchez [7].

Finally we mention a very interesting result in a different direction. Let $T : M \to M$ be a continuous map on a compact manifold and λ an eigenvalue of the map $T_* : H_1(M) \to H_1(M)$ on one-dimensional homology. Then Manning [12] showed $h(T) \geq \log|\lambda|$. It is conceivable that this inequality is true for λ for any $H_k(M)$ (not just $k = 1$) provided T is a diffeomorphism.

1. R. L. Adler, A. G. Konheim and M. H. McAndrew: Topological entropy, *Trans. Amer. Math. Soc.* **114** (1965), 309-319.

2. P. Billingsley: *Ergodic Theory and Information*, Wiley (1965). ([1])
3. R. Bowen: Entropy-expansive maps, *Trans. Amer. Math. Soc.* **164** (1972), 323-331.
4. R. Bowen: Topological entropy for noncompact sets, *Trans. Amer. Math. Soc.* **184** (1973), 125-136.
5. R. Bowen: Some systems with unique equilibrium states, *Math. Systems Theory* **8** (1974/75), no. 3, 193-202.
6. E. I. Dinaburg: On the relations among various entropy characteristics of dynamical systems, *Math. USSR Izvestia* **5** (1971), 337-378.
7. E. Franco-Sanchez: *Flows with unique equilibrium states*, Ph.D. thesis. Univ. of California, Berkeley, 1974.
8. T. N. T. Goodman: Relating topological entropy and measure entropy, *Bull. London Math. Soc.* **3** (1971), 176-180.
9. T. N. T. Goodman: Maximal measures for expansive homeomorphisms, *J. London Math. Soc.* (2) **5** (1972), 439-444.
10. L. W. Goodwyn: Topological entropy bounds measure-theoretic entropy, *Proc. Amer. Math. Soc.* **23** (1969), 679-688.
11. B. M. Gurevič: Topological entropy of denumerable Markov chains, *Soviet Math. Dokl.* **10** (1969), 911-915.
12. A. Manning: *Topological entropy and the first homology group*, in "Dynamical Systems" - Warwick 1974 (Proc. Symp. Appl. Topology and Dynamical Systems, Univ. Warwick, Coventry, 1973/1974), pp. 185-190. Lect. Notes in Math. **468**, Springer, Berlin, 1975.
13. M. Misiurewicz: Diffeomorphism without any measure with maximal entropy, *Bull. Acad. Polon. Sci. Sér. Sci. Math. Astronom. Phys.* **21** (1973), 903–910.
14. W. Parry: Intrinsic Markov chains, *Trans. Amer. Math. Soc.* **112** (1964), 55-66.
15. D. Ruelle: Statistical mechanics on a compact set with $\mathbb{Z}^\nu$ action satisfying expansiveness and specification, *Trans. Amer. Math. Soc.* **185** (1973), 237-251.
16. P. Walters: A variational principle for the pressure of continuous transformations, *Amer. J. Math.* **97** (1975), no. 4, 937-971.

[1] Reprint: Robert E. Krieger Publishing Co., Huntington, N.Y., 1978 (note of the editor).

3

AXIOM A DIFFEOMORPHISMS

A. Definition

We now suppose that $f : M \to M$ is a diffeomorphism of a compact $\mathscr{C}^\infty$ Riemannian manifold M. Then the derivative of f can be considered a map $df : TM \to TM$ where $TM = \bigcup_{x \in M} T_x M$ is the tangent bundle of M and $df_x : T_x M \to T_{f(x)} M$.

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

$$T_x M = E_x^u \oplus E_x^s$$

of subspaces so that

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

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

and

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

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

Remark. *Condition (c) actually follows from the others. $E^u = \bigcup_{x \in \Lambda} E_x^u$ and $E^s = \bigcup_{x \in \Lambda} E_x^s$ are continuous subbundles of $T_\Lambda M = \bigcup_{x \in \Lambda} T_x M$ and $T_\Lambda M = E^u \oplus E^s$.*

A point $x \in M$ is *non-wandering* if

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

for every neighborhood U of x. The set $\Omega = \Omega(f)$ of all non-wandering points is seen to be closed and f-invariant. A point x is *periodic* if $f^n x = x$ for some $n > 0$; clearly such an x is in $\Omega(f)$.

Definition. *f satisfies* Axiom A *if $\Omega(f)$ is hyperbolic and* $\Omega(f) = \overline{\{x : x \text{ is periodic}\}}$.

This definition is due to Smale [14]. A type of f studied extensively by Russian mathematicians is the Anosov diffeomorphism: f is *Anosov* if M is hyperbolic [2]. We shall see a little later that such diffeomorphisms always satisfy Axiom A. Right now we mention that it is unknown whether $\Omega(f) = M$ for every Anosov f. The reader should study the examples in [14].

The Riemannian metric on M is used to state condition (b) in the definition of hyperbolic set. The truth of this condition does not depend upon which metric is used although the constants c and λ do. A metric is *adapted* (to an Axiom A f) if $\Omega(f)$ is hyperbolic with respect to it with $c = 1$.

3.1. Lemma. *Every Axiom A diffeomorphism has an adapted metric.*

Proof. This lemma is due to Mather. See [8] for a proof. □

We will *always* use an adapted metric. This will keep various estimates a bit simpler. For $x \in M$ define

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

The following stable manifold theorem is the main analytic fact behind the behavior of Axiom A diffeomorphisms.

3.2. Theorem. *Let Λ be a hyperbolic set for a $\mathscr{C}^r$ diffeomorphism f. For small $\varepsilon > 0$*

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

Proof. See Hirsch and Pugh [8]. □

One consequence of 3.2 is that $W^s_\varepsilon(x) \subset W^s(x)$ for $x \in \Lambda$. One then sees that

$$W^s(x) = \bigcup_{n \geq 0} f^{-n} W^s_\varepsilon(f^n x)$$

for $x \in \Lambda$. Similarly

$$W^u(x) = \bigcup_{n \geq 0} f^{n} W^u_\varepsilon(f^{-n} x)\,.$$

3.3. Canonical Coordinates. *Suppose f satisfies Axiom A. For any small $\varepsilon > 0$ there is a $\delta > 0$ so that $W^s_\varepsilon(x) \cap W^u_\varepsilon(y)$ consists of a single point $[x, y]$ whenever $x, y \in \Omega(f)$ and $d(x, y) \le \delta$. Furthermore $[x, y] \in \Omega(f)$ and*

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

is continuous.

Proof. See Smale [14]. The first statement follows because the intersection $W^s_\varepsilon(x) \cap W^u_\varepsilon(x) = \{x\}$ is transversal and such intersections are preserved under small perturbation. To get $[x, y] \in \Omega(f)$ uses that the periodic points are dense in $\Omega(f)$. In the Anosov case one of course has $[x, y] \in M$ and thus canonical coordinates on M (instead of $\Omega(f)$) without any assumption on periodic points. □

3.4. Lemma. *Let Λ be a hyperbolic set. Then there is an $\varepsilon > 0$ so 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. Otherwise $y \in W^s_\varepsilon(x) \cap W^u_\varepsilon(x)$. x is also in this intersection and so $y = x$ by 3.3. □

B. Spectral decomposition

From now on f will always be an Axiom A diffeomorphism.

3.5. Spectral Decomposition. *One can write $\Omega(f) = \Omega_1 \cup \Omega_2 \cup \cdots \cup \Omega_s$ where the 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 pairwise 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 ([14, 4]). For $p \in \Omega$ periodic let $X_p = \overline{W^u(p) \cap \Omega}$. Let δ be as in 3.3. We claim that

$$X_p = B_\delta(X_p) = \{y \in \Omega : d(y, X_p) < \delta\}\,.$$

As periodic points are dense in Ω, it is enough to see that a periodic $q \in B_\delta(X_p)$ is in X_p. Pick $x \in W^u(p) \cap \Omega$ with $d(x, q) < \delta$ and consider $x' = [x, q] \in W^u(p) \cap W^s(q) \cap \Omega$. Letting $f^m p = p$ and $f^n q = q$, one has

$$\begin{aligned} & f^{kmn}x' \in f^{kmn}W^u(p) = W^u(f^{kmn}p) = W^u(p) \\ \text{and} \quad & d(f^{kmn}x', q) = d(f^{kmn}x', f^{kmn}q) \to 0 \quad \text{as } k \to \infty\,. \end{aligned}$$

So $q \in X_p$.

Notice that $fX_p = X_{f(p)}$ since $fW^u(p) = W^u(f(p))$. If $q \in X_p$ as above, then $W^u_\delta(q) \subset X_p$ and

$$\begin{aligned} W^u(q) &= \bigcup_{k\geq 0} f^{kmn} W^u_\delta(q) \\ &\subset \bigcup_{k\geq 0} f^{kmn} X_p = X_p\,. \end{aligned}$$

(Note that $y \in W^u(q)$ iff $f^{-kmn}y \to q$ as $k \to \infty$.) It follows that $X_q \subset X_p$. If x' is as above, then $f^{kmn}x' \in X_q$ for large k as $X_q = B_\delta(X_q)$ is open in Ω. As $f^{imn}X_q = X_{f^{imn}q} = X_q$, one has $f^{jmn}x' \in X_q$ for all j and

$$p = \lim_{j\to -\infty} f^{jmn}x' \in \overline{X}_q = X_q\,.$$

The above argument with the roles of p and q reversed gives $X_p \subset X_q$. In summary, if $q \in X_p$ with p, q periodic, $X_p = X_q$.

Now any two X_p, X_q are either disjoint or equal. For if $X_q \cap X_q \neq \emptyset$, then this intersection is open in Ω and hence contains a periodic point r; then $X_p = X_r = X_q$. Now

$$\Omega = \bigcup_{p \text{ periodic}} B_\delta(X_p) = \bigcup_p X_p,$$

and so by compactness (the X_p are open) let

$$\Omega = X_{p_1} \cup \cdots \cup X_{p_t}$$

with the X_{p_j}'s pairwise disjoint. Then $f(X_{p_j}) = X_{fp_j}$ intersects and hence equals some X_{p_i}. So f permutes the X_{p_j}'s and the Ω_i are just the union of the X_{p_j}'s in the various cycles of the permutation.

The transitivity in (a) is implied by the mixing in (b). We finish by showing $f^N : X_r \to X_r$ is mixing whenever r is periodic and N positive with $f^N X_r = X_r$. Suppose U, V are nonempty subsets of X_r open in X_r (*i.e.*, in Ω). Pick periodic points $p \in U$ and $q \in V$, say $f^m p = p$, $f^n q = q$. For each $0 \leq j < mn$ with $f^j p \in X_r$ one can find a point x'_j as in the beginning of this proof so that

$$x'_j \in f^j U \quad \text{and} \quad f^{kmn}x'_j \in V \text{ for large } k\,.$$

Writing $tN = kmn + j$, $0 \leq j \leq mn$, we have $f^j p = f^{tN} p \in X_r$ and

$$f^{kmn}x'_j = f^{tN}(f^{-jN}x'_j) \in f^{tN}U \cap V$$

provided k is large. Then $f^{tN}U \cap V \neq \emptyset$ for large t and $f^N|_{X_r}$ is topologically mixing. $\square$

The Ω_i in the spectral decomposition of $\Omega(f)$ are called the *basic sets* of f. Notice that if $g = f^n$ and n is a multiple of every n_i, then the basic sets of g are the $X_{j,i}$'s and $g|_{X_{j,i}}$ is mixing. We will at times restrict our attention to mixing basic sets and recover the general case by considering f^n.

A sequence $\underline{x} = \{x_i\}_{i=a}^b$ ($a = -\infty$ or $b = +\infty$ is permitted) 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 $x \in M$ *β-shadows* $\underline{x}$ if

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

3.6. Proposition. *For every $\beta > 0$ there is an $\alpha > 0$ so 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. Let $\varepsilon > 0$ be a small number to be determined later and choose $\delta \in (0, \varepsilon)$ as in 3.3, *i.e.*, $W^s_\varepsilon(x) \cap W^u_\varepsilon(y) \cap \Omega \ne \emptyset$ whenever $x, y \in \Omega$ with $d(x, y) \le \delta$. Pick M so large that $\lambda^M \varepsilon < \delta/2$ and then $\alpha > 0$ so that:

$$\begin{aligned}&\text{if } \{y_i\}_{i=0}^M \text{ is an } \alpha-\text{pseudo-orbit in } \Omega, \text{ then}\\ &d(f^j y_0, y_j) < \delta/2 \;\text{ for all }\; j \in [0, M]\,.\end{aligned}$$

Consider first an α-pseudo-orbit $\{x_i\}_{i=0}^{rM}$ with $r > 0$. Define x'_{kM} recursively for $k \in [0, r]$ by $x'_0 = x_0$ and

$$x'_{(k+1)M} = W^u_\varepsilon(f^M x'_{kM}) \cap W^s_\varepsilon(x_{(k+1)M}) \in \Omega\,.$$

This makes sense: $d(f^M x'_{kM}, f^M x_{kM}) \le \lambda^M \varepsilon < \delta/2$ and $d(f^M x_{kM}, x_{(k+1)M}) < \delta/2$ by the choice of α; so $d(f^M x'_{kM}, x_{(k+1)M}) < \delta$ and $x'_{(k+1)M}$ exists. Now set $x = f^{-rM} x'_{rM}$. For $i \in [0, rM]$ pick s with $i \in [sM, (s+1)M)$, then

$$\begin{aligned}d(f^i x, f^{i-sM} x'_{sM}) &\le \sum_{t=s+1}^{r} d(f^{i-tM} x'_{tM}, f^{i-tM+M} x_{(t-1)M})\\ &\le \sum_{t=s+1}^{r} \varepsilon\; \lambda^{tM-i} \le \frac{\varepsilon\lambda}{1-\lambda}\end{aligned}$$

where we used $x'_{tM} \in W^u_\varepsilon(f^M x'_{(t-1)M})$. Since $x'_{sM} \in W^s_\varepsilon(x_{sM})$

$$d(f^{i-sM} x'_{sM}, f^{i-sM} x_{sM}) \le \varepsilon\,.$$

By the choice of α one has

$$d(f^{i-sM} x_{sM}, x_i) < \delta/2\,.$$

By the triangle inequality

$$d(f^i x, x_i) \le \frac{\varepsilon\lambda}{1-\lambda} + \varepsilon + \frac{\delta}{2}\cdot$$

For small ε this is less that the given β.

Now any α-pseudo-orbit $\{x_i\}_{i=0}^n$ in Ω extends to $\{x_i\}_{i=0}^{rM}$ when $rM \geq n$ by setting $x_i = f^{i-n}x_n$ for $i \in (n, rM]$. An $x \in \Omega$ shadowing this extended pseudo-orbit will shadow the original one. If $\{x_i\}_{i=a}^b$ is a finite α-pseudo orbit, then so is $\{x_{j+a}\}_{j=0}^{b-a}$ and x shadowing this one yields $f^{-a}x$ shadowing the original. Thus the proposition holds for finite pseudo-orbits. If $\{x_i\}_{i=-\infty}^{+\infty}$ is an α-pseudo-orbit in Ω, then find $x^{(m)} \in \Omega$ β-shadowing $\{x_i\}_{i=-m}^m$ and let x be a limit point of the sequence $x^{(m)}$. Then $x \in \Omega$ β-shadows $\{x_i\}_{i=-\infty}^{+\infty}$. □

3.7. Corollary. *Given any $\beta > 0$ there is an $\alpha > 0$ so that the following holds: if $x \in \Omega$ and $d(f^n x, x) < \alpha$, then there is an $x' \in \Omega$ with $f^n x' = x'$ and*

$$d(f^k x, f^k x') \leq \beta \quad \text{for all} \quad k \in [0, n]\,.$$

Proof. Let $x_i = f^k x$ for $i \equiv k \pmod{n}$, $k \in [0, n)$. Then $\{x_i\}_{i=-\infty}^{\infty}$ is an α-pseudo-orbit. Let $x' \in \Omega$ β-shadow it. Then $d(f^i x', f^i f^n x') \leq d(f^i x', x_i) + d(x_i, f^{i+n}x') \leq 2\beta$ and by expansiveness (Lemma 3.4) $f^n x' = x'$. □

3.8. Anosov's Closing Lemma. *If f is an Anosov diffeomorphism, then f satisfies Axiom A.*

Proof. We must show that the periodic points are dense in $\Omega(f)$. We have been assuming f satisfies Axiom A; however 3.3 is true also for Anosov diffeomorphisms and so then is 3.6 and 3.7, using M in place of $\Omega(f)$. If y is a non-wandering point for an Anosov f, then for any γ one can find x with $d(x, y) < \gamma$ and $d(f^n x, x) < \gamma$ for some n. The periodic points x' constructed in 3.7 for such x converge to y. □

3.9. Fundamental Neighborhood. *Let f satisfy Axiom A. There is a neighborhood U of $\Omega(f)$ so that*

$$\bigcap_{n\in\mathbb{Z}} f^n U = \Omega(f)\,.$$

Proof. Let β be small and α as in 3.6. Pick $\gamma < \alpha/2$ so that

$$\forall x, y \in M\,,\ d(x, y) < \gamma \quad \text{implies} \quad d(fx, fy) < \alpha/2\,.$$

Let $U = \{y \in M : d(y, \Omega) < \gamma\}$. If $y \in \bigcap_{n\in\mathbb{Z}} f^n U$, pick $x_i \in \Omega$ with $d(f^i y, x_i) < \gamma$. Then

$$d(f^i y, f^i x) < \beta + \gamma \quad \text{for all } i\,.$$

For small β and γ this implies $y = x \in \Omega$ by 3.4. □

For Ω_j a basic set of an Axiom A diffeomorphism one let

$$\begin{aligned} & W^s(\Omega_j) = \{x \in M : d(f^n x, \Omega_j) \to 0 \text{ as } n \to \infty\} \\ \text{and} \quad & W^u(\Omega_j) = \{x \in M : d(f^{-n} x, \Omega_j) \to 0 \text{ as } n \to \infty\}\,. \end{aligned}$$

Using the definition of non-wandering sets it is easy to check that $f^n x \to \Omega$ and $f^{-n}x \to \Omega$ as $n \to \infty$. As $\Omega = \Omega_1 \cup \cdots \cup \Omega_s$ is a disjoint union of closed invariant sets one then sees that

$$M = \bigcup_{j=1}^{s} W^s(\Omega_j) = \bigcup_{j=1}^{s} W^u(\Omega_j)$$

and that there are disjoint unions.

3.10. Proposition. *$W^s(\Omega_j) = \bigcup_{x\in\Omega_j} W^s(x)$ and $W^u(\Omega_j) = \bigcup_{x\in\Omega_j} W^u(x)$. For $\varepsilon > 0$ there is a neighborhood U_j of Ω_j so that*

$$\bigcap_{k\geq 0} f^{-k}U_j \subset W^s_\varepsilon(\Omega_j) = \bigcup_{x\in\Omega_j} W^s_\varepsilon(x)$$

$$\text{and} \qquad \bigcap_{k\geq 0} f^{k}U_j \subset W^u_\varepsilon(\Omega_j) = \bigcup_{x\in\Omega_j} W^u_\varepsilon(x)\,.$$

Proof. Suppose $f^n y \to \Omega_j$ as $n \to \infty$; say $d(f^n y, \Omega_j) < \gamma$ for all $n \geq N$. Pick $x_n \in \Omega_j$ for $n \geq N$ with $d(x_n, f^n y) \leq \gamma$; for $n < N$ let $x_n = f^{n-N}x_N$. The $\{x_n\}_{n=-\infty}^{\infty}$ is a pseudo-orbit in Ω_j. Letting $x \in \Omega_j$ shadow it, one gets

$$f^N y \in W^s_\varepsilon(f^N x) \subset W^s(f^N x)$$

(provided γ was small enough). Then $y \in f^{-N}W^s(f^N x) = W^s(x)$. The reverse inclusion, $W^s(\Omega_j) \supset \bigcup_{x\in\Omega_j} W^s(x)$, is clear.
The proof for $W^u(\Omega_j)$ is similar and we have proved the second statement with $U_j = \{y \in M : d(y, \Omega_j) < \gamma\}$. $\square$

C. Markov partitions

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

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

R is called *proper* if R is closed and $R = \overline{\operatorname{int}(R)}$ ($\operatorname{int}(R)$ is the interior of R as a subset of Ω_s). For $x \in R$, let

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

where ε is small and the diameter of R is small compared to ε.

3.11. Lemma. *Let R be a closed rectangle. As a subset of Ω_s, R has boundary*

$$\partial R = \partial^s R \cup \partial^u R$$

where

$$\partial^s R = \{x \in R : x \notin \operatorname{int}(W^u(x,R))\}$$
$$\partial^u R = \{x \in R : x \notin \operatorname{int}(W^s(x,R))\}$$

and the interiors of $W^u(x,R)$, $W^s(x,R)$ are as subsets of $W^u_\varepsilon(x)\cap\Omega$, $W^s_\varepsilon(x)\cap\Omega$.

Proof. If $x \in \operatorname{int}(R)$, then $W^u(x,R) = R\cap(W^u_\varepsilon(x)\cap\Omega)$ is a neighborhood of x in $W^u_\varepsilon(x)\cap\Omega$ since R is a neighborhood of x in Ω. Similarly $x \in \operatorname{int}(W^u(x,R))$. Suppose $x \in \operatorname{int}(W^u(x,R))$ and $x \in \operatorname{int}(W^s(x,R))$. For $y \in \Omega_s$ near x the points

$$[x,y] \in W^s_\varepsilon(x)\cap\Omega \quad\text{and}\quad [x,y] \in W^u_\varepsilon(x)\cap\Omega$$

depend continuously on y. Hence for $y \in \Omega_s$ close enough to x, $[x,y] \in R$ and $[y,x] \in R$. Then

$$y' = [[y,x],[x,y]] \in R \cap W^s_\varepsilon(y) \cap W^u_\varepsilon(y)$$

and $y' = y$ as $W^s_\varepsilon(y)\cap W^u_\varepsilon(y) = \{y\}$. Thus $x \in \operatorname{int}(R)$. □

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

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

3.12. Theorem. *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 very small and choose $\alpha > 0$ small as in Proposition 3.6, *i.e.*, every α-pseudo-orbit in Ω_s is β-shadowed in Ω_s. Choose $\gamma < \alpha/2$ so that

$$d(fx,fy) < \alpha/2 \quad\text{when}\quad d(x,y) < \gamma\,.$$

Let $P = \{p_1,\dots,p_r\}$ be a γ-dense subset of Ω_s and

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

For each $\underline{q} \in \Sigma(P)$ there is a unique $\theta(\underline{q}) \in \Omega_s$ which β-shadows $\underline{q}$; for each $x \in \Omega_s$ there are $\underline{q}$ with $x = \theta(\underline{q})$.

For $\underline{q},\underline{q}' \in \Sigma(P)$ with $q_0 = q'_0$ we define $\underline{q}^* = [\underline{q},\underline{q}'] \in \Sigma(P)$ by

$$q^*_j = \begin{cases} q_j & \text{for} \quad j \geq 0 \\ q'_j & \text{for} \quad j \leq 0\,. \end{cases}$$

Then $d(f^j\theta(\underline{q}^*), f^j\theta(\underline{q})) \leq 2\beta$ for $j \geq 0$ and $d(f^j\theta(\underline{q}^*), f^j\theta(\underline{q})) \leq 2\beta$ for $j \leq 0$. So $\theta(\underline{q}^*) \in W^s_{2\beta}(\theta(\underline{q}))\cap W^u_{2\beta}(\theta(\underline{q}'))$, *i.e.*,

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

We now see that $T_s = \{\theta(\underline{q}) : \underline{q} \in \Sigma(P), q_0 = p_s\}$ is a rectangle. For $x, y \in T_s$ we write $x = \theta(\underline{q})$, $y = \theta(\underline{q}')$ with $q_0 = p_s = q_0'$. Then

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

Suppose $x = \theta(\underline{q})$ with $q_0 = p_s$ and $q_1 = p_t$. Consider $y \in W^s(x, T_s)$, $y = \theta(\underline{q}')$, $q_0 = p_s$. Then

$$\begin{aligned} y &= [x, y] = \theta[\underline{q}, \underline{q}'] \quad \text{and} \\ fy &= \theta(\sigma[\underline{q}, \underline{q}']) \in T_t \end{aligned}$$

as $\sigma[\underline{q}, \underline{q}']$ has $\underline{q}' = p_t$ in its zeroth position. Since $fy \in W^s_\varepsilon(fx)$ ($\operatorname{diam}(T_s) \le 2\beta$ is small compared to ε), $fy \in W^s_\varepsilon(fx, T_t)$. We have proved
(i) $fW^s(x, T_s) \subset W^s(fx, T_t)$.
A similar proof shows $f^{-1}W^u(fx, T_t) \subset W^u(x, T_s)$, *i.e.*,
(ii) $fW^u(x, T_s) \supset W^u(fx, T_t)$.
Each T_s is closed; this follows from the following lemma.

3.13. Lemma. *$\theta : \Sigma(P) \to \Omega_s$ is continuous.*

Proof. Otherwise there is a $\gamma > 0$ so that for every N one can find $\underline{q}_N, \underline{q}'_N \in \Sigma(P)$ with $q_{j,N} = q'_{j,N}$ for all $j \in [-N, N]$ but $d(\theta(\underline{q}_N), \theta(\underline{q}'_N)) \ge \gamma$. If $x_N = \theta(\underline{q}_N)$, $y_N = \theta(\underline{q}')$ one has

$$d(f^j x_N, f^j y_N) \le 2\beta \quad \forall j \ \in \ [-N, N] \,.$$

Taking subsequences we may assume $x_N \to x$ and $y_N \to y$ as $N \to \infty$. Then $d(f^j x, f^j y) \le 2\beta$ for all j and $d(x, y) \ge \gamma$; this contradicts expansiveness of $f|_{\Omega_s}$. $\square$

Now $\mathcal{T} = \{T_1, \ldots, T_r\}$ is a covering by rectangles and (i) and (ii) above are like the Markov condition (b). However the T_j's are likely to overlap and not be proper. For each $x \in \Omega_s$ let

$$\mathcal{T}(x) = \{T_j \in \mathcal{T} : x \in T_j\} \text{ and } \mathcal{T}^*(x) = \{T_k \in \mathcal{T} : T_k \cap T_j \ne \emptyset \text{ for some } T_j \in \mathcal{T}(x)\} \,.$$

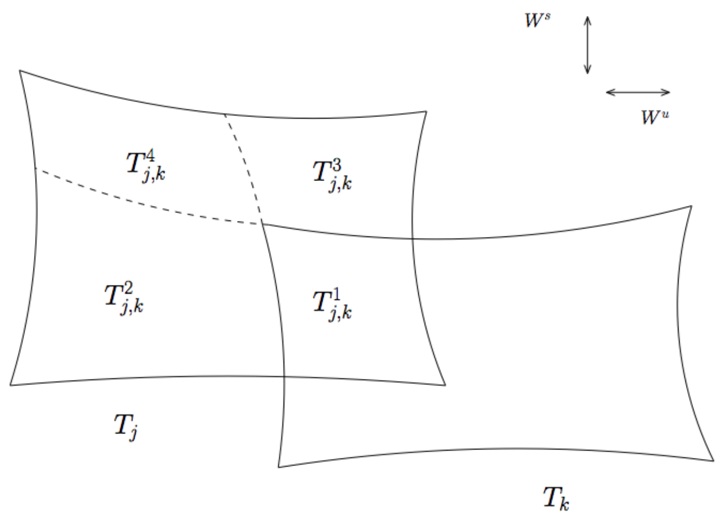
As $\mathcal{T}$ is a closed cover of Ω_s, $Z = \Omega_s \backslash \bigcup_j \partial T_j$ is an open dense subset of Ω_s. In fact, using arguments similar to 3.11, one can show that

$$Z^* = \{x \in \Omega_s : W^s_\varepsilon(x) \cap \partial^s T_k = \emptyset \text{ and } W^u_\varepsilon(x) \cap \partial^u T_k = \emptyset \text{ for all } T_k \in \mathcal{T}^*(x)\}$$

is open and dense in Ω_s.

For $T_j \cap T_k \ne \emptyset$, let

$$\begin{aligned} T^1_{j,k} &= \{x \in T_j : W^u(x, T_j) \cap T_k \ne \emptyset,\ W^s(x, T_j) \cap T_k \ne \emptyset\} = T_j \cap T_k \\ T^2_{j,k} &= \{x \in T_j : W^u(x, T_j) \cap T_k \ne \emptyset,\ W^s(x, T_j) \cap T_k = \emptyset\} \\ T^3_{j,k} &= \{x \in T_j : W^u(x, T_j) \cap T_k = \emptyset,\ W^s(x, T_j) \cap T_k \ne \emptyset\} \\ T^4_{j,k} &= \{x \in T_j : W^u(x, T_j) \cap T_k = \emptyset,\ W^s(x, T_j) \cap T_k = \emptyset\}. \end{aligned}$$



If $x, y \in T_j$, then $W^s([x,y],T_j) = W^s(x,T_j)$ and $W^u([x,y],T_j) = W^u(y,T_j)$; this implies $T^n_{j,k}$ is a rectangle open in Ω_s and each $x \in T_j \cap Z^*$ lies in $\operatorname{int}(T^n_{j,k})$ for some n. For $x \in Z^*$ define

$$R(x) = \bigcap \left\{\operatorname{int}(T^n_{j,k}) : x \in T_j, T_k \cap T_j \neq \emptyset \text{ and } x \in T^n_{j,k}\right\}.$$

Now $R(x)$ is an open rectangle ($x \in Z^*$). Suppose $y \in R(x) \cap Z^*$. Since $R(x) \subset \mathcal{T}(x)$ and $R(x) \cap T_j = \emptyset$ for $T_j \notin \mathcal{T}(x)$, one gets $\mathcal{T}(y) = \mathcal{T}(x)$. For $T_j \in \mathcal{T}(x) = \mathcal{T}(y)$ and $T_k \cap T_j \neq \emptyset$, y lies in the same $T^n_{j,k}$ as x does since $T^n_{j,k} \supset R(x)$; hence $R(y) = R(x)$. If $R(x) \cap R(x') \neq \emptyset$ ($x, x' \in Z^*$), there is a $y \in R(x) \cap R(x') \cap Z^*$; then $R(x) = R(y) = R(x')$. As there are only finitely many $T^n_{j,k}$'s there are only finitely many distinct $R(x)$'s. Let

$$\mathcal{R} = \{\overline{R(x)} : x \in Z^*\} = \{R_1, \ldots, R_m\}.$$

For $x' \in Z^*$, $R(x') = R(x)$ or $R(x') \cap R(x) = \emptyset$; hence $(\overline{R(x)} \setminus R(x)) \cap Z^* = \emptyset$. As Z^* is dense in Ω_s, $\overline{R(x)} \setminus R(x)$ has no interior (in Ω_s) and $R(x) = \operatorname{int}(\overline{R(x)})$. For $R(x) \neq R(x')$

$$\operatorname{int}\left(\overline{R(x)}\right) \cap \operatorname{int}\left(\overline{R(x')}\right) = R(x) \cap R(x') = \emptyset.$$

To show that $\mathcal{R}$ is Markov we are left to verify condition (b).

Suppose $x, y \in Z^* \cap f^{-1}Z^*$, $R(x) = R(y)$ and $y \in W^s_\varepsilon(x)$. We will show $R(fx) = R(fy)$. First $\mathcal{T}(fx) = \mathcal{T}(fy)$. Otherwise assume $fx \in T_j$, $fy \notin T_j$. Let $fx = \theta(\sigma \underline{q})$ with $q_1 = p_j$ and $q_0 = p_s$. Then $x = \theta(\underline{q}) \in T_s$ and by inclusion (i) above

$$fy \in fW^s(x, T_s) \subset W^s(fx, T_j)\,,$$

contradicting $fy \notin T_j$. Now let $fx, fy \in T_j$ and $T_k \cap T_j \neq \emptyset$. We want to show that fx, fy belong to the same $T^n_{j,k}$. As $fy \in W^s_\varepsilon(fx)$ we have $W^s(fy, T_j) = W^s(fx, T_j)$. We will derive a contradiction from

$$W^u(fy, T_j) \cap T_k = \emptyset\,,\ fz \in W^u(fx, T_j) \cap T_k\,.$$

Recall that $fx = \theta(\sigma\underline{q})$, $q_1 = p_j$, $q_0 = p_s$. Then by inclusion (ii)

$$fz \in W^u(fx, T_j) \subset fW^u(x, T_s) \quad \text{or} \quad z \in W^u(x, T_s)\,.$$

Let $fz = \theta(\sigma\underline{q}')$; $q'_1 = p_k$ and $q_0 = p_t$. Then $z \in T_t$ and $fW^s(z, T_t) \subset W^s(fz, T_k)$. Now $T_s \in \mathfrak{T}(x) = \mathfrak{T}(y)$ and $z \in T_t \cap T_s \neq \emptyset$.
Now $z \in W^u(x, T_s) \cap T_t$ and so there is some $z' \in W^u(y, T_s) \cap T_t$ as x, y are in the same $T^n_{s,t}$. Then

$$z'' = [z, y] = [z, z'] \in W^s(z, T_t) \cap W^u(y, T_s),$$

and $fz'' = [fz, fy] \in W^s(fz, T_k) \cap W^u(fy, T_j)$ (using $fz, fy \in T_j$ a rectangle), a contradiction. So $R(fx) = R(fy)$.

For small $\delta > 0$ the sets

$$Y_1 = \bigcup \left\{ W^s_\delta(z) : z \in \bigcup_j \partial^s T_j \right\} \quad \text{and} \quad Y_2 = \bigcup \left\{ W^u_\delta(z) : z \in \bigcup_j \partial^u T_j \right\}$$

are closed and nowhere dense (like in the proof of 3.11). Now $Z^* \supset \Omega_s \backslash (Y_1 \cup Y_2)$ is open and dense. Furthermore if $x \notin (Y_1 \cup Y_2) \cap f^{-1}(Y_1 \cup Y_2)$ then $x \in Z^* \cap f^{-1}Z^*$ and the set of $y \in W^s(x, R(x))$ with $y \in Z^* \cap f^{-1}Z^*$ is open and dense in $W^s(x, \overline{R(x)}\,)$ (as a subset of $W^s_\varepsilon(x) \cap \Omega$). By the previous paragraph $R(fy) = R(fx)$ for such y; by continuity

$$fW^s(x, \overline{R(x)}\,) \subset \overline{R(fx)}\,.$$

As $fW^s(x, \overline{R(x)}\,) \subset W^s_\varepsilon(fx)$, $fW^s(x, \overline{R(x)}\,) \subset W^s(fx, \overline{R(fx)}\,)$.
If $\mathrm{int}(R_i) \cap f^{-1}\mathrm{int}(R_j) \neq \emptyset$, then this open subset of Ω_s contains some x satisfying the above conditions, $R_i = \overline{R(x)}$ and $R_j = \overline{R(fx)}$. For any $x' \in R_i \cap f^{-1}R_j$ one has $W^s(x', R_i) = \{[x', y] : y \in W^s(x, R_i)\}$ and

$$\begin{aligned} fW^s(x', R_i) &= \{[fx', fy] : y \in W^s(x, R_i)\} \\ &\subset \{[fx', z] : z \in W^s(fx, R_j)\} \\ &\subset W^s(fx', R_j)\,. \end{aligned}$$

This completes the proof of half of the Markov conditions (b). The other half is proved similarly and the proof is omitted. Alternatively one could apply the above to f^{-1}, noting that $W^u_f = W^s_{f^{-1}}$. $\square$

D. Symbolic dynamics

Throughout this section $\mathcal{R} = \{R_1, \ldots, R_m\}$ will denote a Markov partition of a basic set Ω_s. One defines the *transition matrix* $A = A(\mathcal{R})$ by

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

3.14. Lemma. *Suppose $x \in R_i$, $fx \in R_j$, $A_{ij} = 1$. Then $fW^s(x, R_i) \subset W^s(fx, R_j)$ and $fW^u(x, R_i) \supset W^u(fx, R_j)$.*

Proof. This is just the same as the last part of the proof of 3.12. $\square$

Definition. $\partial^s\mathcal{R} = \bigcup_j \partial^s R_j$ *and* $\partial^u\mathcal{R} = \bigcup_j \partial^u R_j$.

3.15. Proposition. $f(\partial^s\mathcal{R}) \subset \partial^s\mathcal{R}$ *and* $f^{-1}(\partial^u\mathcal{R}) \subset \partial^u\mathcal{R}$.

Proof. The set $\bigcup_j(\operatorname{int}(R_i) \cap f^{-1}\operatorname{int}(R_j))$ is dense in R_i. For any $x \in R_i$ one can therefore find some j and $x_n \in \operatorname{int}(R_i) \cap f^{-1}\operatorname{int}(R_j)$ with $\lim_{n\to\infty} x_n = x$. Then $A_{ij} = 1$, $x \in R_{ij}$ and $fx \in R_j$. Hence $fW^u(x, R_i) \supset W^u(fx, R_j)$. If $fx \notin \partial^s\mathcal{R}$, then $W^u(fx, R_j)$ is a neighborhood of fx in $W^s_\varepsilon(fx) \cap \Omega$ and so $W^u(x, R_i)$ is a neighborhood of x in $W^s_\varepsilon(x) \cap \Omega_s$ – that is $x \notin \partial^s R_i$. We have shown $f(\partial^s\mathcal{R}) \subset \partial^s\mathcal{R}$. One gets $f^{-1}(\partial^u\mathcal{R}) \subset \partial^u\mathcal{R}$ by a similar argument or by applying the first argument to f^{-1} in place of f. $\square$

3.16. Lemma. *Let $D \subset W^s_\delta(x) \cap \Omega$ and $C \subset W^u_\delta(x) \cap \Omega$. Then the rectangle $[C, D]$ is proper iff $D = \overline{\operatorname{int}(D)}$ and $C = \overline{\operatorname{int}(C)}$ as subsets of $W^s_\delta(x) \cap \Omega$ and $W^u_\delta(x) \cap \Omega$ respectively.*

Proof. This is like 3.11. $\square$

Definition. *Let R, S be two rectangles. S will be called a* u-subrectangle *of R if*

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

3.17. Lemma. *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.*

Proof. Pick $x \in R_i \cap f^{-1}R_j$ and set $D = W^s(x, R_i) \cap S$. Because S is a u-subrectangle (condition (b)) one has

$$S = \bigcup_{y \in D} W^u(y, R_i) = [W^u(x, R_i), D]\,.$$

As S is proper and nonempty, by 3.16 $D \neq \emptyset$ and $D = \overline{\operatorname{int}(D)}$. Now

$$f(S) \cap R_j = \bigcup_{y \in D} (fW^u(y, R_i) \cap R_j) .$$

By 3.14, $f(y) \in R_j$ and $fW^u(y, R_i) \cap R_j = W^u(fy, R_j)$.
So $f(S) \cap R_j = \bigcup_{y' \in f(D)} W^u(y', R_j) = [W^u(fx, R_j), f(D)]$.
Since $R_j = [W^u(fx, R_j), W^s(fx, R_j)]$ is proper, one has $W^u(fx, R_j)$ proper. As f maps $W^s_\varepsilon(x) \cap \Omega$ homeomorphically onto a neighborhood in $W^s_\varepsilon(fx) \cap \Omega$, $f(D) = \overline{\operatorname{int}(f(D))}$ and so $f(S) \cap R_j$ is proper by 3.16. $f(S) \cap R_j \neq \emptyset$ as $f(D) \neq \emptyset$; if $y'' \in f(S) \cap R_j$, then $y'' \in W^u(y', R_j)$ for some $y' \in f(D)$ and $W^u(y'', R_j) = W^u(y', R_j) \subset f(S) \cap R_j$. So $f(S) \cap R_j$ is a u-subrectangle of R_j. $\square$

3.18. Theorem. *For each $\underline{a} \in \Sigma_A$ the set $\bigcap_{j \in \mathbb{Z}} f^{-j} R_{a_j}$ 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 $Y = \Omega_s \setminus \bigcup_{j \in \mathbb{Z}} f^j (\partial^s \mathcal{R} \cup \partial^u \mathcal{R})$.*

Proof. If $a_1 a_2 \cdots a_n$ is a word with $A_{a_j a_{j+1}} = 1$, then inductively using 3.17 one sees 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}. From this one gets that

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

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

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

If $x, y \in K(\underline{a})$, then $f^j x, f^j y \in R_{a_j}$ are close for all $j \in \mathbb{Z}$ and so $x = y$ by expansiveness. As

$$\begin{aligned} K(\sigma \underline{a}) &= \bigcap_j f^{-j} R_{a_{j+1}} = f \left(\bigcap_j f^{-j} R_{a_j} \right) \\ &= f K(\underline{a}), \end{aligned}$$

one has $\pi \circ \sigma = f \circ \pi$. That π is continuous is proved like 3.13. As $\partial^s \mathcal{R} \cup \partial^u \mathcal{R}$ is nowhere dense, Y is residual. For $x \in Y$ pick a_j with $f^j x \in R_{a_j}$. As $x \in Y$, $f^j x \in \operatorname{int}(R_{a_j})$ and so $A_{a_j a_{j+1}} = 1$. Thus $\underline{a} = \{a_j\} \in \Sigma_A$ and $x = \pi(\underline{a})$. If $x = \pi(\underline{b})$, then $f^j x \in R_{b_j}$ and $b_j = a_j$ because $f^j x \notin \partial^s \mathcal{R} \cup \partial^u \mathcal{R}$; so π is injective over Y. As $\pi(\Sigma_A)$ is a compact subset of Ω_s containing a dense set Y, $\pi(\Sigma_A) = \Omega_s$. $\square$

3.19. Proposition. *$\sigma : \Sigma_A \to \Sigma_A$ is topologically transitive. If $f|_{\Omega_s}$ is topologically mixing so is $\sigma : \Sigma_A \to \Sigma_A$.*

Proof. Let U, V be nonempty open in Σ_A. For some $\underline{a}, \underline{b} \in \Sigma_A$ and N one has

$$\begin{aligned} U \supset U_1 &= \{\underline{x} \in \Sigma_A : x_i = a_i \ \forall\ i \in [-N, N]\} \\ V \supset V_1 &= \{\underline{x} \in \Sigma_A : x_i = b_i \ \forall\ i \in [-N, N]\} \,. \end{aligned}$$

Now

$$\emptyset \neq \operatorname{int}(K_N(\underline{a})) = \bigcap_{j=-N}^{N} f^{-j} \operatorname{int}(R_{a_j}) = U_2$$

$$\text{and} \qquad \emptyset \neq \operatorname{int}(K_N(\underline{b})) = \bigcap_{j=-N}^{N} f^{-j} \operatorname{int}(R_{b_j}) = V_2 \,.$$

Also, if $x = \pi(\underline{x}) \in U_2$, then $f^j x \in R_{x_j}$ and $f^j x \in \operatorname{int}(R_{a_j})$ imply $x_j = a_j$; so $\pi^{-1}(U_2) \subset U_1$. Similarly $\pi^{-1}(V_2) \subset V_1$. Since $f|_{\Omega_s}$ is transitive, $f^n U_2 \cap V_2 \neq \emptyset$ for various large n. Then

$$\begin{aligned} \emptyset \neq \pi^{-1}(f^n U_2 \cap V_2) &= \pi^{-1}(f^n U_2) \cap \pi^{-1}(V_2) \\ &\subset f^n U \cap V \,. \end{aligned}$$

This same argument shows that $\sigma|_{\Sigma_A}$ is mixing if $f|_{\Omega_s}$ is. $\square$

References

The basic references are Anosov [2] for Anosov diffeomorphisms and Smale [14] for Axiom A diffeomorphisms. The stable manifold theorem for hyperbolic sets is due to Hirsch and Pugh [8]. Canonical coordinates and spectral decomposition are from [14], with the mixing part of 3.5 from [4].
The idea of pseudo-orbit has probably occurred to many people. Proposition 3.6 is explicitly proved in [6], though earlier similar statements are in [4] and for Anosov diffeomorphisms in [2]. Sinai [12] stated 3.6 explicitly for Anosov diffeomorphisms. Corollary 3.7 is in [2] for Anosov diffeomorphisms and [3] for Axiom A diffeomorphisms. As the name implies, 3.8 is due to Anosov. Results 3.9 and 3.10 are from [7], [15] and for their proofs we have followed [6].
Symbolic dynamics for certain geodesic flows goes back to Hadamard and was developed by Morse [9]. It was carried out for the horseshoe by Smale [13] and for automorphisms of the torus by Adler and Weiss [1]. Sinai [10], [11] did Sections 3 and 3 for Anosov diffeomorphisms and this was carried over to Axiom A diffeomorphisms in [5].

1. R. Adler, B. Weiss: *Similarity of automorphisms of the torus*, Memoirs of the American Mathematical Society **98**, American Mathematical Society, Providence, R.I. 1970.
2. D. Anosov: Geodesic flows on closed Riemann manifolds with negative curvature, *Proc. Steklov Inst. Math.* **90** (1967).
3. R. Bowen: Topological entropy and Axiom A, *Proc. Sympos. Pure Math., Vol. XIV, Berkeley, Calif., 1968*, pp. 23–41, Amer. Math. Soc., Providence, R.I., 1970.
4. R. Bowen: Periodic points and measures for Axiom A diffeomorphisms, *Trans. Amer. Math. Soc.* **154** (1971), 377-397.
5. R. Bowen: Markov partitions for Axiom A diffeomorphisms, *Amer. J. Math.* **92** (1970), 725-747.
6. R. Bowen: ω-limit sets for Axiom A diffeomorphisms, *J. Differential Equations* **18** (1975), no. 2, 333–339.
7. M. Hirsch, J. Palis, C. Pugh and M. Shub: Neighborhoods of hyperbolic sets, *Invent. Math.* **9** 1969/1970 121–134.
8. M. Hirsch and C. Pugh: Stable manifolds and hyperbolic sets, *Proc. Sympos. Pure Math., Vol. XIV, Berkeley, Calif., 1968*, pp. 133–163, Amer. Math. Soc., Providence, R.I., 1970.

9. M. Morse: A One-to-One Representation of Geodesics on a Surface of Negative Curvature, *Amer. J. Math.* **43** (1921), no. 1, 33–51.
10. Ya. Sinai: Markov partitions and C-diffeomorphisms, *Func. Anal. and its Appl.* **2** (1968), no. 1, 64-89.
11. Ya. Sinai: Construction of Markov partitions, *Func. Anal. and its Appl.* **2** (1968), no. 2, 70-80.
12. Ya. Sinai: Gibbs measures in ergodic theory, *Uspehi Mat. Nauk* **27** (1972), no. 4, 21–64. English translation: *Russian Math. Surveys* **27** (1972), no. 4, 21–69.
13. S. Smale: Diffeomorphisms with many periodic points, in "Differential and Combinatorial Topology". A Symp. in Honor of Marston Morse, pp. 63-80, Princeton Univ. Press, Princeton, N.J., 1965.
14. S. Smale: Differentiable dynamical systems, *Bull. Amer. Math. Soc.* **73** (1967), 747-817.
15. S. Smale: The Ω-stability theorem, *Proc. Sympos. Pure Math., Vol. XIV, Berkeley, Calif., 1968*, pp. 289–298, Amer. Math. Soc., Providence, R.I.

4

ERGODIC THEORY OF AXIOM A DIFFEOMORPHISMS

A. Equilibrium states for basic sets

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

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

4.1. Theorem. *Let Ω_s be a basic set for an Axiom A diffeomorphism f and $\phi : \Omega_s \to \mathbb{R}$ Hölder continuous. Then ϕ has a unique equilibrium state μ_ϕ (w.r.t. $f|_{\Omega_s}$). Furthermore μ_ϕ is ergodic; μ_ϕ is Bernoulli if $f|_{\Omega_s}$ is topologically mixing.*

4.2. Lemma. *There are $\varepsilon > 0$ and $\alpha \in (0, 1)$ for which the following are true: if $x \in \Omega_s$, $y \in M$, and $d(f^k x, f^k y) \leq \varepsilon$ for all $k \in [-N, N]$, then $d(x, y) < \alpha^N$.*

Proof. See p. 140 of [12]. □

Proof of 4.1. Let $\mathcal{R}$ be a Markov partition for Ω_s of diameter at most ε, A the transition matrix for $\mathcal{R}$ and $\pi : \Sigma_A \to \Omega_s$ as in 3.D. Let $\phi^* = \phi \circ \pi$. If $\underline{x}, \underline{y} \in \Sigma_A$ have $x_k = y_k$ for $k \in [-N, N]$, then

$$f^k\pi(\underline{x})\ ,\ f^k\pi(\underline{y}) \in R_{x_k} = R_{y_k} \text{ for } k \in\ [-N, N]\,.$$

This gives $d(\pi(\underline{x}), \pi(\underline{y})) < \alpha^N$, $|\phi^*(\underline{x}) - \phi^*(\underline{y})| \leq a\ (\alpha^\theta)^N$ and $\phi^* \in \mathscr{F}_A$.

First we assume $f|_{\Omega_s}$ is mixing. Then $\sigma|_{\Sigma_A}$ is mixing by 3.19 and we have a Gibbs measure μ_{ϕ^*} as in Chapter 1. Let $D_s = \pi^{-1}(\partial^s\mathcal{R})$ and $D_u = \pi^{-1}(\partial^u\mathcal{R})$. Then D_s and D_u are closed subsets of Σ_A, each smaller than Σ_A, and $\sigma D_s \subset D_s$, $\sigma^{-1}D_u \subset D_u$. As μ_{ϕ^*} is σ-invariant, $\mu_{\phi^*}(\sigma^n D_s) = \mu_{\phi^*}(D_s)$; using $\sigma^{n+1}D_s \subset \sigma^n D_s$ one has

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

Now $\bigcap_{n\geq 0}\sigma^n D_s$ has measure 0 or 1 as it is σ-invariant and μ_{ϕ^*} is ergodic (see 1.14); since its complement (a nonempty open set) has positive measure by 1.4, one gets $\mu_{\phi^*}(D_s)=0$. Similarly one sees $\mu_{\phi^*}(D_u)=0$. Now let $\mu_\phi=\pi^*\mu_{\phi^*}$, *i.e.*, $\mu_\phi(E)=\mu_{\phi^*}(\pi^{-1}E)$. Then μ_ϕ is f-invariant and the automorphisms of the measure spaces (σ,μ_{ϕ^*}), (f,μ_ϕ) are conjugate since π is one-to-one except on the null set $\bigcup_{n\in\mathbb{Z}}\sigma^n(D_s\cup D_u)$. In particular $h_{\mu_\phi}(f)=h_{\mu_{\phi^*}}(\sigma)$ and so [1]

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

Hence [2] $P_\sigma(\phi^*)=P_f(\phi)$ and μ_ϕ is an equilibrium state for ϕ by Chapter 2. Now μ_ϕ is Bernoulli because of 1.25.

4.3. Lemma. *For any $\mu\in\mathscr{M}_f(\Omega_s)$ there is a $\nu\in\mathscr{M}_\sigma(\Sigma_A)$ with $\pi^*\nu=\mu$.*

Proof. This well-known fact is proved as follows. $F(g\circ\pi)=\mu(g)=\int g d\mu$ defines a positive linear functional on a subspace of $\mathscr{C}(\Sigma_A)$. By a modification of the Hahn-Banach Theorem F extends to $\mathscr{C}(\Sigma_A)$, still positive. As $F(1)=F(1\circ\pi)=1$, F is identified with some $\beta\in\mathscr{M}(\Sigma_A)$. By compactness let $\nu=\lim_{k\to\infty}\frac{1}{n_k}(\beta+\sigma^*\beta+\cdots+(\sigma^{n_k-1})^*\beta)$. Then $\sigma^*\nu=\nu$ and $\pi^*\nu=\mu$ (using $\pi^*(\sigma^k)^*\beta=(f^k)^*\pi^*\beta=(f^k)^*\mu=\mu$). □

Proof of 4.1 (continued). Suppose μ is any equilibrium state of ϕ and pick $\nu\in\mathscr{M}_\sigma(\Sigma_A)$ with $\pi^*\nu=\mu$. Then $h_\nu(\sigma)\geq h_\mu(f)$ and so

$$h_\nu(\sigma)+\int\phi^*\ d\nu\geq h_\mu(f)+\int\phi\ d\mu=P(\phi)=P(\phi^*)\,.$$

Thus ν is an equilibrium state for ϕ^* and $\nu=\mu_{\phi^*}$ by 1.22. Then $\mu=\pi^*\mu_{\phi^*}=\mu_\phi$.

We have left the case $\Omega_s=X_1\cup\cdots\cup X_m$ with $fX_k=X_{k+1}$ and $f^m|_{X_1}$ mixing. For $\mu\in\mathscr{M}_f(\Omega_s)$, one has $\mu(X_1)=\frac{1}{m}$ and so $\mu'=m\ \mu|_{X_1}\in\mathscr{M}_{f^m}(X_1)$. Conversely, if $\mu'\in\mathscr{M}_{f^m}(X_1)$, then $\mu\in\mathscr{M}_f(\Omega_s)$ where

$$\mu(E)=\frac{1}{m}\sum_{k=0}^{m-1}\mu'(X_1\cap f^kE)\,.$$

One checks that $\mu\leftrightarrow\mu'$ defines a bijection $\mathscr{M}_f(\Omega_s)\leftrightarrow\mathscr{M}_{f^m}(X_1)$, $h_{\mu'}(f^m)=mh_\mu(f)$, and $\int S_m\phi\ d\mu'=m\int\phi\ d\mu$. Finding μ maximizing $h_\mu(f)+\int\phi\ d\mu$ is equivalent therefore to finding μ' maximizing $h_{\mu'}(f^m)+\int S_m\phi\ d\mu'$. For ϕ Hölder on Ω_s, $S_m\phi$ will be Hölder on X_1 and therefore one is done since X_1 is a mixing basic set of f^m. □

---

[1] The second equality follows from Theorem 1.22; The inequality comes from Proposition 2.13 (note of the editor).

[2] Using the variational principle 2.17 (note of the editor).

4.4. Proposition. *Let $\phi : \Omega_s \to \mathbb{R}$ be Hölder continuous and $P = P_{f|\Omega_s}(\phi)$. For small $\varepsilon > 0$ there is a $b_\varepsilon > 0$ so that, for any $x \in \Omega_s$ and for all n,*

$$\mu_\phi \left\{y \in \Omega_s : d(f^k y, f^k x) < \varepsilon \;\; \forall\, k \in [0,n]\right\} \geq b_\varepsilon \;\; \exp(-Pn + S_n\phi(x))\,.$$

Proof. Choose the Markov partition $\mathcal{R}$ above to have $\operatorname{diam}(\mathcal{R}) < \varepsilon$. Assume first $f|_{\Omega_s}$ is mixing. Pick $\underline{x} \in \Sigma_A$ with $\pi(\underline{x}) = x$. Then

$$\begin{aligned} B &= \left\{y \in \Omega_s : d(f^k y, f^k x) < \varepsilon \;\; \forall\, k \in [0,n]\right\} \\ &\supset \pi\left\{\underline{y} \in \Sigma_A : y_k = x_k \;\; \forall\, k \in [0,n]\right\}\,. \end{aligned}$$

Applying 1.4 and $P(\phi^*) = P(\phi) = P$ one gets

$$\mu_\phi(B) \geq c_1 \;\; \exp(-Pn + S_n\phi(x))\,.$$

We leave it to the reader to reduce the general case to the mixing one as in the proof of 4.1. □

4.5. Proposition. *Let $\phi, \psi : \Omega_s \to \mathbb{R}$ be two Hölder continuous functions. Then the following are equivalent:*

(i) $\mu_\phi = \mu_\psi$.
(ii) *There are constants K and L so that $|S_m\phi(x) - S_m\psi(x) - Km| \leq L$ for all $x \in \Omega_s$ and all $m \geq 0$.*
(iii) *There is a constant K so that $S_m\phi(x) - S_m\psi(x) = Km$ when $x \in \Omega_s$ with $f^m x = x$.*
(iv) *There is a Hölder function $u : \Omega_s \to \mathbb{R}$ and a constant K so that $\phi(x) - \psi(x) = K + u(fx) - u(x)$.*

If these conditions hold, $K = P(\phi) - P(\psi)$.

Proof. Let $\phi^* = \phi \circ \pi$ and $\psi^* = \psi \circ \pi$. We assume $f|_{\Omega_s}$ is mixing and leave the reduction to this case to the reader. If $\mu_\phi = \mu_\psi$, then $\mu_{\phi^*} = \mu_{\psi^*}$ and by Theorem 1.28 there are K and L so that

$$|S_m\phi^*(\underline{x}) - S_m\psi^*(\underline{x}) - Km| \leq L$$

for $\underline{x} \in \Sigma_A$. For $x \in \Omega_s$, picking $\underline{x} \in \pi^{-1}(x)$, this gives us (ii).

Assume (ii) and $f^m x = x$. Then

$$L \geq |S_{mj}\phi(x) - S_{mj}\psi(x) - mjK| = j|S_m\phi(x) - S_m\psi(x) - mK|\,.$$

Letting $j \to \infty$ we get (iii). If (iv) is true, then

$$\phi^*(\underline{x}) - \psi^*(\underline{x}) = K + u(\pi(\sigma\underline{x})) - u(\pi(\underline{x}))$$

and $\mu_{\phi^*} = \mu_{\psi^*}$ by Theorem 1.28. One then has $\mu_\phi = \pi^*\mu_{\phi^*} = \pi^*\mu_{\psi^*} = \mu_\psi$.

Now we assume (iii) and prove (iv). Let $\eta(x) = \phi(x) - \psi(x) - K$ and pick $x \in \Omega_s$ with dense forward orbit (Lemma 1.29). Let $A = \{f^k x : k \geq 0\}$ and define $u : A \to \mathbb{R}$ by

$$u(f^k x) = \sum_{j=0}^{k-1} \eta(f^j x)\,.$$

For $z \in A$ one has $u(fz) - u(z) = \eta(z)$. Pick ε and α as in 4.2. By 3.7 there is $\delta > 0$ so that if $y \in \Omega_s$ and $d(f^n y, y) < \delta$, then there is a $y' \in \Omega_s$ with $f^n y' = y'$ and $d(f^k y, f^k y') < \frac{\varepsilon}{2}$ for all $k \in [0, n]$.

Let R be the maximum ratio that f expands any distance. Suppose $y = f^k x$, $z = f^m x$ with $k < m$ and $d(y, z) < \epsilon / 2R^N$. Providing N is large one has $z = f^n y$, $n = m - k$, and $d(f^n y, y) < \delta$. Then find y' as above with $f^n y' = y'$. Then, as $S_n \eta(y') = 0$,

$$\begin{aligned} |u(z) - u(y)| &= \left| S_n \eta(y) - S_n \eta(y') \right| \\ &\le \sum_{j=0}^{n-1} \left| \eta(f^j y) - \eta(f^j y') \right|. \end{aligned}$$

By the choice of R and y' one sees that

$$d(f^j y, f^j y') < \varepsilon \quad \text{for all} \quad j \in [-N, n+N]\,.$$

For $j \in [0, n)$ Lemma 4.2 gives

$$d(f^j y, f^j y') < \alpha^{\min\{j+N, N+n-j\}}\,.$$

Because η is Hölder,

$$|\eta(f^j y) - \eta(f^j y')| \le a \ \alpha^{\theta \min\{j+N, N+n-j\}}$$

$$|u(y) - u(z)| \le 2a \sum_{r=N}^{\infty} \alpha^{\theta r} \le a' \alpha^{\theta N}.$$

Pick N so that $d(y, z) \in [\varepsilon / 2R^{N+1}, \varepsilon / 2R^N]$. Taking $\gamma > 0$ so that $(1/R)^\gamma \ge \alpha^\theta$ one has

$$|u(y) - u(z)| \le a'' d(y, z)^\gamma.$$

Thus u is Hölder on A and extends uniquely to a Hölder function on $\bar{A} = \Omega_s$. The formula $\eta(z) = u(fz) - u(z)$ extends to Ω_s by continuity. $\square$

B. The case $\phi = \phi^{(u)}$

Recall that M has a Riemannian structure and this induces a volume measure m on M. We will assume for the remainder of this chapter that $f : M \to M$ is a $\mathscr{C}^2$ Axiom A diffeomorphism and Ω_s is a 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

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

using inner products derived from the Riemannian metric.

4.6. Lemma. *If Ω_s is a $\mathscr{C}^2$ basic set, then $\phi^{(u)} : \Omega_s \to \mathbb{R}$ is Hölder continuous.*

Proof. The map $x \mapsto E^u_x$ is Hölder (see 6.4 of [12]) and $E^u_x \mapsto \phi^{(u)}(x)$ is differentiable, so the composition $x \mapsto \phi^{(u)}(x)$ is Hölder. □

By Theorem 4.1 the function $\phi^{(u)}$ has a unique equilibrium state which we denote $\mu^+ = \mu_{\phi^{(u)}}$. While $\phi^{(u)}$ depends on the metric used, when $f^m x = x$

$$S_m \phi^{(u)}(x) = -\log \mathrm{Jac}(Df^m : E^u_x \to E^u_x)$$

does not depend on the metric (this Jacobian is the absolute value of the determinant). By 4.5 one sees that the measure μ^+ on Ω_s and $P(\phi^{(u)})$ do not depend on which metric is used.

4.7. Volume Lemma. *Let*

$$B_x(\varepsilon, m) = \left\{ y \in M : d(f^k x, f^k y) \le \varepsilon \ \text{ for all } k \in [0, m) \right\} .$$

If $x \in \Omega_s$ is a $\mathscr{C}^2$ basic set and $\varepsilon > 0$ is small, then there is a constant C_ε so that

$$m(B_x(\varepsilon, m)) \in [C_\varepsilon^{-1}, C_\varepsilon] \ \exp(S_m \phi^{(u)}(x))$$

for all $x \in \Omega_s$.

Proof. See 4.2 of [9]. □

4.8. Proposition. *Let Ω_s be a $\mathscr{C}^2$ basic set.*

(a) *Letting $B(\varepsilon, n) = \bigcup_{x \in \Omega_s} B_x(\varepsilon, n)$, one has (for small $\varepsilon > 0$)*

$$P_{f|\Omega_s}(\phi^{(u)}) = \lim_{n \to \infty} \frac{1}{n} \log m(B(\varepsilon, n)) \le 0 \, .$$

(b) *Let $W^s_\varepsilon(\Omega_s) = \bigcup_{x \in \Omega_s} W^s_\varepsilon(x)$. If $m(W^s_\varepsilon(\Omega_s)) > 0$, then*

$$P_{f|\Omega_s}(\phi^{(u)}) = 0 \quad \text{and} \quad h_{\mu^+}(f) = -\int \phi^{(u)} d\mu^+ \, .$$

Proof. Call $E \subset M$ (n, ε)-*separated* if whenever y, z are two distinct points in E, one can find $k \in [0, n)$ with $d(f^k y, f^k z) > \delta$. Choose $E_n(\delta)$ maximal among the (n, ε)-separated subsets of Ω_s. For $x \in \Omega_s$ one has $x \in B_y(\varepsilon, n)$ for some $y \in E_n(\delta)$; otherwise $E_n(\delta) \cup \{x\}$ is (n, ε)-separated. Then $B_x(\varepsilon, n) \subset B_y(\delta + \varepsilon, n)$, $B(\varepsilon, n) \subset \bigcup_{y \in E_n(\delta)} B_y(\delta + \varepsilon, n)$ and by 4.7

$$(\star) \qquad m(B(\varepsilon, n)) \le C_{\delta + \varepsilon} \ \textstyle\sum_{y \in E_n(\delta)} \exp(S_n \phi^{(u)}(y)).$$

For $\delta \le \varepsilon$, $\bigcup_{y \in E_n(\delta)} B_y(\delta/2, n) \subset B(\varepsilon, n)$ is a disjoint union and so

$$(\star\star) \qquad m(B(\varepsilon,n)) \geq C_{\delta/2}^{-1} \; \textstyle\sum_{y\in E_n(\delta)} \exp(S_n\phi^{(u)}(y)).$$

Since $\phi^{(u)}$ is Hölder, we have

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

for some $a, \theta > 0$ and all $x, y \in \Omega_s$. Suppose $x \in B_y(\varepsilon, n) \cap \Omega_s$. Then for $j \in [0, n)$

$$d(f^j x, f^j y) < \alpha^{\min\{j, n-j-1\}}$$

by Lemma 4.2. Hence

$$\begin{aligned} |S_n\phi(x) - S_n\phi(y)| &\leq \sum_{j=0}^{n-1} |\phi^{(u)}(f^j x) - \phi^{(u)}(f^j y)| \\ &\leq 2a \sum_{k=0}^{\infty} \alpha^{k\theta} = \gamma \, . \end{aligned}$$

Fix $\delta \leq \varepsilon$ and let $\mathcal{U}$ be an open cover of Ω_s with $\operatorname{diam}(\mathcal{U}) < \delta$. Let $\Gamma \subset \mathcal{U}^n$ cover Ω_s. For each $y \in E_n(\delta)$ pick $\underline{\mathrm{U}}_y \in \Gamma$ with $y \in X(\underline{\mathrm{U}}_y)$. Then $S_n\phi^{(u)}(\underline{\mathrm{U}}_y) \geq S_n\phi^{(u)}(y)$. If $\underline{\mathrm{U}}_y = \underline{\mathrm{U}}_{y'}$, then $d(f^k y, f^k y') \leq \operatorname{diam}(\mathcal{U}) < \delta$ and $y = y'$ as $E_n(\delta)$ is (n,δ)-separated. Thus

$$\sum_{\underline{\mathrm{U}}\in\Gamma} \exp(S_n\phi^{(u)}(\underline{\mathrm{U}})) \geq \sum_{y\in E_n(\delta)} \exp(S_n\phi^{(u)}(y)).$$

Using this together with $(\star)$ above one gets

$$\begin{aligned} P(\phi^{(u)}, \mathcal{U}) &= \lim_{n\to\infty} \frac{1}{n} \log \inf_{\Gamma} \sum_{\underline{\mathrm{U}}\in\Gamma} \exp(S_n\phi^{(u)}(\underline{\mathrm{U}})) \\ &\geq \limsup_{n\to\infty} \frac{1}{n} \log m(B(\varepsilon,n)) \, . \end{aligned}$$

Letting $\operatorname{diam}(\mathcal{U}) \to 0$, one replaces $P(\phi^{(u)}, \mathcal{U})$ with $P(\phi^{(u)})$.

Now let $\mathcal{U}$ be an open cover and let δ be a Lebesgue number for $\mathcal{U}$. For each $y \in E_n(\delta)$ one can pick $\underline{\mathrm{U}}_y \in \mathcal{U}^n$ with $B_y(\delta, n) \cap \Omega_s \subset X(\underline{\mathrm{U}}_y)$.
Let $\Gamma = \{\underline{\mathrm{U}}_y : y \in E_n(\delta)\}$. Then Γ covers Ω_s since every $x \in \Omega_s$ lies in some $B_y(\delta, n)$ with $y \in E_n(\delta)$. Also

$$S_n\phi^{(u)}(\underline{\mathrm{U}}_y) \leq S_n\phi^{(u)}(y) + \gamma$$

and so

$$\begin{aligned} Z_n(\phi^{(u)}, \mathcal{U}) &\leq \sum_{\underline{\mathrm{U}}_y\in\Gamma^n} \exp(S_n\phi^{(u)}(\underline{\mathrm{U}}_y)) \\ &\leq e^\gamma \sum_{y\in E_n(\delta)} \exp(S_n\phi^{(u)}(y)) \, . \end{aligned}$$

Using $(\star\star)$ we get

$$\begin{aligned}P(\phi^{(u)}, \mathfrak{U}) &= \lim_{n\to\infty} \frac{1}{n} \log Z_n(\phi^{(u)}, \mathfrak{U}) \\ &\le \liminf_{n\to\infty} \frac{1}{n} \log m(B(\varepsilon, n))\,.\end{aligned}$$

Noting that $m(B(\varepsilon,n)) \le m(M)$, the proof of (a) is complete.

Now $m(B(\varepsilon,n)) \ge m(W^s_\varepsilon(\Omega_s))$ as $W^s_\varepsilon(\Omega_s) \subset B(\varepsilon,n)$. If $m(W^s_\varepsilon(\Omega_s)) > 0$, the formula from (a) yields $P(\phi^{(u)}) = 0$. Since μ^+ is an equilibrium state for $\phi^{(u)}$,

$$h_{\mu^+}(f) + \int \phi^{(u)} d\mu^+ = P(\phi^{(u)}) = 0\,. \quad \square$$

A basic set Ω_s is an *attractor* if it has small neighborhood U with $f(U) \subset U$. By Proposition 3.10 this is equivalent to $W^s_\varepsilon(\Omega_s)$ being a neighborhood of Ω_s.

4.9. Lemma. *Let Ω_s be a $\mathscr{C}^1$ basic set. If $W^u_\varepsilon(x) \subset \Omega_s$ for some $x \in \Omega_s$, then Ω_s is an attractor. If Ω_s is not an attractor there exists $\gamma > 0$ such that for every $x \in \Omega_s$, there is $y \in W^u_\varepsilon(x)$ with $d(y, \Omega_s) > \gamma$.*

Proof. If $W^u_\varepsilon(\Omega_s) \subset \Omega_s$, then

$$U_x = \bigcup \{W^s_\varepsilon(y) : y \in W^u_\varepsilon(x)\}$$

is a neighborhood of x in M (see Lemma 4.1 of [11]). Choose a periodic point $p \in U_x$, say $f^m x = x$. For some small β one has $W^u_\beta(p) \subset U_x$; if $z \in W^u_\beta(p)$ lies in $W^s_\varepsilon(y)$ $(y \in W^u_\varepsilon(\Omega_s))$, then $d(f^n z, f^n y) < \varepsilon$ and $d(f^{-n}z, f^{-n}p) < \beta$ for $n \ge 0$. By Theorem 3.9 one has $z \in \Omega_s$ and $W^u_\beta(p) \subset \Omega_s$. Then also $W^u(p) = \bigcup_{k\ge 0} f^{mk} W^u_\beta(p) \subset \Omega_s$.

Now $X_p = \overline{W^u(p)}$ and $\Omega_s = X_p \cup fX_p \cup \cdots \cup f^N X_p$ for some N. For each $x \in \bigcup_{k=0}^N f^k W^u(p) = Y$ one has $W^u_\varepsilon(x) \subset \Omega_s$ and so U_x as defined above is a neighborhood of x in M. Since $W^s_\varepsilon(x)$, $W^u_\varepsilon(x)$ depend continuously on $x \in \Omega_s$, one can find a $\delta > 0$ independent of x so that U_x contains the 2δ-ball $B_x(2\delta)$ about x in M for all $x \in Y$ (see [11, Lemma 4.1]). Then

$$B_{\Omega_s}(\delta) \subset \bigcup \{U_x : x \in Y\} \subset W^s_\varepsilon(\Omega_s)$$

and Ω_s is an attractor.

To prove (b) notice that the set

$$V_\gamma = \{x \in \Omega_s : d(y, \Omega_s) > \gamma \text{ for some } y \in W^u_\varepsilon(x)\}$$

is open because $W^u_\varepsilon(x)$ varies continuously with x. Also V_γ increases when γ decreases and $\bigcup_{\gamma>0} V_\gamma = \Omega_s$ by statement (a). By compactness $V_\gamma = \Omega_s$ for some $\gamma > 0$. $\square$

4.10. Second Volume Lemma. *Let Ω_s be a $\mathscr{C}^2$ basic set. For small $\varepsilon, \delta > 0$ there is a $d = d(\varepsilon, \delta) > 0$ so that*

$$m(B_y(\delta, n)) \geq d\ m(B_x(\varepsilon, n))$$

whenever $x \in \Omega_s$ and $y \in B_x(\varepsilon, n)$.

Proof. See 4.3 of [9]. □

4.11. Theorem. *Let Ω_s be a $\mathscr{C}^2$ basic set. The following are equivalent:*

(a) *Ω_s is an attractor.*
(b) *$m(W^s(\Omega_s)) > 0$.*
(c) *$P_{f|\Omega_s}(\phi^{(u)}) = 0$.*

Proof. As $W^s(\Omega_s) = \bigcup_{n=0}^{\infty} f^{-n} W^s_\varepsilon(\Omega_s)$, (b) is equivalent to $m(W^s_\varepsilon(\Omega_s)) > 0$. If Ω_s is an attractor, then (b) is true since $W^s_\varepsilon(\Omega_s)$ is a neighborhood of Ω_s. (c) follows from (b) by Proposition 4.8 (b). We finish by assuming Ω_s is not an attractor and showing $P(\phi^{(u)}) < 0$.

Given a small $\varepsilon > 0$ choose γ as in 4.9. Pick N so that

$$f^N W^u_{\gamma/4}(x) \supset W^u_\varepsilon(f^N x)$$

for all $x \in \Omega_s$. Let $E \subset \Omega_s$ be (γ, n)-separated. For $x \in E$ there is a $y(x, n) \in B_x(\gamma/4, n)$ with

$$d(f^{n+N} y(x, n), \Omega_s) > \gamma$$

(since $f^n B_x(\gamma/4, n) \supset W^u_{\gamma/4}(f^n x)$ and $f^N W^u_{\gamma/4}(f^n x) \supset W^u_\varepsilon(f^{N+n} x)$). Choose $\delta \in (0, \gamma/4)$ so that $d(f^N z, f^N y) < \gamma/2$ whenever $d(z, y) < \delta$. Then

$$B_{y(x,n)}(\delta, n) \subset B_x(\gamma/2, n),$$

$$f^{n+N} B_{y(x,n)}(\delta, n) \cap B_{\Omega_s}(\gamma/2) = \emptyset\,.$$

Hence $B_{y(x,n)}(\delta, n) \cap B(\gamma/2, n + N) = \emptyset$. Using the Second Volume Lemma

$$\begin{aligned} m(B(\gamma/2, n)) - m(B(\gamma/2, n + N)) &\geq \sum_{x \in E} m(B_{y(x,n)}(\delta, n)) \\ &\geq d(3\gamma/2, \delta) \sum_{x \in E} m(B_x(3\gamma/2, n)) \\ &\geq d(3\gamma/2, \delta)\ m(B(\gamma/2, n))\,. \end{aligned}$$

Therefore, setting $d = d(3\gamma/2, \delta)$

$$m(B(\gamma/2, n + N)) \leq (1 - d)\ m(B(\gamma/2, n))$$

and by Proposition 4.8 (a)

$$P_{f|\Omega_s}(\phi^{(u)}) \leq \frac{1}{N} \log(1 - d) < 0\,. \quad \square$$

Remark. *It is possible to find a $\mathscr{C}^1$ basic set (a horseshoe) which is not an attractor but nevertheless has $m(W^s(\Omega_s)) > 0$ [8].*

C. Attractors and Anosov diffeomorphisms

Because $M = \bigcup_{k=1}^{r} W^s(\Omega_k)$, Theorem 4.11 implies that m-almost all $x \in M$ approach an attractor under the action of a $\mathscr{C}^2$ Axiom A diffeomorphism f. This leads us next to the following result.

4.12. Theorem. *Let Ω_s be a $\mathscr{C}^2$ 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 $g : M \to \mathbb{R}$ (i.e., *x is a generic point for μ^+*).

Proof. Let us write $\overline{g}(n,x) = \frac{1}{n}\sum_{k=0}^{n-1} g(f^k x)$ and $\overline{g} = \int g \; d\mu^+$. Fix $\delta > 0$ and define the sets

$$\begin{aligned} C_n(g,\delta) &= \{x \in M : |\overline{g}(n,x) - \overline{g}| > \delta\} \\ E(g,\delta) &= \{x \in M : |\overline{g}(n,x) - \overline{g}| > \delta \text{ for infinitely many } n\} \\ &= \bigcap_{N=1}^{\infty} \bigcup_{n=N}^{\infty} C_n(g,\delta)\,. \end{aligned}$$

Choose $\varepsilon > 0$ so that $|g(x) - g(y)| < \delta$ when $d(x,y) < \varepsilon$.

Now fix $N > 0$ and choose sets $R_N, R_{N+1}, \ldots$ successively as follows. Let R_n $(n \geq N)$ be a maximal subset of $\Omega_s \cap C_n(g, 2\delta)$ satisfying the conditions:

(a) $B_x(\varepsilon,n) \cap B_y(\varepsilon,k) = \emptyset$ for $x \in R_n$, $y \in R_k$, $N \leq k < n$,
(b) $B_x(\varepsilon,n) \cap B_{x'}(\varepsilon,n) = \emptyset$ for $x, x' \in R_n$, $x \neq x'$.

If $y \in W^s_\varepsilon(\Omega_s) \cap C_n(g, 3\delta)$ $(n \geq N)$ and $y \in W^s_\varepsilon(z)$ with $z \in \Omega_s$, then $z \in C_n(g, 2\delta)$ by the choice of ε. By the maximality of R_n one has

$$B_z(\varepsilon,n) \cap B_x(\varepsilon,k) \neq \emptyset \quad \text{for some } x \in R_k, N \leq k \leq n\,.$$

Then $y \in B_z(\varepsilon,n) \subset B_z(\varepsilon,k) \subset B_x(2\varepsilon,k)$ and so

$$W^s_\varepsilon(\Omega_s) \cap \bigcup_{n=N}^{\infty} C_n(g,3\delta) \subset \bigcup_{k=N}^{\infty} \bigcup_{x\in R_k} B_x(2\varepsilon,k)\,.$$

Using the Volume Lemma 4.7 one gets

$$m\left(W^s_\varepsilon(\Omega_s) \cap \bigcup_{n=N}^{\infty} C_n(g,3\delta)\right) \leq c_{2\varepsilon} \sum_{k=N}^{\infty} \sum_{x\in R_k} \exp(S_k\phi^{(u)}(x)).$$

The definition of R_n shows that $V_N = \bigcup_{k=N}^{\infty} \bigcup_{x\in R_k} B_x(\varepsilon,k)$ is a disjoint union. The choice of ε gives $B_x(\varepsilon,k) \subset C_k(g,\delta)$ for $x \in R_k \subset C_k(g,2\delta)$ and so

$V_N \subset \bigcup_{k=N}^{\infty} C_k(g, \delta)$. Since the measure μ^+ is ergodic, the Ergodic Theorem implies

$$0 = \mu^+(E(g, \delta)) = \lim_{n \to \infty} \mu^+ \left(\bigcup_{n=N}^{\infty} C_n(g, \delta) \right)$$

and thus $\lim_{N \to \infty} \mu^+(V_N) = 0$. By 4.8 (b) one has $P_{f|\Omega_s}(\phi^{(u)}) = 0$ and then by 4.4

$$\mu^+(V_N) \geq b_\varepsilon \sum_{k=M}^{\infty} \sum_{x \in R_k} \exp(S_k \phi^{(u)}(x)).$$

As $\mu^+(V_N) \to 0$, the sum on the right approaches 0 as $N \to \infty$. Using the inequality of the preceding paragraph one sees

$$\lim_{N \to \infty} m \left(W_\varepsilon^s(\Omega_s) \cap \bigcup_{n=N}^{\infty} C_n(g, 3\delta) \right) = 0 \,.$$

This in turn implies $m(W_\varepsilon^s(\Omega_s) \cap E(g, 3\delta)) = 0$.

For $\delta' > 3\delta$ the set $E(g, \delta') \cap f^{-n} W_\varepsilon^s(\Omega_s) \subset f^{-n}(E(g, 3\delta) \cap W_\varepsilon^s(\Omega_s))$ has measure 0 since f preserves measure (w.r.t. m). Thus

$$m(E(g, \delta') \cap W^s(\Omega_s)) \leq \sum_{n=0}^{\infty} m \left(E(g, \delta') \cap f^{-n} W_\varepsilon^s(\Omega_s) \right) = 0 \,.$$

Fixing g still but letting $\delta' = \frac{1}{m} \to 0$ one gets $\lim_n \overline{g}(n, x) = \overline{g}$ for all $x \in W^s(\Omega_s)$ outside an m-null set $A(g)$. Let $\{g_k\}_{k=1}^{\infty}$ be a dense subset of $\mathscr{C}(M)$; for $x \in W^s(\Omega_s) \setminus \bigcup_{k=1}^{\infty} A(g_k)$ one gets that $\lim_n \overline{g}(n, x) = \overline{g}$ for all $g \in \mathscr{C}(M)$. $\square$

4.13. Corollary. *Suppose $f : M \to M$ is a transitive $\mathscr{C}^2$ Anosov diffeomorphism. If f leaves invariant a probability measure μ which is absolutely continuous with respect to m, then $\mu = \mu^+$.*

Proof. In this case $M = \Omega = \Omega_1$ is the spectral decomposition. Let $g \in \mathscr{C}(M)$. By the Ergodic Theorem there is a function $g^* : M \to \mathbb{R}$ so that

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n-1} g(f^k x) = g^*(x)$$

for μ-almost all x. Let A be the set of $x \in M$ with

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

Because $m(A) = 1$ and $\mu \ll m$, $\mu(A) = 1$. It follows that $g^*(x) = \int g d\mu^+$ for μ-almost all x. Then

$$\int g d\mu = \int g^* d\mu = \int g d\mu^+ .$$

As this holds for all $g \in \mathscr{C}(M)$, $\mu = \mu^+$. $\square$

Remark. *If M is connected, then $M = \Omega_1 = X_1$ and f is mixing. So μ above is Bernoulli.*

4.14. Theorem. *Let f be a transitive $\mathscr{C}^2$ Anosov diffeomorphism. The following are equivalent:*

(a) *f admits an invariant measure of the form $d\mu = hdm$ with h a positive Hölder function.*
(b) *f admits an invariant measure μ absolutely continuous w.r.t. m.*
(c) *$Df^n : T_xM \to T_xM$ has determinant 1 whenever $f^n x = x$.*

Proof. Clearly (a) implies (b). Assume (b) holds. Let $\lambda^{(s)}(x)$ be the Jacobian of $Df : E^s_{f^{-1}x} \to E^s_x$ and $\phi^{(s)}(x) = \log \lambda^{(s)}(x)$. Now f^{-1} is Anosov with $E^u_{x,f^{-1}} = E^s_{x,f}$ an $E^s_{x,f^{-1}} = E^u_{x,f}$. Also

$$\begin{aligned} \lambda^{(u)}_{f^{-1}}(x) &= \text{Jacobian } \ Df^{-1} : E^s_x \to E^s_{f^{-1}x} \\ &= \lambda^{(s)}(x)^{-1} \end{aligned}$$

and so $\phi^{(u)}_{f^{-1}}(x) = -\log \lambda^{(u)}_{f^{-1}}(x) = \phi^{(s)}(x)$. There is an invariant measure μ^- so that

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

for m-almost all x; μ^- is the unique equilibrium state for $\phi^{(u)}_{f^{-1}}$ w.r.t. f^{-1}.

Notice that equilibrium states w.r.t. f^{-1} are the same as those w.r.t. f; for $\mathscr{M}_f(M) = \mathscr{M}_{f^{-1}}(M)$ and $h_\nu(f) = h_\nu(f^{-1})$. So $\mu^- = \mu_{\phi^{(s)}}$. Applying 4.13 to both f and f^{-1} we see

$$\mu_{\phi^{(u)}} = \mu^+ = \mu = \mu^- = \mu_{\phi^{(s)}} .$$

By 4.8 (b), $P(\phi^{(u)}) = 0 = P(\phi^{(s)})$. By 4.5 one has, for $f^n x = x$,

$$\sum_{k=0}^{n-1} \phi^{(u)}(f^k x) - \sum_{k=0}^{n-1} \phi^{(s)}(f^k x) = 0.$$

Exponentiating,

$$1 = (\det\ Df^n|_{E^u_x}) \ (\det\ Df^n|_{E^s_x}) = (\det\ Df^n|_{T_xM}) .$$

Now assume (c) and let $\phi(x) = \log \operatorname{Jac}\ (Df : T_xM \to T_{fx}M)$. Then

$$\begin{aligned} S_n\phi(x) &= \log \prod_{k=0}^{n-1} \text{Jac } (Df : T_{f^k x}M \to T_{f^{k+1}x}M) \\ &= \log \text{Jac } (Df^n : T_x M \to T_{f^k x}M) = 0 \end{aligned}$$

when $f^k x = x$. By Proposition 4.5 (with $\psi = 0$, $K = 0$) there is a Hölder $u : M \to \mathbb{R}$ with $\phi(x) = u(fx) - u(x)$. Let $h(x) = e^{u(x)}$.
Thinking of $\mu = hdm$ as the absolute value of a form

$$\begin{aligned} f^*(hdm)(fx) &= h(x)\ e^{\phi(x)}\ dm(fx) \\ &= h(fx)\ dm(fx) = (h\ dm)(fx)\,. \end{aligned}$$

So μ is f-invariant. □

Remark. *Actually h above will be $\mathscr{C}^1$. See [13, 14].*

4.15. Corollary. *Among the $\mathscr{C}^2$ Anosov diffeomorphisms the ones that admit no invariant measures $\mu \ll m$ are open and dense.*

Proof. [18], page 36. We use condition 4.14 (c). Suppose $f^n x = x$ and $\det(Df^n|_{T_x M}) \neq 1$. For f_1 near f, f_1 will be Anosov and have a periodic point x_1 near x with $f_1^n x_1 = x_1$. Then $\det(Df_1^n|_{T_{x_1} M})$ will be near $\det(Df^n|_{T_x M})$ and not equal to one. We are using stability theory not covered in these notes. On the other hand, if $f^n x = x$ with $\det(Df^n|_{T_x M}) = 1$, this condition is destroyed by the right small perturbation of f. □

In the case where f admits no invariant $\mu \ll m$, the measure m is actually dissipative [10].

References

This chapter contains the main theorems of these notes. The notes as a whole constitute a version of Sinai's program [18] for applying statistical mechanics to diffeomorphisms. It was Ruelle who carried Sinai's work on Anosov diffeomorphisms over to Axiom A attractors [11] and brought in the formalism of equilibrium states [15].

Theorem 4.1 is from [6, 7]; see [18] for the Anosov case. Results 4.5, 4.13, 4.14 and 4.15 came from [13, 14, 18]. Section 4 is taken verbatim from [9]. Theorem 4.12 is due to Ruelle [16] (we followed the proof in [9]). Ruelle [16] along these lines also proved that $f^n\mu \to \mu^+$ when $\mu \ll m$ has support in a neighborhood of an attractor; for the Anosov case this result is due to Sinai [17]. That μ^+ is Bernoulli in the transitive Anosov case is due to Azencott [4]. The ergodic theory of Anosov diffeomorphisms with $\mu^+ \ll m$ has been studied in many papers; see for instance [2] or [3].

In case $\phi^{(u)}$ is a constant function μ^+ is the unique invariant measure which maximizes entropy ($\phi = 0$ and $\phi^{(u)}$ have the same equilibrium state). For hyperbolic automorphisms of the 2-torus μ^+ is Haar measure and construction in 4.1 is due to Adler and Weiss [1]. This paper is very important in the development of the subject and is good reading. When $\phi^{(u)} \neq C$ the measure μ which maximizes entropy still has the following geometrical significance: the periodic points of Ω_s are equidistributed with respect to μ [5]. K. Sigmund [19] studied the *generic* properties of measures on Ω_s.

1. R. Adler, B. Weiss: *Similarity of automorphisms of the torus*, Memoirs of the American Mathematical Society **98**, American Mathematical Society, Providence, R.I. 1970.
2. D. Anosov: Geodesic flows on closed Riemann manifolds with negative curvature, *Proc. Steklov Inst. Math.* **90** (1967).
3. D. Anosov, Ya. Sinai: Certain smooth ergodic systems, *Uspehi Mat. Nauk* **22** (1967) no. 5 (137), 107–172. English translation: *Russ. Math. Surveys* **22** (1967), no. 5, 103-167.
4. R. Azencott: Difféomorphismes d'Anosov et schémas de Bernoulli, *C. R. Acad. Sci. Paris, Sér. A-B* **270** (1970), A1105-A1107.
5. R. Bowen: Periodic points and measures for Axiom A diffeomorphisms, *Trans. Amer. Math. Soc.* **154** (1971), 377-397.

6. R. Bowen: Some systems with unique equilibrium states, *Math. Systems Theory* **8** (1974/75), no.3, 193-202.
7. R. Bowen: Bernoulli equilibrium states for Axiom A diffeomorphisms, *Math. Systems Theory* **8** (1974/75), no. 4, 289–294.
8. R. Bowen: A horseshoe with positive measure, *Invent. Math.* **29** (1975), no. 3, 203-204.
9. R. Bowen, D. Ruelle: The ergodic theory of Axiom A flows, *Invent. Math.* **29** (1975), no. 3, 181-202.
10. B. Gurevič, V. Oseledec: Gibbs distributions, and the dissipativity of C-diffeomorphisms, *Soviet Math. Dokl.* **14** (1973), 570–573.
11. M. Hirsch, J. Palis, C. Pugh and M. Shub: Neighborhoods of hyperbolic sets, *Invent. Math.* **9** (1969/1970), 121–134.
12. M. Hirsch and C. Pugh: Stable manifolds and hyperbolic sets, *Proc. Sympos. Pure Math., Vol. XIV, Berkeley, Calif., 1968*, pp. 133–163, Amer. Math. Soc., Providence, R.I., 1970.
13. A. N. Livšic: Cohomology of dynamical systems, *Math. USSR-Izv.* **6** (1972), 1278-1301.
14. A. N. Livšic, Ya. Sinai: Invariant measures that are compatible with smoothness for transitive C-systems, *Soviet Math. Dokl.* **13** (1972), 1656-1659.
15. D. Ruelle: Statistical mechanics on a compact set with $\mathbb{Z}^\nu$ action satisfying expansiveness and specification, *Trans. Amer. Math. Soc.* **185** (1973), 237-251.
16. D. Ruelle: A measure associated with Axiom A attractors, *Amer. J. Math.* **98** (1976), no. 3, 619-654.
17. Ya. Sinai: Markov partitions and C-diffeomorphisms, *Func. Anal. and its Appl.* **2** (1968), no. 1, 64-89.
18. Ya. Sinai: Gibbs measures in ergodic theory, *Uspehi Mat. Nauk* **27** (1972), no. 4, 21–64. English translation: *Russian Math. Surveys* **27** (1972), no. 4, 21–69.
19. K. Sigmund: Generic properties of invariant measures for Axiom A diffeomorphisms, *Invent. Math.* **11** (1970), 99-109.

Index