8. Topological entropy

§8.1 Introduction

In the previous lecture we defined the entropy of a measure-preserving transformations on a probability space. In this lecture we study entropy in the context of continuous transformations of compact metric spaces. In particular, given a continuous transformation \( T : X \to X \) of a compact metric space \( X \), we can study the map \( \mu \mapsto h_\mu(T) : \mathcal{M}(X,T) \to \mathbb{R} \). For a wide class of dynamical system this entropy function is bounded and attains its supremum; we obtain ‘measures of maximal entropy’.

Entropy as we’ve defined it so far is a purely measure-theoretic construction. Indeed, one often refers to \( h_\mu(T) \) as the metric entropy of \( T \); here ‘metric’ is a contraction of ‘measure-theoretic’, it does not refer to a metric or topological structure on the probability space \( X \). There is a related concept of topological entropy. It turns out that topological entropy and measure-theoretic entropy are very closely related by a variational principle.

§8.2 Recap on entropy

Let \((X, \mathcal{B}, \mu)\) be a probability space. Let \( T : X \to X \) be a measure-preserving transformation.

In the previous lecture we used \( \alpha, \beta, \ldots \) to denote finite or countable partitions of \( X \). In this lecture we will use \( \zeta, \eta \).

Let \( \zeta = \{A_i\}, A_i \in \mathcal{B} \) be a finite or countable partition of \( X \). We define

\[
H_\mu(\zeta) = -\sum_{A \in \zeta} \mu(A) \log \mu(A)
\]

to be the entropy of \( \zeta \).

If \( \zeta \) and \( \eta \) are two partitions of \( X \) then we let \( \zeta \lor \eta \), the join of \( \zeta \) and \( \eta \), denote the partition \( \{A \cap B \mid A \in \zeta, B \in \eta\} \).

Let \( T^{-1}\zeta = \{T^{-1}A \mid A \in \zeta\} \). We define

\[
h_\mu(T, \zeta) = \lim_{n \to \infty} \frac{1}{n} H_\mu \left( \bigvee_{j=0}^{n-1} T^{-j}\zeta \right)
\]

to be the entropy of \( T \) relative to \( \zeta \); the limit exists by subadditivity.

Finally, we define

\[
h_\mu(T) = \sup h_\mu(T, \zeta) \tag{8.1}
\]
where the supremum is taken over all finite or countable partitions $\zeta$ with $H_\mu(\zeta) < \infty$. Indeed, although we will not prove it, it is sufficient to take the supremum just over finite partitions. We call $h_\mu(T)$ the entropy of $T$ with respect to $\mu$; if $T$ is fixed then we sometimes refer to $h_\mu(T)$ as the entropy of $\mu$.

Certain partitions $\zeta$ achieve the supremum in (8.1). A partition $\zeta$ is a generator if

$$\bigvee_{j=-n}^n T^{-j} \zeta \not\in \mathcal{B}$$

as $n \to \infty$. Equivalently, $\zeta$ is a generator if $\bigvee_{j=-n}^n T^{-j} \zeta$ separates $\mu$-almost every pair of points: for almost every $x, y \in X$, $x \neq y$, there exists $n$ such that $x$ and $y$ are in different elements of the partition $\bigvee_{j=-n}^n T^{-j} \zeta$. In this case, we have the following theorem:

**Theorem 8.1 (Sinai’s theorem)**

Let $T$ be a measure-preserving transformation of a probability space $(X, \mathcal{B}, \mu)$. Let $\zeta$ be a generator with $H_\mu(\zeta) < \infty$. Then $h_\mu(T) = h_\mu(T, \zeta)$.

§8.3 The entropy map and expansive homeomorphisms

§8.3.1 The weak* topology

Let $X$ be a compact metric space equipped with the Borel $\sigma$-algebra $\mathcal{B}$. Let $C(X, \mathbb{R})$ denote the space of all continuous real-valued functions on $X$. Let $M(X)$ denote the set of all Borel probability measures on $X$. Recall that we equip $M(X)$ with the weak* topology as follows: if $\mu_n, \mu \in M(X)$ then we say that $\mu_n \rightharpoonup \mu$ if $\int f \, d\mu_n \to \int f \, d\mu$ as $n \to \infty$ for all $f \in C(X, \mathbb{R})$. The space $M(X)$ is weak* compact.

A measure $\mu$ is $T$-invariant if $\mu(T^{-1}B) = \mu(B)$ for all $B \in \mathcal{B}$. Let $M(X, T)$ denote the space of all $T$-invariant Borel probability measures. We saw in Lecture 6 that $M(X, T)$ is a non-empty compact convex subset of $M(X)$.

§8.3.2 The entropy map

Let $T$ be a continuous transformation of a compact metric space $X$. For each invariant probability measure $\mu \in M(X, T)$ we can calculate the entropy $h_\mu(T)$. How does the entropy vary as a function of $\mu$?

In general, the map $\mu \mapsto h_\mu(T)$ is not continuous, as the following example shows.

**Proposition 8.2**

Let $\Sigma_2 = \{(x_j)_{j=0}^\infty \mid x_j \in \{0, 1\}\}$ denote the full one-sided 2-shift and let $\sigma : \Sigma_2 \to \Sigma_2$ denote the shift map. Then the map $\mu \mapsto h_\mu(\sigma)$ is not weak*
continuous.

**Proof.** Recall that \( x = (x_j) \) is a periodic point for \( \sigma \) with period \( n \) if \( x_j = x_{j+n} \) for all \( j \in \mathbb{N} \). Thus a periodic point is determined by the first \( n \) terms in the sequence \( x \). Hence for each \( n \geq 1 \), there are \( 2^n \) points of period \( n \) for \( \sigma \). Let

\[
\mu_n = \frac{1}{2^n} \sum_{x=\sigma^n x} \delta_x
\]

denote the measure supported on the set of periodic points of period \( n \), giving each periodic point mass \( \frac{1}{2^n} \). It is clear that \( \mu_n \in M(X,T) \).

As \( \mu_n \) is atomic, it is a straightforward calculation to check that \( h_{\mu_n}(\sigma) = 0 \).

Let \( \mu \) denote the Bernoulli \((1/2,1/2)\)-measure. We will show below that \( \mu_n \rightharpoonup \mu \). The proposition follows as \( h_{\mu}(\sigma) = \log 2 \).

To see that \( \mu_n \rightharpoonup \mu \) we have to show that \( \int f \, d\mu_n \rightarrow \int f \, d\mu \) for all continuous function \( f : \Sigma_2 \rightarrow \mathbb{R} \). It is clear that we only need check this for a dense subset of continuous functions. By the Stone-Weierstrass Theorem, functions \( f : \Sigma_2 \rightarrow \mathbb{R} \) that depend only on finitely many co-ordinates, that is functions of the form \( f(x) = f(x_0, \ldots, x_m) \) for some \( m \geq 0 \), are dense in \( C(\Sigma_2, \mathbb{R}) \). (We often refer to such functions as being locally constant.) Hence we only need check that \( \int f \, d\mu_n \rightarrow \int f \, d\mu \) when \( f \) is locally constant. As locally constant functions are finite linear combinations of characteristic functions of cylinders, we need only check that \( \int \chi_{[x_0, \ldots, x_m]} \, d\mu_n \rightarrow \int \chi_{[x_0, \ldots, x_m]} \, d\mu \) for each cylinder \([x_0, \ldots, x_m]\). However, it is clear that if \( n \geq m \) then

\[
\int \chi_{[x_0, \ldots, x_m]} \, d\mu_n = \frac{1}{2^m} = \int \chi_{[x_0, \ldots, x_m]} \, d\mu.
\]

Although in general the map \( \mu \mapsto h_{\mu}(T) \) is not continuous, in many situations it is upper semi-continuous.

**Definition.** The entropy map \( \mu \mapsto h_{\mu}(T) \) is upper semi-continuous if whenever \( \mu_n \rightharpoonup \mu \) we have \( \limsup_{n \to \infty} h_{\mu_n}(T) \leq h_{\mu}(T) \).

(There are examples to show that in general the entropy map is not upper semi-continuous.)

**Proposition 8.3**

Suppose \( T : X \rightarrow X \) is a homeomorphism with the following property: there exists \( \delta > 0 \) such that every finite partition \( \zeta \) of \((X, \mathcal{B})\) with \( \text{diam} \zeta < \delta \) is a generator:

\[
\bigvee_{j=-\infty}^{\infty} T^{-j} \zeta = \mathcal{B}.
\]

Then then entropy map \( \mu \mapsto h_{\mu}(T) \) is upper semi-continuous.
Remark Recall that if $\mu_n \to \mu$ then it does not follow that $\mu_n(B) \to \mu(B)$ for all $B \in \mathcal{B}$. However, one can prove that if $\mu_n \to \mu$ then $\mu_n(B) \to \mu(B)$ for all $B \in \mathcal{B}$ with $\mu(\partial B) = 0$. (Indeed, this is equivalent to $\mu_n \to \mu$.)

Proof. We first claim that one can find a finite partition $\zeta = \{B_1, \ldots, B_k\}$ of $X$ such that $\text{diam} \, \zeta < \delta$ and $\mu(\partial B_i) = 0$. To see this, first note that for any $x \in X$ there exists $\delta' < \delta$ such that $\mu(\partial B(x, \delta')) = 0$ (as we cannot have an uncountable collection of disjoint sets of positive measure). There is a finite open cover $\beta = \{A_1, \ldots, A_r\}$ by open balls of radius $< \delta/2$ and $\mu(\partial A_i) = 0$. Let $B_1 = A_1$ and define inductively $B_n = A_n \setminus (B_1 \cup \cdots \cup B_{n-1})$. Then $\zeta = \{B_1, \ldots, B_n\}$ forms a partition of $X$ such that $\text{diam} \, B_i \leq \text{diam} \, A_i \leq \delta$ and, as $\partial B_i \subset \bigcup_{j=1}^r \partial A_i$, $\mu(\partial B_i) = 0$.

Let $\zeta$ be a finite partition of $X$ with $\text{diam} \, \zeta < \delta$. Then, as $\zeta$ is a generator, by Sinai’s theorem we have that for each $\mu \in M(X,T)$,

$$h_\mu(T) = h_\mu(T, \zeta) \leq H_\mu(\zeta) \leq \log \text{card} \, \zeta < \infty.$$ 

Suppose that $\mu_j \in M(X,T)$ is such that $\mu_j \to \mu$. Choose a finite partition $\zeta = \{B_1, \ldots, B_\ell\}$ with $\text{diam} \, \zeta < \delta$ and $\mu(\partial B_i) = 0$, $1 \leq i \leq \ell$. Then

$$h_{\mu_j}(T) = h_{\mu_j}(T, \zeta) \leq \frac{1}{n} H_{\mu_j} \left( \bigcup_{j=0}^{n-1} T^{-j} \zeta \right). \quad (8.2)$$

Note that, if $\mu(\partial B_i) = 0$ for all $B_i \in \zeta$ then

$$H_{\mu_j}(\zeta) = -\sum_{i=1}^\ell \mu_j(B_i) \log \mu_j(B_i) \to -\sum_{i=1}^\ell \mu(B_i) \log \mu(B_i) = H_\mu(\zeta).$$

More generally, as a typical element of the partition $\bigcup_{j=0}^{n-1} T^{-j} \zeta$ has the form $\bigcap_{k=0}^{n-1} T^{-k} B_{i_k}$ and $\partial(\bigcap_{k=0}^{n-1} T^{-k} B_{i_k}) \subset \bigcup_{k=0}^{n-1} T^{-k} \partial B_{i_k}$, we have

$$H_{\mu_j} \left( \bigcup_{j=0}^{n-1} T^{-j} \zeta \right) \to H_{\mu} \left( \bigcup_{j=0}^{n-1} T^{-j} \zeta \right).$$

Letting $j \to \infty$ in (8.2) we see that for each $n$,

$$\limsup_{j \to \infty} h_{\mu_j}(T) \leq \frac{1}{n} H_{\mu} \left( \bigcup_{j=0}^{n-1} T^{-j} \zeta \right).$$

Letting $n \to \infty$ we obtain

$$\limsup_{j \to \infty} h_{\mu_j}(T) \leq h_\mu(T).$$

$\square$
§8.3.3 Expansive homeomorphisms

We want to find a wide class of examples of dynamical systems that satisfy the hypotheses of Proposition 8.3.

**Definition.** Let $X$ be a compact metric space. A homeomorphism $T : X \to X$ is said to be expansive if there exists $\delta > 0$ such that if $d(T^n x, T^n y) \leq \delta$ for all $n \in \mathbb{Z}$ then $x = y$. We call $\delta$ an expansive constant.

**Remark** Thus $T$ is expansive if the orbits of any two distinct points $x, y \in X$ are, at some time, distance at least $\delta$ apart.

**Proposition 8.4**
A two-sided shift of finite type is expansive.

**Proof.** Let $\sigma : \Sigma_A \to \Sigma_A$ be a shift of finite type. Recall that the metric $d$ on $\Sigma_A$ is defined, essentially, as $d(x, y) = 1/2^{|n|}$ where $n$ is the first place in which the sequences $x, y$ disagree.

Let $\delta = 1/2$. Suppose $x = (x_j)_{j=-\infty}^\infty, y = (y_j)_{j=-\infty}^\infty \in \Sigma_A$. If $x \neq y$ then there exists $n$ such that $x_n \neq y_n$. Hence $d(\sigma^n x, \sigma^n y) = 1 \geq \delta$. \hfill \Box

**Proposition 8.5**
Let $A$ be a $k \times k$ matrix with integer coefficients and $\det A = \pm 1$. Define the linear toral automorphism $T : \mathbb{R}^k / \mathbb{Z}^k \to \mathbb{R}^k / \mathbb{Z}^k$ by $Tx = Ax \mod 1$. Then $T$ is expansive if and only if $A$ is hyperbolic.

**Proof.** Omitted. \hfill \Box

**Remark** More generally, Anosov diffeomorphisms are expansive, as are Axiom A diffeomorphisms on their basic sets.

We will need the following technical result.

**Lemma 8.6**
Let $T$ be an expansive homeomorphism of a compact metric space with expansive constant $\delta > 0$.

(i) The following holds: $\forall \varepsilon > 0, \exists N \in \mathbb{N}$ such that if $d(T^j x, T^j y) \leq \delta$ for $-N \leq j \leq N$ then $d(x, y) < \varepsilon$.

(ii) Let $\zeta = \{B_1, \ldots, B_k\}$ be a finite partition of $X$ with $\text{diam} \zeta \leq \delta$. Then

$$\text{diam} \bigvee_{j=-n}^n T^{-j} \zeta \to 0$$

as $n \to \infty$. 

5
(iii) If \( \{ \zeta_n \} \) is a sequence of finite partitions of \( X \) such that \( \text{diam} \zeta_n \to 0 \) then the smallest \( \sigma \)-algebra containing all the elements of \( \bigvee_{j=1}^{\infty} \zeta_j \), \( n \geq 1 \), is \( \mathcal{B} \), i.e.

\[
\bigvee_{j=1}^{\infty} \zeta_n = \mathcal{B}.
\]

**Proof.** We prove (i). Suppose that the statement in (i) fails: i.e. \( \exists \varepsilon > 0 \) such that \( \forall N \in \mathbb{N} \), \( \exists x_N, y_N \) with \( d(T^j x_N, T^j y_N) \leq \delta, -N \leq j \leq N \) but \( d(x_N, y_N) \geq \varepsilon \). By compactness, we can choose subsequences \( x_{N_k} \to x, y_{N_k} \to y \). Then \( d(T^j x, T^j y) \leq \delta \) for all \( j \in \mathbb{Z} \) but \( d(x, y) \geq \varepsilon > 0 \), contradicting expansiveness.

We prove (ii). First note that \( \text{diam} \bigvee_{j=-N}^{N} T^{-j} \zeta \) decreases as \( n \) increases. Let \( \varepsilon > 0 \). Choose \( N \) as in part (i). Note that if \( x, y \) are in the same element of the partition \( \bigvee_{j=-N}^{N} T^{-j} \zeta \) then \( T^j x, T^j y \) are in the same element of \( \zeta \) for \( -N \leq j \leq N \). Hence \( d(T^j x, T^j y) \leq \delta \) for \( -N \leq j \leq N \). Hence, by part (i), \( d(x, y) < \varepsilon \). Hence \( \text{diam} \bigvee_{j=-N}^{N} T^{-j} \zeta < \varepsilon \). As this is a decreasing sequence, \( \text{diam} \bigvee_{j=-n}^{n} T^{-j} \zeta < \varepsilon \) for all \( n \geq N \).

We prove (iii). It is sufficient to prove that every open ball \( B(x, \varepsilon) \) is in \( \bigvee_{n=1}^{\infty} \zeta_n \). For each \( k \), choose \( N_k \) such that if \( n \geq N_k \) then \( \text{diam} \zeta_n \leq 1/k \). For \( k > [1/\varepsilon] \), let \( E_k \) denote the union of all sets in \( \zeta_{N_k} \) that intersect \( B(x, \varepsilon - 1/k) \). Then

\[
B(x, \varepsilon - 1/k) \subset E_k \subset B(x, \varepsilon).
\]

Hence \( \bigcup_{k=1}^{\infty} E_k = B(x, \varepsilon) \). Since \( E_k \) is in the \( \sigma \)-algebra generated by \( \zeta_{N_k} \) we have that

\[
\bigcup_{k=1}^{\infty} E_k \in \bigvee_{n=1}^{\infty} \zeta_n.
\]

**Proposition 8.7**

Let \( T \) be an expansive homeomorphism of a compact metric space \( X \) and let \( \delta \) be an expansive constant for \( T \). Let \( \zeta \) be a finite partition of \( X \) with \( \text{diam} \zeta < \delta \). Then

\[
\bigvee_{j=-\infty}^{\infty} T^{-j} \zeta = \mathcal{B}.
\]

**Proof.** Let \( \zeta_n = \bigvee_{j=-n}^{n} T^{-j} \zeta \). Then, by Lemma 8.6(ii), \( \text{diam} \zeta_n \to 0 \) as \( n \to \infty \). By Lemma 8.6(iii) we have

\[
\mathcal{B} = \bigvee_{n=1}^{\infty} \zeta_n = \bigvee_{j=-\infty}^{\infty} T^{-j} \zeta.
\]

\( \square \)
Combining Proposition 8.3 and Proposition 8.7 we see that the entropy map for expansive homeomorphisms is upper semi-continuous.

**Theorem 8.8**

Let $T$ be an expansive homeomorphism of the compact metric space $X$. Then the entropy map

$$\mu \mapsto h_\mu(T) : M(X,T) \to \mathbb{R}$$

is upper semi-continuous in the weak* topology.

§8.4 Topological entropy

Let $X$ be a compact metric space and let $T : X \to X$ be continuous. We introduce topological entropy - a topological analogue of metric entropy.

§8.4.1 Definition in terms of open covers

Recall that an open cover $\alpha$ of a metric space $X$ is a collection of open sets $\alpha = \{U_i, i \in I\}$ such that $\bigcup_{U_i \in \alpha} U_i = X$. In most cases, the sets $U_i \in \alpha$ will not be pairwise disjoint.

First recall that if $X$ is a compact metric space then every open cover has a finite subcover. For a given open cover there may be lots of different ways of choosing a finite subcover. However, given an open cover we can always choose a finite subcover of smallest cardinality.

**Definition.** Let $\alpha$ be an open cover of $X$. We define the entropy of $\alpha$ to be

$$H_{\text{top}}(\alpha) = \log N(\alpha)$$

where $N(\alpha)$ is the cardinality of the smallest finite subcover of $\alpha$.

We can form joins and refinements as with metric entropy:

**Definition.** Let $\alpha, \beta$ be open covers of $X$. We define the join $\alpha \vee \beta$ to be the open cover of $X$ by sets of the form $T^{-1}A$, $A \in \alpha$.

**Definition.** Let $\alpha, \beta$ be open covers of $X$. We say that $\beta$ is a refinement of $\alpha$ and write $\alpha \leq \beta$ if every member of $\beta$ is a subset of an element of $\alpha$.

Note that $\alpha \leq \alpha \vee \beta$ for all open covers $\alpha, \beta$.

**Definition.** Let $T : X \to X$ be a continuous transformation. We define $T^{-1}\alpha$ to be the open cover of $X$ by sets of the form $T^{-1}A$, $A \in \alpha$.

It is straightforward to check from the definitions that $T^{-1}(\alpha \vee \beta) = T^{-1}\alpha \vee T^{-1}\beta$, and if $\alpha \leq \beta$ then $T^{-1}\alpha \leq T^{-1}\beta$.

Here are some easy, but useful, properties of $H(\alpha)$. 


Lemma 8.9  
(i) If $\alpha \leq \beta$ then $H_{\text{top}}(\alpha) \leq H_{\text{top}}(\beta)$.

(ii) If $\alpha$ and $\beta$ are open covers then $H_{\text{top}}(\alpha \vee \beta) \leq H_{\text{top}}(\alpha) + H_{\text{top}}(\beta)$.

(iii) If $T$ is continuous then $H_{\text{top}}(T^{-1} \alpha) \leq H_{\text{top}}(\alpha)$. If, in addition, $T$ is surjective then $H_{\text{top}}(T^{-1} \alpha) = H_{\text{top}}(\alpha)$.

Proof.  
(i) Suppose that $\alpha \leq \beta$. Choose a subcover $\{B_1, \ldots, B_{N(\beta)}\}$ of $\beta$ of minimal cardinality. For each $i$, choose $A_i \in \alpha$ such that $B_i \subset A_i$. Then $\{A_1, \ldots, A_{N(\beta)}\}$ is a finite subcover of $\alpha$ of cardinality $N(\beta)$. Hence $N(\alpha) \leq N(\beta)$.

(ii) Let $\alpha = \{A_1, \ldots, A_{N(\alpha)}\}$, $\beta = \{B_1, \ldots, B_{N(\beta)}\}$ be subcovers of minimal cardinality of $\alpha$, $\beta$, respectively. Then

$$\{A_i \cap B_j \mid 1 \leq i \leq N(\alpha), 1 \leq j \leq N(\beta)\}$$

is a finite subcover of $\alpha \vee \beta$ of cardinality $N(\alpha)N(\beta)$. Hence $N(\alpha \vee \beta) \leq N(\alpha)N(\beta)$. Taking logs we see that $H_{\text{top}}(\alpha \vee \beta) \leq H_{\text{top}}(\alpha) + H_{\text{top}}(\beta)$.

(iii) Let $\{A_1, \ldots, A_{N(\alpha)}\}$ be a finite subcover of $\alpha$ of minimal cardinality. Then $\{T^{-1}A_1, \ldots, T^{-1}A_{N(\alpha)}\}$ is a subcover of $T^{-1}\alpha$. Hence $N(T^{-1}\alpha) \leq N(\alpha)$.

Conversely, suppose that $\{T^{-1}A_1, \ldots, T^{-1}A_{N(T^{-1}\alpha)}\}$ is a subcover of $T^{-1}\alpha$ of minimal cardinality. As $T$ is surjective, $\{A_1, \ldots, A_{N(T^{-1}\alpha)}\}$ is a subcover of $\alpha$. Hence $N(\alpha) \leq N(T^1\alpha)$.

We can now define the entropy of $T$ relative to an open cover $\alpha$.

Definition. Let $T : X \to X$ be a continuous transformation of a compact metric space $X$ and let $\alpha$ be an open cover. Define

$$h_{\text{top}}(T, \alpha) = \lim_{n \to \infty} \frac{1}{n} H_{\text{top}} \left( \bigvee_{j=0}^{n-1} T^{-j} \alpha \right)$$  \hspace{1cm} (8.3)$$

to be the topological entropy of $T$ relative to $\alpha$.

To check that the limit exists it is sufficient to check that quantities in (8.3) form a subadditive sequence. This follows as

$$H_{\text{top}} \left( \bigvee_{j=0}^{n+m-1} T^{-j} \alpha \right) = H_{\text{top}} \left( \bigvee_{j=0}^{n-1} T^{-j} \alpha \vee \bigvee_{j=n}^{n+m-1} T^{-j} \alpha \right)$$

$$\leq H_{\text{top}} \left( \bigvee_{j=0}^{n-1} T^{-j} \alpha \right) + H_{\text{top}} \left( \bigvee_{j=0}^{m-1} T^{-j} \alpha \right)$$

8
\[
\leq H_{\text{top}} \left( \bigvee_{j=0}^{n-1} T^{-j} \alpha \right) + H_{\text{top}} \left( \bigvee_{j=0}^{m-1} T^{-j} \alpha \right)
\]
by Lemma 8.9.

Finally, we can now define the topological entropy of \( T \).

**Definition.** Let \( T \) be a continuous transformation of a compact metric space \( X \). We define the *topological entropy of \( T \)* to be

\[
h_{\text{top}}(T) = \sup \{ h_{\text{top}}(T, \alpha) \mid \alpha \text{ is an open cover of } X \}.
\]

§8.4.2 Bowen’s definition

Let \( X \) be a compact metric space with metric \( d \). For \( x \in X, r > 0 \), we define the open and closed balls with centre \( x \) and radius \( r \) to be

\[
B(x, r) = \{ y \in X \mid d(x, y) < r \}
\]
\[
D(x, r) = \{ y \in X \mid d(x, y) \leq r \}.
\]

Let \( T : X \to X \) be continuous. For each \( n \geq 1 \) we define the metric \( d_n \) by

\[
d_n(x, y) = \max_{0 \leq j \leq n-1} d(T^j x, T^j y).
\]

Thus an \( d_n \)-open ball with centre \( x \) and radius \( \varepsilon \) is

\[
B_n(x, \varepsilon) = \bigcap_{j=0}^{n-1} T^{-j} B(T^j x, \varepsilon).
\]

Alternatively, a point \( y \) is within \( d_n \)-distance \( \varepsilon \) of \( x \) if the first \( n \) iterates of both \( x \) and \( y \) remain within distance \( \varepsilon \) of each other; i.e. \( d_n(x, y) < \varepsilon \) if and only if \( d(T^j x, T^j y) < \varepsilon \) for \( 0 \leq j \leq n - 1 \).

**Definition.** Let \( n \geq 1, \varepsilon > 0 \). A subset \( F \subset X \) is said to be \((n, \varepsilon)\)-span \( X \) with respect to \( T \) if the set of \( d_n \)-balls of radius \( \varepsilon \) and centres in \( F \) covers \( X \). That is,

\[
X = \bigcup_{x \in F} B_n(x, \varepsilon).
\]

We want to make \((n, \varepsilon)\)-spanning sets as small as possible. This motivates the following definition.

**Definition.** Let \( n \geq 1, \varepsilon > 0 \). Let \( p_n(\varepsilon) \) denote the smallest cardinality of an \((n, \varepsilon)\)-spanning set with respect to \( T \).

**Remarks.**
(i) Note that as $X$ is compact with respect to the $d_n$-metric, it follows that $p_n(\varepsilon) < \infty$.

(ii) It is clear from the definition that if $\varepsilon_1 < \varepsilon_2$ then $p_n(\varepsilon_1) \geq p_n(\varepsilon_2)$.

As $n$ increases, we would expect the quantity $p_n(\varepsilon)$ to grow exponentially fast. This motivates the following definition:

**Definition.** Let $\varepsilon > 0$. Define

$$p(\varepsilon) = \limsup_{n \to \infty} \frac{1}{n} p_n(\varepsilon).$$

Possibly $p(\varepsilon)$ could be $\infty$.

As $p_n(\varepsilon)$ increases as $\varepsilon \searrow 0$, it follows that $p(\varepsilon)$ increases as $\varepsilon \searrow 0$. Hence we can make the following definition.

**Definition.** We define the entropy of $T$ to be

$$h_{\text{top}}(T) = \lim_{\varepsilon \to 0} p(\varepsilon).$$

We shall show below that this definition agrees with the definition in 8.4.1. Before we do this, it is useful to have another definition of $h(T)$.

**Definition.** Let $n \geq 1, \varepsilon > 0$. A subset $E \subset X$ is said to $(n, \varepsilon)$-separate $X$ if the $d_n$-distance between any two points in $E$ is greater than $\varepsilon$: i.e. for all $x, y \in E$, $x \neq y$, $d_n(x, y) > \varepsilon$.

We want to make $(n, \varepsilon)$-separated sets as large as possible. Thus we define:

**Definition.** Let $n \geq 1, \varepsilon > 0$. Let $q_n(\varepsilon)$ denote the largest cardinality of an $(n, \varepsilon)$-separated set.

There is a nice relation between $q_n(\varepsilon)$ and $p_n(\varepsilon)$.

**Lemma 8.10**

Let $\varepsilon > 0$. Then $p_n(\varepsilon) \leq q_n(\varepsilon) \leq p_n(\varepsilon/2)$.

**Proof.** Let $E$ be an $(n, \varepsilon)$-separated set of maximal cardinality $q_n(\varepsilon)$. Then $E$ must be $(n, \varepsilon)$-spanning (if not, there would be a point in $X$ of $d_n$-distance at least $\varepsilon$ from all the points in $E$, contradicting maximality). Hence $p_n(\varepsilon) \leq q_n(\varepsilon)$.

For the other inequality, again suppose that $E$ is an $(n, \varepsilon)$-separated set of maximal cardinality $q_n(\varepsilon)$. Let $F$ be an $(n, \varepsilon/2)$-spanning set. Let $x \in E$. Then there exists $y \in F$ such that $d_n(x, y) < \varepsilon/2$. The map that sends $x \mapsto y : E \to F$ is injective. (If not, then two different points $x, x' \in E$ could map to the same $y \in F$. Then $d_n(x, x') \leq d_n(x, y) + d_n(y, x') < \varepsilon$, contradicting the fact that $E$ is $(n, \varepsilon)$-separated. Hence the cardinality of $F$ is greater than $q_n(\varepsilon)$, that is, $q_n(\varepsilon) \leq p_n(\varepsilon/2)$. \qed
It follows that if we define
\[ q(\varepsilon) = \limsup_{n \to \infty} \frac{1}{n} \log q_n(\varepsilon) \]
then
\[ h_{\text{top}}(T) = \lim_{\varepsilon \to 0} q(\varepsilon). \]

§8.4.3 The two definitions of topological entropy coincide

We will show that the definition of topological entropy using open covers and the definition using separated/spanning sets coincide. Let us (temporarily) introduce the notation \( h_{oc}(T) \) to denote the topological entropy of \( T \) using open covers, and \( h_s(T) \) to denote the topological entropy of \( T \) using either separated or spanning sets.

It will be useful to recall the notion of a Lebesgue number for an open cover.

**Definition.** Let \( \alpha \) be an open cover of a metric space \( X \). We say that \( r > 0 \) is a Lebesgue number for \( \alpha \) if for every \( x \in X \) there exists \( A \in \alpha \) such that \( B(x, r) \subset A \).

If \( X \) is compact then every open cover has a finite Lebesgue number.

If \( \alpha \) is an open cover then we define the diameter of \( \alpha \) to be \( \text{diam } \alpha = \sup_{A \in \alpha} \text{diam } A \). Suppose that \( \alpha \) is an open cover for \( X \) with Lebesgue number \( r > 0 \). Suppose that \( \beta \) is another open cover of \( X \) and \( \text{diam } \beta < r \). Then every element of \( \beta \) is contained in an element of \( \alpha \). Hence \( \alpha \leq \beta \).

The following result is useful in calculating \( h_{oc}(T) \); it is the topological analogue of Abramov’s theorem in the calculation of metric entropy.

**Proposition 8.11**

Let \( T : X \to X \) be a continuous transformation of a compact metric space \( X \). Suppose that \( \alpha_n \) is a sequence of open covers of \( X \) such that \( \text{diam } \alpha_n \to 0 \). Then
\[ h_{oc}(T) = \lim_{n \to \infty} h_{oc}(T, \alpha_n). \]

**Proof.** Let \( \varepsilon > 0 \). Choose an open cover \( \beta \) such that
\[ h_{oc}(T, \beta) = \begin{cases} h_{oc}(T) - \varepsilon & \text{if } h_{oc}(T) < \infty \\ 1/\varepsilon & \text{if } h_{oc}(T) = \infty. \end{cases} \]

Let \( r \) be a Lebesgue number for \( \beta \). Choose \( N \) such that if \( n \geq N \) then \( \text{diam } \alpha_n < r \). Hence \( \beta \leq \alpha_n \) for all \( n \geq N \). Hence \( h_{oc}(T, \alpha_n) \geq h_{oc}(T, \beta) \).

Letting \( n \to \infty \) and then letting \( \varepsilon \to 0 \) shows that \( \lim_{n \to \infty} h_{oc}(T, \alpha_n) = h_{oc}(T) \).

The following is the key technical result.

11
Proposition 8.12
Let $X$ be a compact metric space and let $T : X \to X$ be continuous.

(i) Let $\alpha$ be an open cover of $X$ with Lebesgue number $r > 0$. Then

$$N \left( \bigvee_{j=0}^{n-1} T^{-j} \alpha \right) \leq p_n(r/2) \leq q_n(r/2).$$

(ii) Let $\beta$ be an open cover with $\text{diam} \beta < \varepsilon$. Then

$$p_n(\varepsilon) \leq q_n(\varepsilon) \leq N \left( \bigvee_{j=0}^{n-1} T^{-j} \beta \right).$$

Proof. We already know that $p_n(\varepsilon) \leq q_n(\varepsilon)$.

We prove (i). Let $F$ be an $(n, r/2)$-spanning set of minimal cardinality $p_n(r/2)$ so that

$$X = \bigcup_{x \in F} B_n(x, r/2) = \bigcup_{x \in F} \bigcap_{j=0}^{n-1} T^{-j} B(T^j x, r/2).$$

For each $j$, $B(T^j x, r/2)$ is an open set of diameter $r$. As $r$ is a Lebesgue number for $\alpha$, it follows that, for each $j$, $B(T^j x, r/2)$ is a subset of an element of $\alpha$. Hence

$$N \left( \bigvee_{j=0}^{n-1} T^{-j} \alpha \right) \leq p_n(r/2).$$

We prove (ii). Let $E$ be an $(n, \varepsilon)$-separated set of cardinality $q_n(\varepsilon)$. No member of the open cover

$$\bigvee_{j=0}^{n-1} T^{-j} \beta$$

can contain two elements of $E$. Hence

$$q_n(\varepsilon) \leq N \left( \bigvee_{j=0}^{n-1} T^{-j} \beta \right).$$

\hfill $\square$

Finally we can prove the two definitions of topological entropy coincide.

Theorem 8.13
Let $X$ be a compact metric space and let $T : X \to X$ be continuous. Then

$$h_{oc}(T) = h_s(T).$$
Proof. Let $\varepsilon > 0$. Let $\alpha_\varepsilon$ denote the open cover of $X$ using all open balls (using the metric $d$) of radius $3\varepsilon$. Note that $\alpha_\varepsilon$ has Lebesgue number $2\varepsilon$. Let $\beta_\varepsilon$ be any cover by open balls of radius $\varepsilon/2$. Then by Proposition 8.12 we have

$$N \left( \bigvee_{j=0}^{n-1} T^{-j} \alpha_\varepsilon \right) \leq p_n(\varepsilon) \leq q_n(\varepsilon) \leq N \left( \bigvee_{j=0}^{n-1} T^{-j} \beta_\varepsilon \right).$$  

Taking logarithms, dividing by $n$ and letting $n \to \infty$ in (8.4) gives

$$h_{oc}(T, \alpha_\varepsilon) \leq p(\varepsilon) \leq q(\varepsilon) \leq h_{oc}(T, \beta_\varepsilon).$$

Let $\varepsilon = 1/k$. Letting $k \to \infty$ and using Proposition 8.11 we obtain

$$h_{oc}(T) \leq h_s(T) \leq h_{oc}(T)$$

and the result follows.

§8.5 Calculating topological entropy

Let $T$ be a continuous homeomorphism of a compact metric space $X$. Let $\alpha = \{A_1, \ldots, A_k\}$ be a finite open cover for $X$. For each point $x \in X$ we can look at the sequence of elements of $\alpha$ that the orbit of $x$ visits (in the same way that we coded, for example, the orbits of the doubling map in Lecture 1). Note that $x \in X$ has coding $(i_j)_{j=-\infty}^{\infty}$ precisely when $x \in \bigcap_{j=-\infty}^{\infty} T^{-j} A_{i_j}$. Note that, in general, two points may have the same coding.

Definition. We say that a finite open cover $\alpha$ is a (topological) generator if for each sequence $(i_j)_{j=-\infty}^{\infty} \in \{1, \ldots, k\}^\mathbb{Z}$ we have

$$\text{card} \bigcap_{j=-\infty}^{\infty} T^{-j} \overline{A_{i_j}} = 0 \text{ or } 1.$$

(Here, $\overline{A}$ denotes the closure of $A$.)

Remark. In other words, $\alpha$ is a generator if every possible sequence has at most one point with that coding.

The existence of a generator is very closely related to expansivity properties of the dynamics.

Proposition 8.14
The following are equivalent:

(i) $T$ is an expansive homeomorphism;

(ii) there exists a (topological) generator;
(iii) there exists a (topological) weak generator, i.e. a finite open cover 
\( \alpha = \{ A_1, \ldots, A_k \} \) such that for every sequence \( (i_j)_{j=-\infty}^{\infty} \in \{1, \ldots, k\}^\mathbb{Z} \) we have
\[
\text{card} \bigcap_{j=-\infty}^{\infty} T^{-j} A_{i_j} = 0 \text{ or } 1.
\]

**Proof.** Omitted. See Walters. \( \square \)

The following is the analogue of Sinai’s theorem for topological entropy.

**Proposition 8.15**

Let \( T \) be an expansive homeomorphism of a compact metric space \( X \) and let \( \alpha \) be a (topological) generator. Then \( h_{\text{top}}(T) = h_{\text{top}}(T, \alpha) \).

**Proof.** By definition, \( h_{\text{top}}(T, \alpha) \leq h_{\text{top}}(T) \).

We first claim that \( \text{diam} \bigvee_{j=-n}^{n} T^{-j} \alpha \to 0 \) as \( n \to \infty \). Certainly this sequence decreases, so suppose for a contradiction that \( \text{diam} \bigvee_{j=-n}^{n} T^{-j} \alpha \to \varepsilon_0 > 0 \) as \( n \to \infty \). So for all \( n \geq 1 \), there exists \( x_n, y_n \) in the same element of \( \bigvee_{j=-n}^{n} T^{-j} \alpha \). Hence \( d(x_n, y_n) \geq \varepsilon_0 / 2 \). By compactness, we may take a subsequence so that \( x_n \to x \), \( y_n \to y \). Let \( x_{n_j}, y_{n_j} \in \bigcap_{k=-n_j}^{n_j} T^{-k} \alpha \). Fix \( k \). Then infinitely many of the \( A_{n_j,k} \) are the same, as \( \alpha \) is a finite open cover, \( A_{n_j,k} = A'_{n_j} \) say. Then \( x, y \in T^{-j} A'_{n_j} \). Hence \( x, y \in \bigcap_{j=-\infty}^{\infty} T^{-j} A'_{n_j} \), so \( x = y \), a contradiction.

Let \( \beta \) be any open cover of \( X \) and let \( r > 0 \) be a Lebesgue number for \( \beta \). Choose \( N \) such that \( \text{diam} \bigvee_{j=-N}^{N} T^{-j} \alpha \leq r \). Then \( \beta < \bigvee_{j=-N}^{N} T^{-j} \alpha \).

Hence
\[
h_{\text{top}}(T, \beta) \leq h_{\text{top}} \left( T, \bigvee_{j=-N}^{N} T^{-j} \alpha \right)
= \lim_{n \to \infty} \frac{1}{n} H_{\text{top}} \left( \bigvee_{i=0}^{n-1} T^{-i} \bigvee_{j=-N}^{N} T^{-j} \alpha \right)
= \lim_{n \to \infty} \frac{1}{n} H_{\text{top}} \left( \bigvee_{j=-N}^{N+n-1} T^{-j} \alpha \right)
= \lim_{n \to \infty} \frac{2N + n - 1}{n} \frac{1}{2N + n - 1} H_{\text{top}} \left( \bigvee_{j=0}^{2N+n-1} T^{-j} \alpha \right)
= h_{\text{top}}(T, \alpha).
\]

Hence \( h_{\text{top}}(T, \beta) \leq h_{\text{top}}(T, \alpha) \). Taking the supremum over all open covers \( \beta \), we have \( h_{\text{top}}(T) \leq h_{\text{top}}(T, \alpha) \). \( \square \)

We can now calculate the topological entropy for some shifts.
Proposition 8.16
Let $\sigma : \Sigma_k \to \Sigma_k$ denote the full two-sided $k$-shift on symbols $\{1, \ldots, k\}$. Then $h_{\text{top}}(T) = \log k$.

Proof. Let $\alpha = \{[1], \ldots, [k]\}$ denote the partition of $\Sigma_k$ into cylinders of length 1. Note that $\alpha$ is an open cover of $\Sigma_k$. As cylinders are also closed, it is straightforward to see that $\alpha$ is a (topological) generator.

Note that the elements of $\bigvee_{j=0}^n \sigma^{-j} \alpha$ is the partition of $\Sigma_k$ into cylinders of length $n$, of which there are $k^n$. Hence

$$h_{\text{top}}(\sigma) = h(\sigma, \alpha)$$
$$= \lim_{n \to \infty} \frac{1}{n} H_{\text{top}} \left( \bigvee_{j=0}^{n-1} \sigma^{-j} \alpha \right)$$
$$= \lim_{n \to \infty} \frac{1}{n} \log k^n$$
$$= \log k.$$

\[ \square \]

More generally, we can calculate the topological entropy of a shift of finite type.

Proposition 8.17
Let $A$ be an irreducible $k \times k$ matrix with entries in $\{0, 1\}$ and let $\sigma : \Sigma_A \to \Sigma_a$ be the corresponding shift of finite type. Then $h_{\text{top}}(\sigma) = \log \lambda$ where $\lambda$ is the largest positive eigenvalue of $A$.

Remark The largest (in modulus) eigenvalue of an irreducible matrix is always real by the Perron-Frobenius theorem.

Proof. Let $\alpha = \{[1], \ldots, [k]\}$. Again, $\alpha$ is a finite open cover and a (topological) generator.

We have already seen that $\bigvee_{j=0}^{n-1} \sigma^{-j} \alpha$ is the partition of $\Sigma_A$ into cylinders of length $n$. Hence

$$H_{\text{top}} \left( \bigvee_{j=0}^{n-1} \sigma^{-j} \alpha \right) = \log \text{card } \{ \text{no. of cylinders of length } n \text{ in } \Sigma_A \}. $$

Note that the cylinder $[i_0, \ldots, i_{n-1}] \cap \Sigma_A \neq \emptyset$ if and only if $A_{i_0, i_1} A_{i_1, i_2} \cdots A_{i_{n-2}, i_{n-1}} = 1$. Hence the number of cylinders of length $n$ that intersect $\Sigma_A$ is $\|A^n\|$ where $\| \cdot \|$ is the matrix norm given by $\|B\| = \sum_{i,j} |B_{i,j}|$. 

15
Hence
\[ h_{\text{top}}(\sigma) = h_{\text{top}}(\sigma, \alpha) = \lim_{n \to \infty} \frac{1}{n} H_{\text{top}} \left( \sum_{j=0}^{n-1} \sigma^{-j} \alpha \right) = \lim_{n \to \infty} \frac{1}{n} \log \| A^n \| = \log \lambda \]
by the spectral radius formula.

§8.5.1 The variational principle

The following important result relates metric and topological entropy.

**Theorem 8.18 (The variational principle)**

Let \( T : X \to X \) be a continuous transformation of a compact metric space \( X \). Then
\[ h_{\text{top}}(T) = \sup_{\mu \in M(X, T)} h_{\mu}(T). \]

**Proof.** We use \( \alpha, \beta, \ldots \) to denote open covers (used in the calculation of topological entropy) of \( X \), and \( \zeta, \eta, \ldots \) to denote partitions (used in the calculation of metric entropy) of \( X \).

First note that one can also define metric entropy as
\[ h_{\mu}(T) = \sup \mu_{\mu}(T, \zeta) \]
where the supremum is taken over finite partitions. (Our previous definition had the supremum taken over all partitions of finite entropy.)

Let \( \mu \in M(X, T) \). We show that \( h_{\mu}(T) \leq h_{\text{top}}(T) \). Let \( \zeta = \{ A_1, \ldots, A_k \} \) be a finite partition of \( X \). Choose \( \varepsilon > 0 \) such that \( \varepsilon < 1/k \log k \). We can choose compact sets \( B_j \) such that \( B_j \subset A_j \) and \( \mu(A_j \setminus B_j) \leq \varepsilon \), \( 1 \leq j \leq k \).

Let \( \eta \) denote the partition \( \eta = \{ B_0, B_1, \ldots, B_k \} \) where \( B_0 = X \setminus \bigcup_{j=1}^{k} B_j \).

Note that \( \mu(B_0) \leq k \varepsilon \).

We calculate the conditional entropy of \( \zeta \) given \( \eta \). Indeed,
\[ H_{\mu}(\zeta \mid \eta) = -\sum_{i=0}^{k} \mu(B_i) \sum_{j=1}^{k} \frac{\mu(A_j \cap B_i)}{\mu(B_i)} \log \frac{\mu(A_j \cap B_i)}{\mu(B_i)} \]
\[ = -\mu(B_0) \sum_{j=1}^{k} \frac{\mu(A_j \cap B_0)}{\mu(B_0)} \log \frac{\mu(A_j \cap B_0)}{\mu(B_0)} \]
\[ \leq \mu(B_0) \log k \]
\[ \leq k \varepsilon \log k \]
\[ \leq 1 \]
since, for \( i \neq 0 \), \( \mu(A_j \cap B_i) / \mu(B_i) = 0 \) or 1. Hence
\[
h_\mu(T, \zeta) \leq h_\mu(T, \eta) + H_\mu(\zeta | \eta) \leq h_\mu(T, \eta) + 1. \tag{8.5}
\]

For \( i \neq 0 \), we have that \( B_0 \cup B_i = X \setminus \cup_{j \neq i} B_j \) is open. Hence
\[
\beta = \{B_0 \cup B_1, \ldots, B_0 \cup B_k\}
\]
is an open cover of \( X \). Moreover,
\[
H_\mu \left( \bigvee_{j=0}^{n-1} T^{-j} \eta \right) \leq \log \left( \text{card} \bigvee_{j=0}^{n-1} T^{-j} \eta \right) \\
\leq \log \left( 2^n N \left( \bigvee_{j=0}^{n-1} T^{-j} \beta \right) \right) \\
= n \log 2 + \log N \left( \bigvee_{j=0}^{n-1} T^{-j} \beta \right),
\]
so that
\[
h_\mu(T, \eta) \leq \log 2 + h_{\text{top}}(T, \beta) \leq \log 2 + h(T) \tag{8.6}
\]
Combining (8.5) and (8.6) we have that
\[
h_\mu(T, \zeta) \leq h_{\text{top}}(T) + \log 2 + 1.
\]
Taking the supremum over all finite partitions \( \zeta \) we have
\[
h_\mu(T) \leq h_{\text{top}}(T) + \log 2 + 1. \tag{8.7}
\]

This holds for all continuous transformations \( T \). In particular, replacing \( T \) by \( T^n \) for each \( n \geq 1 \) and using exercise 8.4, we have
\[
nh_\mu(T) = h_\mu(T^n) \leq h_{\text{top}}(T^n) + \log 2 + 1 = nh_{\text{top}}(T) + \log 2 + 1.
\]
Dividing by \( n \) and letting \( n \to \infty \) gives \( h_\mu(T) \leq h_{\text{top}}(T) \).

Conversely, we show that \( h_{\text{top}}(T) \geq \sup h_\mu(T) \) where the supremum is taken over all \( \mu \in M(X, T) \). Let \( \varepsilon > 0 \). To show that \( h_{\text{top}}(T) \geq \sup h_\mu(T) \) it is sufficient to construct \( \mu \in M(X, T) \) such that \( h_\mu(T) \geq q(\varepsilon) \).

Let \( E_n \) be an \((n, \varepsilon)\)-separated set of cardinality \( q_n(\varepsilon) \). Define
\[
\sigma_n = \frac{1}{q_n(\varepsilon)} \sum_{x \in E_n} \delta_x \in M(X).
\]
Let
\[
\mu_n = \frac{1}{n} \sum_{j=0}^{n-1} T_j^* \sigma_n.
\]
Since \( M(X) \) is weak* compact, \( \mu_n \) has a weak* convergent subsequence with limit \( \mu \in M(X, T) \) (see Lecture 6). One can then show (with some work) that \( h_\mu(T) \geq q(\varepsilon) \). \( \square \)
§8.5.2 Measures of maximal entropy

We are interested in measures which maximise the entropy.

**Definition.** A measure $\mu \in M(X, T)$ is called a measure of maximal entropy if $h_\mu(T) = h_{\text{top}}(T)$.

Let $M_{\text{max}}(X, T)$ denote the set of all invariant measures whose entropy achieves the maximum:

$$M_{\text{max}}(X, T) = \{ \mu \in M(X, T) \mid h_\mu(T) = h_{\text{top}}(T) \}.$$ 

It is possible for $M_{\text{max}}(X, T)$ to be empty. However, under reasonable assumptions there is always at least one measure of maximal entropy.

**Proposition 8.19**
Suppose that the entropy map $\mu \mapsto h_\mu(T) : M(X, T) \to \mathbb{R}$ is upper semi-continuous. Then $M_{\text{max}}(X, T) \neq \emptyset$.

**Proof.** An upper semi-continuous function on a compact metric space attains its supremum. \qed

More interesting is the case when there is only one invariant measure that maximises entropy.

**Definition.** We say that $T$ has a unique measure of maximal entropy if $M_{\text{max}}(X, T)$ contains exactly one point.

**Lemma 8.20**
Suppose $T$ has a unique measure of maximal entropy $\mu$. Then $\mu$ is ergodic.

**Proof.** Exercise. \qed

We can show that shifts of finite type have measures of maximal entropy.

**Proposition 8.21**
Let $\Sigma_k$ denote the full one-sided $k$-shift with shift map $\sigma : \Sigma_k \to \Sigma_k$. Then the Bernoulli $(1/k, \ldots, 1/k)$-measure is the unique measure of maximal entropy.

**Proof.** We already know that the topological entropy, $h_{\text{top}}(\sigma)$, is equal to $\log k$. The entropy of the Bernoulli $(1/k, \ldots, 1/k)$-measure is also $\log k$.

Let $\mu \in M(X, T)$ and suppose that $h_\mu(\sigma) = \log k$. We need to show that $\mu$ is the Bernoulli $(1/k, \ldots, 1/k)$-measure. By the Kolmogorov Extension Theorem, we need only check that the $\mu$-measure of each cylinder of length $n$ is $1/k^n$. 

18
Let $\zeta$ denote the partition into cylinders of length 1: $\zeta = \{[1], \ldots, [k]\}$. By Sinai’s theorem, $h_\mu(\sigma, \zeta) = h_\mu(\sigma) = \log k$. Moreover, for each $n \geq 1$,

$$
\log k = h_\mu(\sigma, \zeta) \leq \frac{1}{n} H_\mu \left( \bigvee_{j=0}^{n-1} \sigma^{-j} \zeta \right) \leq \frac{1}{n} \log k^n = \log k.
$$

Hence

$$
H_\mu \left( \bigvee_{j=0}^{n-1} T^{-j} \zeta \right) = \log k^n. \tag{8.8}
$$

We need the following fact: if $\phi(x)$ is strictly concave then

$$
\sum_{j=1}^{k} a_j \phi(x_j) \leq \phi \left( \sum_{j=1}^{k} a_j x_j \right) \tag{8.9}
$$

where $a_j \geq 0$, $\sum_{j=1}^{k} a_j = 1$, with equality if and only if all the $x_j$ corresponding to non-zero $a_j$ are equal. Let $\eta = \{A_1, \ldots, A_k\}$ be a partition with $k$ elements. Note that $\phi(x) = -x \log x$ is strictly concave. Putting $a_j = 1/k$ and $x_j = \mu(A_j)$ into (8.9) we see that $H_\mu(\eta) \leq \log k$ with equality precisely when $\mu(A_j) = 1/k$, $1 \leq j \leq k$. As there are $k^n$ elements in the partition $\bigvee_{j=0}^{n-1} T^{-j} \zeta$, it follows from (8.8) that every element in $\bigvee_{j=0}^{n-1} T^{-j} \zeta$ has the same $\mu$-measure, namely $1/k^n$.

Hence $\mu$ agrees with the Bernoulli $(1/k, \ldots, 1/k)$-measure on cylinders. By the Kolmogorov Extension Theorem, $\mu$ is the Bernoulli $(1/k, \ldots, 1/k)$-measure.

One can generalise the above result to shifts of finite type defined by irreducible matrices.

Let $A$ be an irreducible $k \times k$ matrix with entries in $\{0, 1\}$. Let $\Sigma_A$ denote the corresponding shift of finite type with shift map $\sigma : \Sigma_A \rightarrow \Sigma_A$.

Let $P$ be a stochastic matrix and let $p$ be a left probability eigenvector, so that $pP = p$. Recall that we can define a Markov measure $\mu$ on $\Sigma_A$ by defining it on cylinders by

$$
\mu[i_0, \ldots, i_n] = p_{i_0} P_{i_0,i_1} \cdots P_{i_n-1,i_n}.
$$

We will need the following well-known result:

**Theorem 8.22 (Perron-Frobenius)**

Let $A$ 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 $|\lambda_i| < \lambda$,
(ii) the eigenvalue $\lambda$ is simple (i.e. the corresponding eigenspace is one-dimensional),

(iii) there is a unique right-eigenvector $v = (v_1, \ldots, v_k)^T$ such that $v_j > 0$, $\sum_{j=1}^{n} |v_j| = 1$ and $Av = \lambda v$,

(iv) there is a unique positive left-eigenvector $u = (u_1, \ldots, u_k)$ such that $u_j > 0$, $\sum_{j=1}^{n} |u_j| = 1$ and $uA = \lambda u$,

(v) eigenvectors 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.

By the Perron-Frobenius theorem there exists a unique maximal eigenvalue $\lambda$ for $A$ with corresponding left and right eigenvalues $u = (u_1, \ldots, u_k)$ and $v = (v_1, \ldots, v_k)$, respectively. Define

$$P_{i,j} = \frac{A_{i,j}v_j}{\lambda v_i}$$

$$p_i = \frac{u_i v_i}{c}$$

where $c = \sum_{i=1}^{k} u_i v_i$. Then $P$ is a stochastic matrix and $pP = p$. Thus $P$ defines a Markov measure $\mu$ on $\Sigma_A$. We call $\mu$ the Parry measure.

**Theorem 8.23**

Let $\sigma : \Sigma_A \to \Sigma_A$ be a two-sided shift of finite type defined by an irreducible matrix $A$. Then the Parry measure is the unique measure of maximal entropy.

**Proof.** Omitted. See Walters. \[\square\]

§8.5.3 References

Topological entropy was first defined (using open covers) in:


The definition using spanning and separated sets is due to Bowen:


The variational principle is due to Walters and can be found in
The Parry measure was first described in:


§8.5.4 Exercises

Exercise 8.1
Let $D = \{0\} \cup \{1/n\}$ and let $X = D^\mathbb{Z}$. Let $\sigma : X \to X$ be the shift map. Let $\mu_n$ be the measure which is the direct product of the measure on $D$ which assigns measure $1/2$ to each of the atoms $1/n, 1/n + 1$. Show that $\mu_n \in M(X,T)$ and that $\sigma$, with respect to $\mu_n$, is isomorphic to the full 2-shift. Show that $\mu_n \rightharpoonup \delta_{0\infty}$ where $\delta_{0\infty}$ is the Dirac $\delta$-measure supported on $(\ldots, 0, 0, 0, \ldots)$. Show that the entropy map is not upper semi-continuous.

Exercise 8.2
Show that the topological entropy of a homeomorphism of a circle is 0.

Exercise 8.3
Two continuous transformations $T_i$ of metric spaces $X_i$, $i = 1, 2$, are said to be topologically conjugate if there exists a homeomorphism $\phi : X_1 \to X_2$ such that $T_2 \phi = \phi T_1$. Show that topological entropy is an invariant of topological conjugacy.

Exercise 8.4
Show that $h_{\text{top}}(T^m) = mh_{\text{top}}(T)$ for $m \geq 1$. If $T$ is a homeomorphism show that $h_{\text{top}}(T) = h_{\text{top}}(T^{-1})$.

Exercise 8.5
Prove Lemma 8.20. (Hint: there are two cases. When $h_{\text{top}}(T) = \infty$ show that if $T$ has a unique measure of maximal entropy then it is uniquely ergodic. When $h_{\text{top}}(T) < \infty$ show that $M_{\text{max}}(X,T)$ is convex and that the extremal points in $M_{\text{max}}(X,T)$ are precisely the ergodic measures in $M_{\text{max}}(X,T)$.)