## 8. Thermodynamic formalism

## §8.1 Introduction

Historically, ergodic theory was a branch of thermodynamics and statistical mechanics. Although we will not describe the precise relationships between thermodynamics and ergodic theory (see Ruelle's book in the references for this), the connections motivate several interesting constructions which turn out to have much wider applications. This is the heart of thermodynamic formalism.

In this lecture we study the thermodynamic formalism for aperiodic shifts of finite type. In Lecture 3 we constructed Markov and Bernoulli measures and, in Lecture 4, saw that they were ergodic; in this lecture we will introduce a much wider class of invariant ergodic measures known as equilibrium states. We will define these as eigenmeasures of (the dual operator of) a linear operator acting on an appropriate space of functions. This linear operator, known as a transfer operator, is a key object in the study of thermodynamic formalism and hyperbolic dynamics.

## $\S 8.2$ Shifts of finite type

In this section we recall some of the definitions related to shifts of finite type that we have already seen.

## §8.2.1 Aperiodic shifts of finite type

Let $A$ be a $k \times k$ matrix with entries in in $\{0,1\}$. We will assume throughout that $A$ is aperiodic, namely that there exists $n \geq 1$ such that every entry in $A^{n}$ is positive.

Define the (one-sided) shift of finite type

$$
\Sigma=\left\{\left(x_{j}\right)_{j=0}^{\infty} \mid x_{j} \in\{1, \ldots, k\}, A_{x_{j}, x_{j+1}}=1\right\}
$$

Thus $\Sigma$ consists of all infinite sequences of symbols chosen from $\{1, \ldots, k\}$ subject to the restriction that symbol $i$ can be followed by symbol $j$ if and only if the $(i, j)$ th entry of the matrix $A$ is equal to 1 . In previous lectures we have used $\Sigma_{A}^{+}$to denote this space; in this lecture the choice of $A$ will be fixed and we will only work with one-sided shifts, hence we shall use the simpler notation.

## $\S$ 8.2.2 A metric on $\Sigma$

We can define a metric on $\Sigma$ as follows. Fix a choice of $\theta \in(0,1)$. For two sequences $x=\left(x_{j}\right)_{j=0}^{\infty}, y=\left(y_{j}\right)_{j=0}^{\infty} \in \Sigma$ we define

$$
n(x, y)= \begin{cases}n & \text { if } x_{j}=y_{j} \text { for } j=0, \ldots, n-1, x_{n} \neq y_{n} \\ \infty & \text { if } x=y\end{cases}
$$

so that $n(x, y)$ is the first place in which the sequences $x, y$ disagree. We then define a metric by

$$
d_{\theta}(x, y)=\theta^{n(x, y)}
$$

Then $\Sigma$ is a compact, totally disconnected, perfect metric space.
Different choices of $\theta$ will give different metrics; however they all give the same topology. Below we shall be interested in functions $f: \Sigma \rightarrow \mathbb{R}$ that are Lipschitz continuous with respect to $d_{\theta}$. The class of these functions does, however, depend on $\theta$.

## $\S$ 8.2.3 Cylinder sets

Recall that if $i_{0}, \ldots, i_{n-1} \in\{1, \ldots, k\}$ and $m \geq 0$ then we define the cylinder set $\left[i_{0}, \ldots, i_{n-1}\right]$ to be the set of all sequences $\left(x_{j}\right)_{j=0}^{\infty} \in \Sigma$ with the property that $x_{j}=i_{j}$ for $0 \leq j \leq n-1$. We call this a cylinder of length $n$ or rank $n$.

Notice that the cylinder $\left[i_{0}, \ldots, i_{n-1}\right]$ is non-empty if and only if $A_{i_{j}, i_{j+1}}=$ 1 for $j=0, \ldots, n-2$. Also note that cylinders are both open and closed.

The collection of all cylinders forms an algebra which generates the Borel $\sigma$-algebra.

## §8.2.4 The shift map

Define the shift map $\sigma: \Sigma \rightarrow \Sigma$ by

$$
\sigma\left(x_{0}, x_{1}, x_{2}, \ldots\right)=\left(x_{1}, x_{2}, x_{3}, \ldots\right)
$$

Note that $\sigma$ is not invertible. Indeed, given $x=\left(x_{0}, x_{1}, \ldots\right) \in \Sigma$ there are at most $k$ preimages under $\sigma$. This is because the preimages of $x$ have the form

$$
\sigma^{-1}(x)=\left\{\left(i, x_{0}, x_{1}, \ldots\right) \mid i \in\{1, \ldots, k\}, A_{i, x_{0}}=1\right\}
$$

Note that $\sigma$ is a continuous transformation using the metric $d_{\theta}$. In fact, $\sigma$ satisfies a stronger property: $\sigma$ is Lipschitz continuous with respect to $d_{\theta}$ :

$$
d_{\theta}(\sigma(x), \sigma(y)) \leq \theta^{-1} d_{\theta}(x, y)
$$

## §8.3 Hölder function spaces

## §8.3.1 Hölder continuous and Lipschitz continuous functions

We will be interested in spaces of functions defined on $\Sigma$. In Lecture 6 we worked with the space $C(\Sigma, \mathbb{R})$ of continuous function defined on $\Sigma$. Here we need to impose extra regularity conditions on the functions we consider. We will work with functions that satisfy a Hölder continuity condition.

Let $f: \Sigma \rightarrow \mathbb{C}$ be complex-valued function defined on $\Sigma$. Define

$$
\operatorname{var}_{n}(f)=\sup \left\{|f(x)-f(y)| \mid x, y \in \Sigma, x_{j}=y_{j} \text { for } j=0, \ldots, n-1\right\}
$$

to be the $n$th variation of $f$. Note that $\operatorname{var}_{n}(f)$ measures how much $f$ can vary on cylinders of length $n$.

Definition. For $f: \Sigma \rightarrow \mathbb{C}$ define

$$
|f|_{\theta}=\sup \left\{\left.\frac{\operatorname{var}_{n}(f)}{\theta^{n}} \right\rvert\, n=0,1,2, \ldots\right\}
$$

to be the least Hölder constant of $f$.
It is easy to see that $|f|_{\theta}<\infty$ if and only if there exists $C>0$ such that

$$
\begin{equation*}
|f(x)-f(y)| \leq C d_{\theta}(x, y) \text { for all } x, y \in \Sigma \tag{8.1}
\end{equation*}
$$

and $|f|_{\theta}$ is the least such $C>0$ for which (8.1) holds. Note that the condition (8.1) says that $f$ is Lipschitz continuous with respect to the metric $d_{\theta}$. It is customary in thermodynamic formalism to say instead that $f$ is Hölder of exponent $\theta$. This is because if $f$ is Hölder continuous of exponent $\alpha$ with respect to $d_{\theta}$, i.e. $|f(x)-f(y)| \leq C d_{\theta}(x, y)^{\alpha}$, then $f$ is Lipschitz continuous with respect to $d_{\theta^{\alpha}}$. Notice that if $f$ is Hölder of exponent $\theta$ then it is necessarily continuous.

Define

$$
F_{\theta}(\mathbb{C})=\left\{f: \Sigma \rightarrow \mathbb{C} \mid\|f\|_{\theta}<\infty\right\}
$$

to be the space of all Hölder continuous functions of exponent $\theta$. We shall also be interested in the space $F_{\theta}(\mathbb{R})$ of real-valued Hölder functions of exponent $\theta$.

Note that $|\cdot|_{\theta}$ is a semi-norm, but is not a norm. This is because $|f|_{\theta}=0$ if $f$ is a constant function. We define a norm on $F_{\theta}$ by setting

$$
\|f\|_{\theta}=|f|_{\infty}+|f|_{\theta}
$$

where $|f|_{\infty}=\sup |f(x)|$ is the uniform norm of $f$. We have the following important result.

## Proposition 8.1

The space $F_{\theta}(\mathbb{C})$ is a complex Banach space with respect to the norm $\|\cdot\|_{\theta}$.

Remark Notice that if $\theta_{1} \leq \theta_{2}$ then $F_{\theta_{1}}(\mathbb{C}) \supset F_{\theta_{2}}(\mathbb{C})$.

## §8.3.2 Locally constant functions

A particularly important and tractable class of functions are those which only depend on finitely many co-ordinates.

Let $f: \Sigma \rightarrow \mathbb{C}$. We say that $f$ is locally constant if $f$ depends on only finitely many co-ordinates of $\Sigma$. That is, there exists $n \geq 0$ such that $f(x)=f\left(x_{0}, x_{1}, \ldots, x_{n-1}\right)$. Equivalently, $f$ is constant on cylinders of length $n$.

Clearly, if $f$ is locally constant, then with $n$ as above, $\operatorname{var}_{m}(f)=\operatorname{var}_{n}(f)$ for all $m \geq n$. Hence $|f|_{\theta}<\infty$ for any $\theta \in(0,1)$. Hence $f \in F_{\theta}(\mathbb{C})$ for all $\theta \in(0,1)$.

## §8.4 Transfer operators

Let $f \in F_{\theta}(\mathbb{R})$. We define the transfer operator or Ruelle operator to be the map

$$
L_{f}: F_{\theta}(\mathbb{C}) \rightarrow F_{\theta}(\mathbb{C})
$$

given by

$$
L_{f} w(x)=\sum_{y \in \Sigma, \sigma(y)=x} e^{f(y)} w(y)=\sum_{i \text { s.t. } A_{i, x_{0}}=1} e^{f\left(i, x_{0}, x_{1}, \ldots\right)} w\left(i, x_{0}, x_{1}, \ldots\right)
$$

That is, given a function $w \in F_{\theta}(\mathbb{C})$, we define a new function as follows: given $x$, consider all the preimages of $x$ under $\sigma$, then evaluate $w$ at each of these preimages and sum them, each weighted according to the function $\exp (f)$. With this in mind, we often refer to $f$ as a weight function.

Note that the iterates of $L_{f}$ have the form

$$
L_{f}^{n} w(x)=\sum_{y \in \Sigma, \sigma^{n}(y)=x} e^{f^{n}(y)} w(y)
$$

where $f^{n}(y)=\sum_{j=0}^{n-1} f\left(\sigma^{j} y\right)$.
It is clear that $L_{f}$ is a linear operator on the Banach space $F_{\theta}(\mathbb{C})$. It is straightforward to check that $L_{f}$ is bounded.

## Proposition 8.2

Let $f \in F_{\theta}(\mathbb{R})$. Then the transfer operator $L_{f}: F_{\theta}(\mathbb{R}) \rightarrow F_{\theta}(\mathbb{R})$ is a bounded linear operator.

We will be interested in understanding the spectral properties of $L_{f}$, i.e. we want to determine the eigenvalues of $L_{f}$. Before we do this, it is instructive to consider a specific example.

## §8.5 A finite-dimensional transfer operator

Let $\Sigma=\left\{\left(x_{j}\right)_{j=0}^{\infty} \mid x_{j} \in\{0,1\}\right\}$ denote the full 1-sided 2-shift on the two symbols 0,1 . Let $\sigma: \Sigma \rightarrow \Sigma$ be the shift map.

Note that each $x=\left(x_{j}\right)_{j=0}^{\infty} \in \Sigma$ has exactly two pre-images under $\sigma$ : $\left(0, x_{0}, x_{1}, \ldots\right)$ and $\left(1, x_{0}, x_{1}, \ldots\right)$. We denote these points by $(0 x)$ and ( $1 x$ ) respectively.

If $f: \Sigma \rightarrow \mathbb{R}$ is a locally constant function depending only on the first two co-ordinates. Then we can write $f(x)=f\left(x_{0}, x_{1}, \ldots\right)=f\left(x_{0}, x_{1}\right)$. That is, the function is determined once we know the values of $f$ on the cylinders [00], [01], [10], [11].

Fix $p, q \in(0,1)$. We define a locally constant function $f$ depending on the first two co-ordinates by

$$
f(00)=\log p, f(01)=\log (1-p), f(10)=\log q, f(11)=\log (1-q)
$$

We will consider the action of the transfer operator $L_{f}$ acting on the space of functions that depend only on the first co-ordinate. Note that if $w(x)$ depends only on the first co-ordinate, $w(x)=w\left(x_{0}\right)$, then $w$ is determined by knowing the values of $w(0), w(1)$. Thus we can view $L_{f}$ as a linear operator on $\mathbb{R}^{2}$ (or $\mathbb{C}^{2}$ if we are working with complex valued functions).

Suppose that $w$ depends only on the co-ordinate $x_{0}$. Then

$$
L_{f} w\left(x_{0}\right)=\sum_{y: \sigma(y)=x} e^{y} w(y)=e^{f\left(0 x_{0}\right)} w(0)+e^{f\left(1 x_{0}\right)} w(1)
$$

Thus $L_{f} w$ is also a function that depends only on the co-ordinate $x_{0}$, indeed

$$
\binom{L_{f} w(0)}{L_{f} w(1)}=\left(\begin{array}{cc}
p & 1-p \\
q & 1-q
\end{array}\right)\binom{w(0)}{w(1)} .
$$

Thus $L_{f}$ acts by the matrix

$$
P=\left(\begin{array}{ll}
p & 1-p \\
q & 1-q
\end{array}\right)
$$

Note that $P$ is a positive matrix. Hence the spectral properties of $P$ are determined by the Perron-Frobenius theorem, which we recall here.

## Theorem 8.3 (Perron-Frobenius)

Let $B$ be a non-negative aperiodic $k \times k$ matrix. Then
(i) there exists a positive eigenvalue $\lambda>0$ such that all other eigenvalues $\lambda_{i} \in \mathbb{C}$ satisfy $\left|\lambda_{i}\right|<\lambda ;$
(ii) the eigenvalue $\lambda$ is simple;
(iii) there is a unique right-eigenvector $v=\left(v_{1}, \ldots, v_{k}\right)^{T}$ such that $v_{j}>0$, $\sum_{j} v_{j}=1$, and $B v=\lambda v ;$
(iv) there is a unique left-eigenvector $u=\left(u_{1}, \ldots, u_{k}\right)$ such that $u_{j}, \sum_{j} u_{j}=$ 1 and $u B=\lambda u$;
(v) eigenvalues corresponding to eigenvalues other than $\lambda$ are not positive: i.e. at least one co-ordinate is positive and at least one co-ordinate is negative.

Note that for the function that is constantly equal to 1 we have $L_{f} 1=1$. (Equivalently, $(1,1)^{T}$ is a right-eigenvector for $P$ with eigenvalue 1, so that $P$ is a stochastic matrix.) If $L_{f} 1=1$ then we will say that the weight function $f$ is normalised.

The corresponding left-eigenvector of $P$ can easily be computed. Indeed, let

$$
p=\left(p_{0}, p_{1}\right)=\left(\frac{q}{q+1-p}, \frac{1-p}{q+1-p}\right) .
$$

Then it easy to check that $p P=p$ and $p_{0}+p_{1}=1$. Hence $p, P$ determine a Markov measure $\mu_{f}$ defined on cylinders by

$$
\mu_{f}\left[i_{0}, \ldots, i_{n}-1\right]=p_{i_{0}} P_{i_{0}, i_{1}} P_{i_{1}, i_{2}} \ldots P_{i_{n-2}, i_{n-1}}
$$

There is a close connection between $\mu_{f}$ and $L_{f}$. Suppose $w$ is a locally constant functions depending only on $x_{0}$. Then $L_{f} w$ takes the value $L_{f} w(0)$ on the cylinder [0] and the value $L_{f} w(1)$ on the cylinder [1]. Hence

$$
\begin{aligned}
\int L_{f} w d \mu_{f} & =\mu_{f}[0] L_{f} w_{0}+\mu_{f}[1] L_{f} w(1) \\
& =p_{0}(p w(0)+(1-p) w(1))+p(1)(q w(0)+(1-q) w(1)) \\
& =p P w \text { where } w=\binom{w(0)}{w(1)} \\
& =p w \\
& =p_{0} w(0)+p_{1} w(1) \\
& =\int w d \mu_{f}
\end{aligned}
$$

That is, $\mu_{f}$ is invariant under the action of $L_{f}$ (at least on functions of 1 variable).

One can also connect the measure $\mu_{f}$ to the ergodic sums of $f$. Let $x=\left(x_{0}, x_{1}, \ldots\right)$. Fix $n \geq 0$. Then

$$
\begin{aligned}
\mu_{f}\left[x_{0}, x_{1}, \ldots, x_{n-1}\right] & =p_{x_{0}} P_{x_{0}, x_{1}} P_{x_{1}, x_{2}} \cdots P_{x_{n-2}, x_{n-1}} \\
& =p_{x_{0}} e^{f\left(x_{0}, x_{1}\right)} e^{f\left(x_{1}, x_{2}\right)} \cdots e^{f\left(x_{n-2}, x_{n-1}\right)} \\
& =p_{x_{0}} e^{f^{n}(x)}
\end{aligned}
$$

where $f^{n}(x)=\sum_{j=0}^{n-1} f^{n}(x)$. Hence, taking $C=\max \left\{p_{0}, p_{1}\right\}$ we have that

$$
\begin{equation*}
\frac{1}{C} \leq \frac{\mu_{f}\left[x_{0}, x_{1}, \ldots, x_{n-1}\right]}{e^{f^{n}(x)}} \leq C \tag{8.2}
\end{equation*}
$$

This often allows us to replace the measure of a cylinder by an ergodic sum, which are normally easier to deal with.

## §8.6 Ruelle's Perron-Frobenius Theorem

In this section we study the case where the weight function $f$ is real and we consider the associated transfer operator acting on the real Banach space $F_{\theta}(\mathbb{R})$ of real-valued functions.

The first step to this is the following result. We assume, for convenience, that $L_{f} 1=1$; we shall see later that this is not a real restriction.

## Proposition 8.4 (Lasota-Yorke inequality)

Let $f \in F_{\theta}(\mathbb{C})$ and suppose that $L_{f} 1=1$. Then for all $w \in F_{\theta}(\mathbb{C})$ and $n \geq 0$ we have

$$
\left\|L_{f}^{n} w\right\|_{\theta} \leq C|w|_{\infty}+\theta^{n}|w|_{\theta}
$$

where $C>0$ depends only on $f$ and $\theta$.
Proof. Throughout, if $x=\left(x_{0}, x_{1}, \ldots\right)$ then $i x=\left(i, x_{0}, x_{1}, \ldots\right)$ (and we assume that $\left.A_{i, x_{0}}=1\right)$. Note that if $x, y \in \Sigma$ then $d(i x, i y) \leq \theta d(x, y)$.

The proof is by induction on $n$. When $n=1$, we estimate

$$
\begin{aligned}
\left|L_{f} w(x)-L_{f} w(y)\right| \leq & \sum\left|e^{f(i x)} w(i x)-e^{f(i y)} w(i y)\right| \\
\leq & \sum e^{f(i y)}\left|e^{f(i x)-f(i y)}-1\right||w(i x)| \\
& +\sum e^{f(i y)}|w(i x)-w(i y)|
\end{aligned}
$$

(the sums are all over $i$ for which $A_{i, x_{0}}=1$ ). Noting that

$$
\sup _{x \neq y} \frac{\left|e^{f(i x)-f(i y)}-1\right|}{d(x, y)} \leq \sum_{r=1}^{\infty} \frac{\theta^{r}|f|_{\theta}^{r} d(x, y)^{r-1}}{r!} \leq C_{0}
$$

for some constant $C_{0}>0$ and recalling that $\sum e^{f(i y)}=1$ (as $L_{f} 1=1$ ), we obtain

$$
\left|L_{f} w\right|_{\theta} \leq C_{0}|w|_{\infty}+\theta|w|_{\theta}
$$

Using induction, we assume that $\left|L_{f}^{n} w\right|_{\theta} \leq C_{n}|w|_{\theta}+\theta^{n}|w|_{\theta}$. Then

$$
\begin{aligned}
\left|L_{f}^{n+1} w\right|_{\theta} & \leq C_{n}\left|L_{f}^{n} w\right|_{\infty}+\theta^{n}\left|L_{f}^{n} w\right|_{\theta} \\
& \leq C_{n}|w|_{\infty}+\theta^{n}\left(C_{0}|w|_{\infty}+\theta|w|_{\theta}\right)
\end{aligned}
$$

Hence we can take $C_{n+1}=C_{n}+\theta^{n+1} C_{0} \leq C_{0} /(1-\theta)$. Hence, by induction

$$
\left|L_{f}^{n} w\right|_{\theta} \leq C|w|_{\infty}+\theta^{n}|w|_{\theta}
$$

and

$$
\left\|L_{f}^{n} w\right\|_{\theta} \leq(C+1)|w|_{\infty}+\theta^{n}|w|_{\theta} .
$$

We study the spectrum of $L_{f}$ in the case where $f \in F_{\theta}(\mathbb{R})$ is real-valued.

## Theorem 8.5 (Ruelle's Perron-Frobenius Theorem)

Let $A$ be an aperiodic matrix with entries in $\{0,1\}$, with associated shift of finite type $\Sigma$. Let $f \in F_{\theta}(\mathbb{R})$.
(i) There is a simple maximal positive eigenvalue $\lambda$ of $L_{f}: F_{\theta}(\mathbb{C}) \rightarrow F_{\theta}(\mathbb{C})$ with a corresponding strictly positive eigenfunction $h$.
(ii) The remainder of the spectrum is contained inside a disc in $\mathbb{C}$ of radius strictly smaller than $\lambda$.
(iii) There is a unique probability measure $\nu$ such that

$$
\int L_{f} v d \nu=\lambda \int v d \nu \text { for all } v \in C(\Sigma, \mathbb{R})
$$

Moreover, if $h$ is as in (i) and $\int h d \nu=1$ then the measure $\mu$ defined by $d \mu=h d \nu$ is a $\sigma$-invariant probability measure.
(iv) If $h$ is as in (ii) and $\int h d \nu=1$ then

$$
\frac{1}{\lambda^{n}} L_{f}^{n} v \rightarrow h \int v d \nu
$$

uniformly for all $v \in C(\Sigma, \mathbb{R})$.
Remark We give a complete proof (taken from the Astérisque book by Parry and Pollicott, see the references) of this below. However, there is the following abstract result due to Hennion that guarantees quasi-compactness.

Suppose that there are two complex Banach spaces $\left(B_{1},\|\cdot\|_{1}\right),\left(B_{2},\|\cdot\|_{2}\right)$ and an inclusion $\iota: B_{1} \rightarrow B_{2}$. Let $L: B_{1} \rightarrow B_{1}$ be a bounded linear operator. Suppose that
(i) $\iota: B_{1} \rightarrow B_{2}$ is a compact operator,
(ii) there exist sequences $r_{n}, R_{n}$ such that

$$
\left\|L^{n} w\right\|_{1} \leq R_{n}\|w\|_{2}+r_{n}\|w\|_{1} .
$$

Suppose that $\liminf _{n \rightarrow \infty} r_{n}^{1 / n}<\rho(L)$ (where $\rho(L)$ denotes the spectral radius of $L$ as an operator on $B_{1}$ ). Then $L$ has a spectral gap.

In our setting, $B_{1}=F_{\theta}(\mathbb{C})$ and $B_{2}=C(\Sigma, \mathbb{C})$.

Remark If a linear operator $L$ has the spectral properties given in (i) and (ii) of Theorem 8.5 then we say that $L$ has a spectral gap or is quasi-compact. Recall if a linear operator is compact, then it has at most countably many eigenvalues, the non-zero eigenvalues have finite multiplicity, and the only possible limit point of the eigenvalues is 0 . Thus, compact operators are quasi-compact (but obviously not conversely in general).

Remark In fact, one can obtain far more information about the spectrum of $L$ if we assume that (i) and (ii) in the first remark above hold. Define the essential spectrum to be the set of limit points of the spectrum of $L$; equivalently, these are the points in the spectrum with infinite multiplicity. Define the essential spectral radius $\rho_{\text {ess }}(L)$ to be the supremum of the modulus of the essential spectrum. Thus $L$ is quasi-compact if $\rho_{\text {ess }}(L)<\rho(L)$ ( $L$ is compact if $\left.\rho_{\text {ess }}(L)=0\right\}$ ). The theorem of Hennion alluded to above actually proves a stronger statement: that $\rho_{\text {ess }}(L)=\lim \inf _{n \rightarrow \infty} r_{n}^{1 / n}$. Thus outside the disc of radius $\rho_{\text {ess }}(L)+\varepsilon$ there are only finitely many eigenvalues and each eigenvalue has finite multiplicity. (A direct proof of this in the context of shifts of finite type is given in Parry and Pollicott.)

Proof. Step 1: The Schauder-Tychonov fixed point theorem. The Schauder-Tychonov theorem is a surprisingly general fixed point theorem in the context of convex sets that generalises the Brouwer fixed point theorem. It says the following: Let $\Lambda$ be a convex compact subset of a normed vector space $X$ and suppose that $L: \Lambda \rightarrow \Lambda$ is a continuous transformation. Then $L$ has a fixed point in $\Lambda$.

Step 2: Existence of $\lambda$ and $h$ Let

$$
\begin{gathered}
\Lambda=\left\{g \in C(\Sigma, \mathbb{R}) \mid 0 \leq g(x) \leq 1, g(x) \leq g(y) \exp \left(\frac{|f|_{\theta} \theta^{n}}{1-\theta}\right)\right. \\
\text { whenever } \left.x_{j}=y_{j}, j=0,1, \ldots, n-1\right\} .
\end{gathered}
$$

We claim that $\Lambda$ satisfies the hypotheses of the Schauder-Tychonov theorem.

It is easy to see that $\Lambda$ is convex and uniformly closed.
Suppose that $x, y \in \Sigma$ are such that $x_{j}=y_{j}, j=0,1, \ldots, n-1$. Then from the definition of $\Lambda$ it follows that

$$
\begin{aligned}
|g(x)-g(y)| & \leq|g(y)|\left(\exp \left(\frac{|f|_{\theta} \theta^{n}}{1-\theta}\right)-1\right) \\
& \leq|g|_{\infty} \frac{|f|_{\theta} \theta^{n}}{1-\theta} \exp \left(\frac{|f|_{\theta} \theta^{n}}{1-\theta}\right)
\end{aligned}
$$

Hence $\Lambda$ is uniformly equicontinuous.

As $\Lambda$ is uniformly closed and uniformly equicontinuous, it follows form the Arzelà-Ascoli theorem that $\Lambda$ is uniformly compact.

Also note from (8.3) that $\Lambda \subset F_{\theta}(\mathbb{R})$.
Define a family of linear operators $\Lambda \rightarrow \Lambda$ as follows. For $n \geq 1$ define

$$
L_{n} g(x)=\frac{L_{f}(g(x)+1 / n)}{\left|L_{f}(g(x)+1 / n)\right|_{\infty}}
$$

Clearly $\left|L_{n} g\right|_{\infty}=1$. Suppose $x, y \in \Sigma$ are such that $x_{j}=y_{j}$ for $0 \leq j \leq k$. Then

$$
L_{f}(g+1 / n)(x) \leq L_{f}(g+1 / n)(y) \exp \left(\frac{\theta^{k}}{1-\theta}|f|_{\theta}\right)
$$

In particular,

$$
L_{n} g(x) \leq L_{n} g(y) \exp \left(\frac{\theta^{k}}{1-\theta}|f|_{\theta}\right)
$$

so that $L_{n}$ is a well-defined operator $\Lambda \rightarrow \Lambda$, for each $n$.
Since $\Lambda$ is a convex uniformly compact subset of $C(\Sigma, \mathbb{R})$, we can apply the Schauder-Tychonov theorem to each $L_{n}: \Lambda \rightarrow \Lambda$. Hence, for each $n \geq 1$, there exists $h_{n} \in \Lambda$ with $L_{f}\left(h_{n}+1 / n\right)=\lambda_{n} h_{n}$, where $\lambda_{n}=\left|L_{f}\left(h_{n}+1 / n\right)\right|_{\infty}$.

As $\Lambda$ is uniformly compact, $h_{n}$ has a uniformly convergent subsequence with limit $h \in \Lambda$. As $\Lambda \subset F_{\theta}(\mathbb{R})$, we have $h \in F_{\theta}(\mathbb{R})$.

By continuity, $L_{f} h=\lambda h$ where $\lambda=\left|L_{f} h\right|_{\infty}$.

## Step 3: The eigenvalue $\lambda$ is positive

To see that $\lambda$ is positive, note that

$$
\begin{aligned}
\lambda_{n} h_{n}(x) & =\sum_{\sigma(y)=x} e^{f(y)}\left(h_{n}(y)+1 / n\right) \\
& \geq\left(\inf h_{n}+1 / n\right) e^{-|f|_{\infty}}
\end{aligned}
$$

Hence

$$
\lambda_{n} \inf h_{n} \geq\left(\inf h_{n}+1 / n\right) e^{-|f|_{\infty}}
$$

Hence $\lambda_{n} \geq e^{-|f|_{\infty}}$ for each $n \geq 1$. Hence $\lambda \geq e^{-|f|_{\infty}}>0$.
Step 4: The eigenfunction $h$ can be taken to be strictly positive.
As $h_{n} \in \Lambda$ it follows from the definition of $\Lambda$ that $h_{n}(x) \geq 0$. Hence $h(x) \geq 0$ for all $x \in \Sigma$. Suppose for a contradiction that there exists $x_{0}$ for which $h\left(x_{0}\right)=0$. Iterating the eigenvalue equation $L_{f} h=\lambda h$ gives that

$$
\sum_{y: \sigma^{n} y=x_{0}} e^{f^{n} y} h(y)=\lambda^{n} h\left(x_{0}\right)=0
$$

In particular, $h(y)=0$ whenever $\sigma^{n} y=x_{0}$. As the set of such $y$ is dense in $\Sigma$, by the aperiodicity of $A$, it follows that $h$ is identically zero, a contradiction.

Step 5: $\lambda$ is simple. We know that $L_{f} h=\lambda h$. Suppose that $g$ is another continuous eigenfunction for $L_{f}$ corresponding to the eigenvalue $\lambda$. Let $t=\inf g(x) / h(x)$. By compactness, this infimum is achieved at some point: $t=g\left(x_{0}\right) / h\left(x_{0}\right)$, say. Then $g\left(x_{0}\right)-t h\left(x_{0}\right)=0$. Repeating the argument from Step 4 shows that $g(y)-t h(y)=0$ whenever $y \in \Sigma$ is such that $\sigma^{n} y=x$. Again, by aperiodicity the set of such $y$ is dense, hence $g(x)-t h(x)=0$ for all $x$, i.e. $g$ is a scalar multiple of $h$. Hence the eigenspace corresponding to $\lambda$ is one-dimensional.

Step 6: Reduction to the normalised case. Let $h, \lambda$ be as above, so that $L_{f} h=\lambda h$ and $h>0$. Define

$$
g=f-\log h \sigma+\log h-\log \lambda
$$

Then

$$
\begin{aligned}
L_{g} w(x) & =\sum_{y: \sigma(y)=x} e^{g(y)} w(y) \\
& =\frac{1}{\lambda} \sum_{y: \sigma(y)=x} e^{f(y)} \frac{h(y)}{h(\sigma(y))} w(y) \\
& =\frac{1}{\lambda} \frac{1}{h(x)} \sum_{y: \sigma(y)=x} e^{f(y)} h(y) w(y)
\end{aligned}
$$

Hence if we let $M_{h}$ denote the linear operator that multiplies a function by $h$ (i.e. $\left.\left(M_{h} w\right)(x)=h(x) w(x)\right)$ then

$$
\begin{equation*}
L_{g}=\lambda^{-1} M_{h}^{-1} L_{f} M_{h} \tag{8.3}
\end{equation*}
$$

As $L_{f} h=\lambda h$, it follows from (8.3) that $L_{g} 1=1$, i.e. $g$ is normalised.
Since the spectrum of $L_{f}$ is the spectrum of $L_{g}$ scaled by a factor of $1 / \lambda$, it is sufficient to prove the remainder of the theorem under the hypothesis that $L_{f} 1=1$.

## Step 7: Existence of $\nu$.

The operator $L_{f}^{*}$ acts on $C(\Sigma, \mathbb{R})^{*}$ and preserves the convex weak-* compact subset of functionals that correspond to $\sigma$-invariant probability measures. By the Schauder-Tychonov theorem, $L_{f}^{*}$ has a fixed point $\nu$.

## Step 8: Uniqueness of $\nu$.

Note that

$$
\operatorname{var}_{k} L_{f}^{n} w \leq\left|L_{f}^{n} w\right|_{\theta} \theta^{k} \leq C \theta^{k}|w|_{\infty}+\theta^{n+k}|w|_{\theta}
$$

by Proposition 8.4. Hence, for fixed $w \in F_{\theta}(\mathbb{R})$, the set $\left\{L_{f}^{n} w\right\}_{n=1}^{\infty}$ is a uniformly equicontinuous subset of $C(\Sigma, \mathbb{R})$ and so has a convergent subsequence, $L_{f}^{n_{k}} w \rightarrow w^{*}$ uniformly. We claim that $w^{*}$ is constant.

To see that $w^{*}$ is constant note that, as $L_{f}$ is a convex combination of preimages, we have that $\sup w \geq \sup L_{f} w \geq \cdots$. Hence $\sup L_{f}^{k_{n}} w^{*}=\sup w^{*}$. Choose $x_{n_{k}} \in \Sigma$ such that $L_{f}^{n_{k}} w^{*}\left(x_{n_{k}}\right)=\sup w^{*}$ (so that, in particular, $\left.w^{*}\left(x_{0}\right)=\sup w^{*}\right)$. Then

$$
L_{f}^{n_{k}} w^{*}\left(x_{n_{k}}\right)=\sum_{\sigma^{n} k y=x_{n_{k}}} e^{f^{n_{k}}(y)} w^{*}(y)=w^{*}\left(x_{0}\right) .
$$

This is a convex combination of the points $w^{*}(y)$. Hence $w^{*}(y)=w^{*}\left(x_{0}\right)$ whenever $\sigma^{n_{k}}(y)=x_{n_{k}}$. As the set of such $y$ is dense, it follows that $w^{*}$ is constant.

To see that $\nu$ is unique, note that

$$
w^{*}=\int w^{*} d \nu=\lim _{k \rightarrow \infty} \int L_{f}^{n_{k}} w d \nu=\int w d \nu
$$

We can repeat this argument through any subsequence to see that $L_{f}^{n} w \rightarrow$ $\int w d \nu$ for all $w \in F_{\theta}(\mathbb{R})$. By approximation, this is also true for all $w \in C(\Sigma, \mathbb{R})$. Hence by the Riesz Representation Theorem, $\nu$ is uniquely determined by the condition that $L_{f}^{*} \nu=\nu$.

## Step 9: Estimation of the remainder of the spectrum.

We have seen in Step 8 that if $w \in F_{\theta}(\mathbb{R})$ then $L_{f}^{n} w \rightarrow \int w d \nu$. Thus the constant functions are eigenfunctions with eigenvalue 1. To show that the remainder of the spectrum of $L_{f}: F_{\theta}(\mathbb{R}) \rightarrow F_{\theta}(\mathbb{R})$ lies in a disc of radius strictly less than 1 it is sufficient to prove that $L_{f}$, acting on the space $\mathbb{C}^{\perp}=\left\{w \in F_{\theta}(\mathbb{R}) \mid \int w d \nu=0\right\}$, has spectral radius strictly less than 1.

By Proposition 8.4, we have

$$
\left|L_{f}^{n+k} w\right|_{\theta} \leq C\left|L_{f}^{k} w\right|_{\infty}+\theta^{n}\left|L_{f}^{k} w\right|_{\theta} \leq C\left|L_{f}^{k} w\right|_{\infty}+C \theta^{n}|w|_{\infty}+\theta^{n+k}|w|_{\theta}
$$

Moreover, by Step 8 , as $w \in \mathbb{C}^{\perp}$ we have that $L_{f}^{n} w \rightarrow 0$ on the uniformly compact set $\left\{w \in \mathbb{C}^{\perp} \mid\|w\|_{\theta}<1\right\}$. Fix a choice of $\varepsilon>0$. Then $\left\|L_{f}^{k} w\right\|_{\theta}<\varepsilon$ provided $k$ is sufficiently large. The spectral radius formula tells us that the spectral radius of $L_{f}$ on $\mathbb{C}^{\perp}$ is bounded above by

$$
\inf \left\{\left\|L_{f}^{n+k} w\right\|_{\theta}^{1 /(n+k)} \mid w \in C^{\perp},\|w\|_{\theta} \leq 1\right\} \leq \varepsilon^{1 /(n+k)}
$$

The claim follows.

## $\S 8.7$ Decay of correlations

Let $\mu$ be the equilibrium state corresponding to a Hölder potential $u \in$ $F_{\theta}(\mathbb{R})$. We can assume that $u$ is normalised so that $L_{u} 1=1$. Then $L_{u}^{*} \mu=\mu$.

We have already seen that $\mu$ is an invariant measure, although we recall the proof here. Let $w \in C(\Sigma, \mathbb{R})$ be continuous. Then

$$
\begin{aligned}
\int w \sigma d \mu & =\int L_{u}(w \sigma) d \mu \\
& =\int \sum_{y: \sigma y=x} e^{u(y)} w(\sigma y) d \mu \\
& =\int w(x) L_{u} 1 d \mu \\
& =\int w d \mu
\end{aligned}
$$

We will show that $\mu$ is strong-mixing. Indeed, we will show that $\mu$ satisfies the following stronger property.

Definition. Let $\mathcal{F}$ be class of functions equipped with a norm $\|\cdot\|$. We say that a dynamical system $T$ with invariant measure $\mu$ has exponential decay of correlations on $\mathcal{F}$ if there exist constants $C>0$ and $\rho \in(0,1)$ such that

$$
\left|\int f \sigma^{n} \cdot g d \mu-\int f d \mu \int g d \mu\right| \leq C\|f\|\|g\| \rho^{n}
$$

Recall that a measure-preserving transformation $T$ on a probability space ( $X, \mathcal{B}, \mu$ ) is strong-mixing if for all $A, B \in \mathcal{B}$ we have

$$
\mu\left(T^{-n} A \cap B\right) \rightarrow \mu(A) \mu(B)
$$

as $n \rightarrow \infty$. By the definition of the Lebesgue integral, this is easily seen to be equivalent to

$$
\int f T^{n} \cdot g d \mu \rightarrow \int f d \mu \int g d \mu
$$

as $n \rightarrow \infty$, for all $f, g \in L^{2}(X, \mathcal{B}, \mu)$.
If $\mathcal{F}$ is $L^{2}$-dense in $L^{2}(X, \mathcal{B}, \mu)$ and $T$ has exponential decay of correlations on $\mathcal{F}$ with respect to $\mu$, then it follows by approximation that $T$ is strong-mixing with respect to $\mu$. In particular, $T$ is ergodic with respect to $\mu$.

## Theorem 8.6

Let $\mu$ be the equilibrium state corresponding to a Hölder potential $u \in$ $F_{\theta}(\mathbb{R})$. Then $\sigma$ has exponential decay of correlations on $F_{\theta}(\mathbb{R})$ with respect to $\mu$.

Proof. Let $f, g \in F_{\theta}(\mathbb{R})$.
Write $L_{u}=\mu+Q$ where $Q$ denotes the projection onto $\left\{w \in F_{\theta}(\mathbb{C}) \mid\right.$ $\left.\int w d \mu=0\right\}$. Then $Q$ has spectral radius at most $\rho$, for some $\rho \in(0,1)$.

Note that $L_{u}^{n}=\nu+Q^{n}$. Then, as $\mu(g)=\mu\left(g \sigma^{n}\right)$, we have

$$
\begin{aligned}
\left|\int f \sigma^{n} \cdot g d \mu-\mu(f) \mu(g)\right| & =\left|\int f \sigma^{n}(g-\mu(g)) d \mu\right| \\
& =\left|\int L_{u}^{n}\left(f \sigma^{n}(g-\mu(g))\right) d \mu\right| \\
& =\left|\int f L_{u}^{n}(g-\mu(g)) d \mu\right| \\
& =\left|\int f Q^{n} g d \mu\right| \\
& \leq C\|f\|_{\infty}\|g\|_{\theta} \rho^{n}
\end{aligned}
$$

and the result follows.

## §8.8 Gibbs measures and equilibrium states

We are interested in characterising the measures that appear as eigenmeasures for $L_{f}$. Recall that $f^{n}(x)=\sum_{j=0}^{n-1} f\left(\sigma^{j} x\right)$. The following definition does not require $f$ to be Hölder.

Definition. Let $f: \Sigma \rightarrow \mathbb{R}$ be continuous. A probability measure $m$ is called a Gibbs measure with potential $f$ if there exist constants $P \in \mathbb{R}$ and $C_{1}, C_{2}>0$ such that

$$
\begin{equation*}
C_{1} \leq \frac{m\left[x_{0}, \ldots, x_{n-1}\right]}{e^{f^{n}(x)-n P}} \leq C_{2} \tag{8.4}
\end{equation*}
$$

where $x \in \Sigma, x=\left(x_{0}, x_{1}, \ldots\right)$.

## Remarks.

(i) We do not necessarily assume that $m$ is $\sigma$-invariant.
(ii) Suppose that $m$ is a Gibbs measure for the potential $f$ with constants $C_{1}, C_{2}, P$. If $d m_{1}=h d m$ where $h$ is uniformly bounded away from 0 and $\infty$, then $m_{1}$ is also a Gibbs measure for the potential $f$; the constant $P$ will remain the same but $C_{1}, C_{2}$ could change.
(iii) The condition (8.4) says the following: given $x \in \Sigma$, the measure of the cylinder of length $n$ that contains $x$ can be approximated by the (exponential) of the ergodic sum of the potential function, subject to some correction $P$.
(iv) Not all constants have equal status! The exact values of the constants $C_{1}, C_{2}$ in (8.4) are unimportant, as long as we know that they are positive. However, the constant $P$ is an important quantity that we shall study further in this lecture and the next.

Given a function $f \in F_{\theta}(\mathbb{R})$ we want to prove the existence of a Gibbs measure. We start in the case when $f$ is normalised. By Theorem 8.5 we know that there exists a unique probability measure $m$ such that $L_{f}^{*} m=m$. Moreover, $m$ is $\sigma$-invariant as

$$
\int w \sigma d m=\int L_{f}(w \sigma) d m=\int w L_{f} 1 d m=\int w d m
$$

## Lemma 8.7

Let $f \in F_{\theta}(\mathbb{R})$ be normalised. Then for each $x=\left(x_{0}, x_{1}, \ldots\right) \in \Sigma$ we have

$$
\begin{equation*}
e^{-|f|_{\theta} \theta^{n}} \leq \frac{m\left[x_{0}, \ldots, x_{n}\right] e^{-f(x)}}{m\left[x_{1}, \ldots, x_{n}\right]} \leq e^{|f|_{\theta} \theta^{n}} \tag{8.5}
\end{equation*}
$$

In particular, $m$ is a Gibbs measure for $f$ with $P=0$.
Proof. As $m$ is $\sigma$-invariant and $L_{f}^{*} m=m$ we have that

$$
\begin{aligned}
m\left[x_{1}, \ldots, x_{n}\right] & =m\left(\bigcup_{x_{0}}\left[x_{0}, x_{1}, \ldots, x_{n}\right]\right) \\
& =\int \sum_{x_{0}} \chi_{\left[x_{0}, x_{1}, \ldots, x_{n}\right]} d m \\
& =\int \sum_{y: \sigma y=x} e^{f(y)} \chi_{\left[x_{0}, x_{1}, \ldots, x_{n}\right]} e^{-f(y)} d m \\
& =\int L_{f}\left(\chi_{\left[x_{0}, x_{1}, \ldots, x_{n}\right]} e^{-f(y)}\right) d m \\
& =\int \chi_{\left[x_{0}, x_{1}, \ldots, x_{n}\right]} d m \\
& =\iint_{\left[x_{0}, x_{1}, \ldots, x_{n}\right]} e^{-f} d m
\end{aligned}
$$

As $f \in F_{\theta}(\mathbb{R})$ it follows that if $z, w \in\left[x_{0}, \ldots, x_{n}\right]$ (so that $z_{j}=w_{j}$ for $j=0, \ldots, n-1)$. Then

$$
e^{-|f|_{\theta} \theta^{n}} \leq e^{f(z)-f(w)} \leq e^{|f|_{\theta} \theta^{n}}
$$

Hence (8.5) follows.
To see that $m$ is a Gibbs measure with $P=0$ note that the above gives a sequence of inequalities

$$
\begin{aligned}
e^{-|f|_{\theta} \theta^{n}} & \leq \frac{m\left[x_{0}, \ldots, x_{n}\right]}{m\left[x_{1}, \ldots, x_{n}\right]} e^{-f(x)}
\end{aligned} \leq e^{|f|_{\theta} \theta^{n}}
$$

Multiplying the above inequalities together gives that

$$
e^{-|f|_{\theta}(1-\theta)^{-1}} \leq \frac{m\left[x_{0}, \ldots, x_{n}\right]}{e^{f^{n+1}(x)}} \leq e^{|f|_{\theta}(1-\theta)^{-1}}
$$

Hence $m$ is a Gibbs measure for $f$.
Let $f \in F_{\theta}(\mathbb{R})$. By normalising $f$ we can apply Lemma 8.7 to obtain the following result.

## Corollary 8.8

Let $f \in F_{\theta}(\mathbb{R})$. Let $\lambda>0$ be the maximal eigenvalue of $L_{f}$. Then $m$ is a Gibbs measure for $f$ with $P=\log \lambda$.

Proof. Let $h$ be as in Theorem 8.5. Then $g=f-\log h \sigma+\log h-\log \lambda$ is normalised. By applying Lemma 8.7 to $g$, it follows that there exist constants $C_{1}^{\prime}, C_{2}^{\prime}$ such that

$$
C_{1}^{\prime} \leq \frac{m\left[x_{0}, \ldots, x_{n}\right]}{e^{f^{n}(x)-n \log \lambda}} \leq C_{2}^{\prime}
$$

## §8.9 Information and entropy of Gibbs measures

Let $\beta=\{[1],[2], \ldots,[k]\}$ denote the state partition of $\Sigma$ into cylinders of length 1. Then $\beta$ is a strong generator for $\sigma$. Recall that this means that $\bigvee_{j=0}^{\infty} \sigma^{-j} \beta=\mathcal{B}$ (in the sense that the smallest $\sigma$-algebra that contains all the sets in each $\bigvee_{j=0}^{n-1} \sigma^{-j} \beta$ is $\mathcal{B}$ ).

Note that

$$
\sigma_{j=1}^{\infty} \sigma^{-j} \beta=\sigma^{-1} \mathcal{B}
$$

We recall the definition of information and entropy from Lecture 7. Let $T$ be a measure-preserving transformation of a probability space $(X, \mathcal{B}, \mu)$. Let $\alpha$ be a finite or countable partition of $X$ into measurable sets. Let $\mathcal{A}$ be a sub- $\sigma$-algebra. We define the conditional information function of $\alpha$ by

$$
I_{\mu}(\alpha \mid \mathcal{A})(x)=-\sum_{A \in \alpha} \chi_{A}(x) \log \mu(A \mid \mathcal{A})(x)
$$

where $\mu(A \mid \mathcal{A})=E_{\mu}\left(\chi_{A} \mid \mathcal{A}\right)$ denotes the conditional probability. Note that we are emphasising the dependence of these quantities on the measure $\mu$ as we shall be considering varying $\mu$ in the sequel.

The conditional entropy of $\alpha$ is then given by

$$
\begin{aligned}
& H_{\mu}(\alpha, \mathcal{A})=-\int I_{\mu}(\alpha \mid \mathcal{A}) d \mu \\
& \quad=-\sum_{A \in \mathcal{A}} \int_{A} \chi_{A} \log \mu(\alpha \mid \mathcal{A}) d \mu=-\sum_{A \in \mathcal{A}} \int_{A} \mu(\alpha \mid \mathcal{A}) \log \mu(\alpha \mid \mathcal{A})
\end{aligned}
$$

The entropy of $T$ relative to $\alpha$ is then defined to be

$$
h_{\mu}(T, \alpha)=H_{\mu}\left(\alpha \mid \bigvee_{j=1}^{\infty} T^{-j} \alpha\right)
$$

We return to the context of shifts of finite type. We know that the state partition $\beta$ is a strong generator. By Sinai's theorem, it follows that

$$
h_{\mu}(\sigma)=h_{\mu}(\sigma, \beta)=H_{\mu}\left(\beta \mid \sigma^{-1} \mathcal{B}\right)=\int I_{\mu}\left(\beta \mid \sigma^{-1} \mathcal{B}\right) d \mu
$$

Let $f \in F_{\theta}(\mathbb{R})$ be normalised and suppose that $L_{f}^{*} m=m$. We can easily calculate the information function and entropy of $m$.

## Proposition 8.9

Let $f \in F_{\theta}(\mathbb{R})$ be normalised and suppose that $L_{f}^{*} m=m$. Then $I_{m}(\beta \mid$ $\left.\sigma^{-1} \mathcal{B}\right)=-f(x)$. Moreover, $h_{m}(\sigma)=-\int f d m$.

Proof. By the Increasing Martingale Theorem we have that

$$
I_{m}\left(\beta \mid \sigma^{-1} \mathcal{B}\right)=\lim _{n \rightarrow \infty} I_{m}\left(\beta \mid \bigvee_{j=1}^{n} \sigma^{-j} \mathcal{B}\right)
$$

both $m$-a.e. and in $L^{1}$.
A straightforward calculation from the definitions shows that

$$
I_{m}\left(\beta \mid \bigvee_{j=1}^{n} \sigma^{-j} \mathcal{B}\right)=-\sum_{x_{0}, x_{1}, \ldots x_{n}} \chi_{\left[x_{0}, x_{1}, \ldots, x_{n}\right]}(x) \log \frac{m\left[x_{0}, x_{1}, \ldots, x_{n}\right]}{m\left[x_{1}, \ldots, x_{n}\right]}
$$

It follows from the proof of Lemma 8.7 that

$$
\frac{m\left[x_{0}, x_{1}, \ldots, x_{n}\right]}{m\left[x_{1}, \ldots, x_{n}\right]} \rightarrow e^{-f(x)}
$$

Hence $I_{m}\left(\beta \mid \sigma^{-1} \mathcal{B}\right)=-f(x)$ and $h_{m}(\sigma)=-\int f d m$.

## §8.10 The variational principle and equilbrium states

It will be useful to see how invariant Gibbs measures are distinguished amongst all $\sigma$-invariant probability measures. This will lead to a characterisation in terms of the so-called variational principle.

A useful trick in the proof is to consider a family of probability distributions on the finite set $\{1, \ldots, k\}$. Given a $\sigma$-invariant probability measure
$\mu$, we have that $\mu\left[x_{0}, \ldots, x_{n}\right]>0$ for $\mu$-a.e. $x \in \Sigma$. Define a probability distribution on $\{1, \ldots, k\}$ by

$$
\mu_{n, x}(i)=\frac{\mu\left[i, x_{0}, \ldots, x_{n-1}\right]}{\mu\left[x_{0}, \ldots, x_{n-1}\right]}=\mu\left([i] \mid \bigvee_{j=1}^{n} \sigma^{-j} \beta\right)
$$

By the Increasing Martingale Theorem, we can let $n \rightarrow \infty$ and obtain a probability distribution on $\{1, \ldots, k\}$ by

$$
\mu_{x}(i)=\lim _{n \rightarrow \infty} \mu_{n, x}(i)
$$

Note that

$$
I_{\mu}\left(\beta \mid \sigma^{-1} \mathcal{B}\right)=-\sum_{i} \chi_{[i]} \log \mu_{x}(i)
$$

We begin with some preparatory results.

## Lemma 8.10

Let $g \in C(\Sigma, \mathbb{R})$ be a continuous function. Then for $\mu$-a.e. $x$ we have

$$
\sum_{i=1}^{k} \int g\left(i, x_{0}, x_{1}, \ldots\right) \mu_{x}(i) d \mu=\int g d \mu
$$

Proof. It is sufficient to prove this result in the case when $g=\chi_{\left[j 0, \ldots, j_{]}\right]}$; the general case follows by approximation. Note that

$$
\begin{aligned}
\sum_{i=1}^{k} \int g\left(i, x_{0}, x_{1}, \ldots\right) \mu_{x}(i) d \mu & =\lim _{n \rightarrow \infty} \sum_{i=1}^{k} \int \chi_{\left[j_{0}, \ldots, j_{\ell}\right]} \frac{\mu\left[i, x_{0}, \ldots, x_{n-1}\right]}{\mu\left[x_{0}, \ldots, x_{n-1}\right]} d \mu \\
& =\mu\left[j_{0}, \ldots, j_{\ell}\right] \sum_{i=1}^{k} \frac{\mu\left[i, x_{0}, \ldots, x_{n-1}\right]}{\mu\left[x_{0}, \ldots, x_{n-1}\right]} \\
& =1
\end{aligned}
$$

## Lemma 8.11

Let $p=\left(p_{1}, \ldots, p_{k}\right), q=\left(q_{1}, \ldots, q_{k}\right)$ be two probability vectors with $p_{j}>0$ for $1 \leq j \leq k$. Then

$$
\begin{equation*}
-\sum_{j=1}^{k} q_{j} \log q_{j}+\sum_{j=1}^{k} q_{j} \log p_{j} \leq 0 \tag{8.6}
\end{equation*}
$$

with equality if and only if $p_{j}=q_{j}$ for $1 \leq j \leq k$.

Proof. The left-hand side of (8.6) can be written as

$$
\begin{equation*}
\sum_{j=1}^{k}-p_{j} \frac{q_{j}}{p_{j}} \log \frac{q_{j}}{p_{j}}=\sum_{j=1}^{k} p_{j} \phi\left(\frac{q_{j}}{p_{j}}\right) \tag{8.7}
\end{equation*}
$$

where $\phi(x)=-x \log x$ (with the usual convention that $0 \times \log 0=0$ ). Note that, as $\phi$ is concave, we have that (8.7) is less than or equal to

$$
\phi\left(\sum_{j=1}^{k} p_{j} \frac{q_{j}}{p_{j}}\right)=\phi(1)=0
$$

with equality if and only if all the $q_{j} / p_{j}$ are equal.
The following gives a preliminary characterisation.

## Proposition 8.12

Let $f \in F_{\theta}(\mathbb{R})$ be normalised and let $L_{f}^{*} m=m$. Then for any $\sigma$-invariant probability measure $\mu$ we have

$$
h_{\mu}(\sigma)+\int f d \mu \leq 0
$$

with equality if and only if $\mu=m$.
Proof. Let $\mu$ be any $\sigma$-invariant probability measure. We define a probability distribution on $\{1, \ldots, k\}$ by $\mu_{x}(i)$, as above. In the case when $\mu=m$, we have the probability distribution $m_{x}(i)=e^{f\left(i, x_{0}, x_{1}, \ldots\right)}$.

By Lemma 8.11, we have for $\mu$-a.e. $x$,

$$
\begin{aligned}
& -\sum_{i=1}^{k} \mu_{x}(i) \log \mu_{x}(i)+\sum_{i=1}^{k} \mu_{x}(i) f\left(i, x_{0}, x_{1}, \ldots\right) \\
& \quad=-\sum_{i=1}^{k} \mu_{x}(i) \log \mu_{x}(i)+\sum_{i=1}^{k} \mu_{x}(i) \log m_{x}(i) \\
& \quad \leq 0
\end{aligned}
$$

with equality for $\mu$-a.e. $x \in \Sigma$ if and only if $\mu_{x}(i)=\exp f\left(i, x_{0}, x_{1}, \ldots\right)$. Integrating this expression with respect to $\mu$ we obtain, by Lemma 8.10,

$$
h_{\mu}(\sigma)+\int f d \mu \leq 0
$$

with equality if and only if $\mu_{x}(i)=\exp f\left(i, x_{0}, x_{1}, \ldots\right)$.
Note that $\mu_{x}(i)=\exp f\left(i, x_{0}, x_{1}, \ldots\right)$ implies that

$$
\int \sum_{i=1}^{k} e^{f\left(i, x_{0}, x_{1}, \ldots\right)} g\left(i, x_{0}, x_{1}, \ldots\right) d \mu=\int g d \mu
$$

for all $g \in C(\Sigma, \mathbb{R})$, i.e. $\int L_{f} g d \mu=\int g d \mu$. This is equivalent to $L_{f}^{*} \mu=\mu$. As $m$ is the unique $\sigma$-invariant probability measure that is fixed by $L_{f}^{*}$, the proposition follows.

We can now give the following characterisation of Gibbs measures.

## Theorem 8.13 (The variational principle)

Let $f \in F_{\theta}(\mathbb{R})$. Then there exists a unique invariant probability measure $\mu_{f}$ such that

$$
\begin{equation*}
P(f)=\sup \left\{h_{\mu}(\sigma)+\int f d \mu\right\}=h_{\mu_{f}}(\sigma)+\int f d \mu_{f} \tag{8.8}
\end{equation*}
$$

where the supremum is taken over all $\sigma$-invariant probability measures. The measure $\mu_{f}$ is a Gibbs measure and is given by $d \mu_{f}=h d m$ where $h$ and $m$ are the eigenfunction and probability eigenmeasure for $L_{f}$ corresponding to the maximal eigenvalue $e^{P(f)}$ for $L_{f}$ and $\int h d m=0$. Moreover, $\mu_{f}$ is the unique $\sigma$-invariant Gibbs measure for the potential $f$.

Proof. Let $f \in F_{\theta}(\mathbb{R})$. With $h, \lambda$ as in Theorem 8.5 and $\lambda=e^{P(f)}$, the function $g=f-\log h \sigma+\log h-P(f)$ is normalised. Let $\mu$ be any $\sigma$-invariant probability measure. Then

$$
h_{\mu}(\sigma)+\int g d \mu=h_{\mu}(\sigma)+\int f d \mu+\int-\log h \sigma+\log h d \mu-P(f) \leq 0
$$

with equality if and only if $\mu$ is the eigenmeasure for $L_{g}$, i.e. $d \mu=h d m$ where $L_{f}^{*} m=\lambda m$.

Definition. A measure $\mu$ that achieves the supremum in (8.8) is called an equilibrium state.

Remark Thus the variational principle says that a function $f \in F_{\theta}(\mathbb{R})$ has a unique equilibrium state. There are examples of continuous, but not Hölder, functions which possess more than one equilibrium state.

## §8.11 Pressure

Recall we defined the pressure $P(f)$ of $f \in F_{\theta}(\mathbb{R})$ by $P(f)=\log \lambda$ where $\lambda$ is the maximal eigenvalue for the transfer operator $L_{f}$. We can regard pressure as a functional $F_{\theta}(\mathbb{R}) \rightarrow \mathbb{R}$. By using the variational principle, one can prove the following properties of this functional.

## Theorem 8.14

(i) Pressure is monotone: if $f, g \in F_{\theta}(\mathbb{R})$ and $f \leq g$ then $P(f) \leq P(g)$.
(ii) Pressure is convex: if $f, g \in F_{\theta}(\mathbb{R})$ and $\alpha \in[0,1]$ then $P(\alpha f+(1-$ $\alpha) g) \leq \alpha P(f)+(1-\alpha) P(g)$.
(iii) If $f$ is cohomologous to $g+c$, where $f, g \in F_{\theta}(\mathbb{R})$ and $c \in \mathbb{R}$ then $P(f)=P(g)+c$.
(iv) The function $f \in F_{\theta}(\mathbb{R})$ is normalised if and only if $P(f)=0$.

Proof. Throughout, let $f, g \in F_{\theta}(\mathbb{R})$.
If $f \leq g$ then

$$
P(f)=\sup \left\{h_{\mu}(\sigma)+\int f d \mu\right\} \leq \sup \left\{h_{\mu}(\sigma)+\int g, d \mu\right\}=P(g)
$$

where the suprema are taken over all $\sigma$-invariant probability measures; hence (i) holds.

Statement (ii) follows by noting that, for $\alpha \in[0,1]$,

$$
\begin{aligned}
P(\alpha f+(1-\alpha) g)= & \sup \left\{h_{\mu}(\sigma)+\int \alpha f+(1-\alpha) g d \mu\right\} \\
\leq & \sup \left\{\alpha\left(h_{\mu}(\sigma)+\int f d \mu\right)\right\} \\
& +\sup \left\{(1-\alpha)\left(h_{\mu}(\sigma)+\int g d \mu\right)\right\} \\
= & \alpha P(f)+(1-\alpha) P(g)
\end{aligned}
$$

where the suprema are taken over all $\sigma$-invariant probability measures.
Statements (iii) and (iv) follow immediately from the variational principle.

As well as the variational principle, there are other characterisations of pressure.

## Proposition 8.15

Let $f \in F_{\theta}(\mathbb{R})$. Then

$$
\begin{equation*}
P(f)=\lim _{n \rightarrow \infty} \frac{1}{n} \log \sum_{\left[i_{0}, \ldots, i_{n-1}\right]} \sup _{x \in\left[i_{0}, \ldots, i_{n-1}\right]} \exp f^{n}(x) \tag{8.9}
\end{equation*}
$$

where the sum is taken over all allowable cylinders of length $n$ in $\Sigma$.
Remark In fact one can replace the supremum over $x \in\left[i_{0}, \ldots, i_{n-1}\right]$ in (8.9) by the infimum, or indeed by any arbitrary choice of point in each cylinder. This is because, as $f \in F_{\theta}(\mathbb{R})$, if $x, y \in\left[i_{0}, \ldots, i_{n-1}\right]$ then $\mid f^{n}(x)-$ $\left.f^{n}(y)\left|\leq \theta(1-\theta)^{-1}\right| f\right|_{\theta}$, a constant independent of $n$.

Proposition 8.16
Let $f \in F_{\theta}(\mathbb{R})$. Then

$$
P(f)=\lim _{n \rightarrow \infty} \frac{1}{n} \log \sum_{\sigma^{n} x=x} \exp f^{n}(x)
$$

where the sum is over all $\sigma$-periodic points of period $n$.
Remark For the proof of both of these propositions see Parry and Pollicott.

## §8.12 References

The exposition above largely follows that in the first three chapters of
W. Parry and M. Pollicott, Zeta functions and the periodic orbit structure of hyperbolic dynamics, Astérisque, 187-188, Société Mathématique de France, 1990,
which in turn is an account of the material in
R. Bowen, Equilibrium states and the ergodic theory of Anosov diffeomorphisms, Lecture Notes in Mathematics 470, Springer, Berlin 1975,
D. Ruelle, Thermodynamic formalism, Addison-Wesley, 1978.

Ruelle's presentation is geared more towards statistical mechanics than to ergodic theory, at least in the way that we have been developing the subject. Another account of this material (with an emphasis on Markov maps of the interval) is given in
V. Baladi, Positive transfer operators and decay of correlations, World Scientific, 2000.

Hennion's theorem (which gives a precise description of the spectrum and essential spectrum of a linear operator satisfying a Lasota-Yorke type inequality together with a compact inclusion from one Banach space to another) is proved in
H. Hennion, Sur un théorème spectral et son application aux noyaux lipschitziens, Proc. Amer. Math. Soc., 118 (1993), 627-634.

There is a version of the variational principle for arbitrary continuous transformations of compact metric spaces due to Walters (building on work of Ruelle). An expository account is given in
P. Walters, An introduction to ergodic theory, Springer, Berlin, 1982.

The standard text on perturbation theory is
T. Kato, Perturbation theory of linear operators, Springer-Verlag, Berlin 1966.

## $\S 8.13$ Exercises

## Exercise 8.1

Show that $F_{\theta}$ is a Banach space with respect to the norm $\|\cdot\|_{\theta}$. Show that the unit ball (with respect to the $|\cdot|_{\infty}$-norm) has compact closure with respect to the $\|\cdot\|_{\theta}$-norm. (Use the Arzelà-Ascoli theorem.)

## Exercise 8.2

Show that $f, g \in F_{\theta}(\mathbb{R})$ have the same equilibrium state if and only if $f$ if cohomologous to $g+c$, for some constant $c$.

