6. Continuous transformations of compact metric spaces

§6.1 Introduction

The context of the previous two lectures was that of a measure-preserving transformation on a probability space. That is, we specified in advance the measure and sought to prove that a given transformation preserved that measure. In this lecture, we shift our focus slightly and consider, for a given transformation $T : X \to X$, the space $M(X, T)$ of all probability measures that are invariant under T. In order to equip $M(X, T)$ with some structure we will need to assume that the underlying space X is itself equipped with some additional structure more specific than merely being a measure space; throughout this lecture we will work in the context of X be a compact metric space and T being a continuous transformation.

§6.2 Probability measures on compact metric spaces

Let X be a compact metric space equipped with the Borel σ-algebra $\mathcal{B}$. (Recall that the Borel σ-algebra is the smallest σ-algebra that contains all the open subsets of X.)

Let $C(X, \mathbb{R}) = \{f : X \to \mathbb{R} \mid f \text{ is continuous}\}$ denote the space of real-valued continuous functions defined on X. Define the *uniform norm* of $f \in C(X, \mathbb{R})$ by

$$\|f\| = \sup_{x\in X} |f(x)|.$$

With this norm $C(X, \mathbb{R})$ is a Banach space.

An important property of $C(X, \mathbb{R})$ that will prove to be useful later on is that it is *separable*, that is, it contains a countable dense subset.

Let $M(X)$ denote the set of all Borel probability measures on $(X, \mathcal{B})$. The following simple fact will be useful later on.

Proposition 6.1
The space $M(X)$ is convex: if $\mu_1, \mu_2 \in M(X)$ and $0 \le \alpha \le 1$ then $\alpha\mu_1 + (1-\alpha)\mu_2 \in M(X)$.

Proof. This is immediate from the definition of a measure. □

It will be very important to have a sensible notion of convergence in $M(X)$; this is called *weak* convergence*. We say that a sequence of probability measures μ_n *weak* converges* to μ, as $n \to \infty$ if, for every $f \in C(X, \mathbb{R})$,

$$\int f\, d\mu_n \to \int f\, d\mu, \quad \text{as } n \to \infty.$$

If μ_n weak* converges to μ then we write $\mu_n \rightharpoonup \mu$. We can make $M(X)$ into a metric space compatible with this definition of convergence by choosing a countable dense subset $\{f_n\}_{n=1}^{\infty} \subset C(X)$ and, for $\mu_1, \mu_2 \in M(X)$, and setting

$$\rho(\mu_1, \mu_2) = \sum_{n=1}^{\infty} \frac{1}{2^n \|f_n\|_\infty} \left| \int f_n \, d\mu_1 - \int f_n \, d\mu_2 \right|.$$

It is easy to check that $\mu_n \rightharpoonup \mu$ if and only if $\rho(\mu_n, \mu) \to 0$.

However, we will not need to work with a particular metric: what is important is the definition of convergence.

Remark Note that with this definition it is not necessarily true that $\mu_n(B) \to \mu(B)$, as $n \to \infty$, for $B \in \mathcal{B}$.

Remark There is a continuous embedding of X in $M(X)$ given by the map $X \to M(X) : x \mapsto \delta_x$, where δ_x is the *Dirac measure* at x:

$$\delta_x(A) = \begin{cases} 1 & \text{if } x \in A, \\ 0 & \text{if } x \notin A, \end{cases}$$

(so that $\int f \, d\delta_x = f(x)$).

§6.3 The Riesz Representation Theorem

Let $\mu \in M(X)$ be a Borel probability measure. Then we can think of μ as a functional that acts on $C(X, \mathbb{R})$, namely

$$C(X, \mathbb{R}) \to \mathbb{R} : f \mapsto \int f \, d\mu.$$

We will often write $\mu(f)$ for $\int f \, d\mu$.

Notice that this map enjoys several natural properties:

(i) the functional defined by μ is linear:

$$\mu(\lambda_1 f_1 + \lambda_2 f_2) = \lambda_1 \mu(f_1) + \lambda_2 \mu(f_2)$$

where $\lambda_1, \lambda_2 \in \mathbb{R}$ and $f_1, f_2 \in C(X, \mathbb{R})$.

(ii) the functional defined by μ is continuous: i.e. if $f_n \in C(X, \mathbb{R})$ and $f_n \to f$ then $\mu(f_n) \to \mu(f)$.

(ii') the functional defined by μ is bounded: i.e. if $f \in C(X, \mathbb{R})$ then $|\mu(f)| \leq \|f\|_\infty$.

(iii) if $f \geq 0$ then $\mu(f) \geq 0$ (i.e. the map μ is *positive*);

(iv) consider the function 1 defined by $1(x) \equiv 1$ for all x; then $\mu(1) = 1$ (i.e. the map μ is *normalised*).

Remark It is well-known that a linear functional is continuous if and only if it is bounded. Thus in the presence of (i), we have that (ii) is equivalent to (ii').

The Riesz Representation Theorem says that the above properties characterise all Borel probability measures on X. That is, if we have a map $w : C(X, \mathbb{R}) \to \mathbb{R}$ that satisfies the above four properties, then w must be given by integrating with respect to a Borel probability measure. This will be a very useful method of constructing measures: we need only construct continuous positive normalised linear functionals.

Theorem 6.2 (Riesz Representation Theorem)
Let $w : C(X, \mathbb{R}) \to \mathbb{R}$ be a functional such that:

(i) w is bounded: i.e. for all $f \in C(X, \mathbb{R})$ we have $|w(f)| \le \|f\|_\infty$;

(ii) w is linear: i.e. $w(\lambda_1 f_1 + \lambda_2 f_2) = \lambda_1 w(f_1) + \lambda_2 w(f_2)$;

(iii) w is positive: i.e. if $f \ge 0$ then $w(f) \ge 0$;

(iv) w is normalised: i.e. $w(1) = 1$.

Then there exists a Borel probability measure $\mu \in M(X)$ such that

$$w(f) = \int f \, d\mu.$$

Moreover, μ is unique.

Thus the Riesz Representation Theorem allows us to identify $M(X)$ with the intersection of the positive cone and the unit ball in the dual space of $C(X, \mathbb{R})$.

Remark The Riesz Representation Theorem in the context of Hilbert spaces is often stated as follows. Let X be a (real) Hilbert space with inner product $\langle \cdot, \cdot \rangle$. Let X^* denote the dual space of X, namely the space of continuous linear functionals $X \to \mathbb{R}$. Then there is a natural isomorphism between X and X^* given by $x \mapsto \langle x, \cdot \rangle$. The Riesz Representation Theorem above says that the probability measures are naturally isomorphic to the intersection of the positive cone with the unit ball in the dual space of $C(X, \mathbb{R})$. More generally, if we allow 'signed' measures (i.e. measures that can take negative as well as positive values) then we obtain a natural isomorphism between the space of signed measures and the dual space of $C(X, \mathbb{R})$.

Corollary 6.3
The space $M(X)$ is weak* compact.

Proof. This is a restatement of Alaoglu's theorem: if X is a Banach space then the unit ball in the dual space X^* is weak* compact.

By the Riesz Representation Theorem, we can regard $M(X)$ as the intersection of the unit ball in the dual space $C(X,\mathbb{R})^*$ with the positive cone. The unit ball in $C(X,\mathbb{R})^*$ is weak* compact by Alaoglu's theorem. Moreover, the intersection of the unit ball with the positive cone is closed. Hence $M(X)$ is weak* compact. □

§6.4 Invariant measures for continuous transformations

Let X be a compact metric space equipped with the Borel σ-algebra and let $T : X \to X$ be a continuous transformation. It is clear that T is measurable.

The transformation T induces a map on the set $M(X)$ of Borel probability measures by defining $T_* : M(X) \to M(X)$ by

$$(T_*\mu)(B) = \mu(T^{-1}B).$$

It is easy to see that $T_*\mu$ is a Borel probability measure.

The following result tells us how to integrate with respect to $T_*\mu$.

Lemma 6.4
For $f \in L^1$ we have

$$\int f \, d(T_*\mu) = \int f \circ T \, d\mu.$$

Proof. From the definition, for $B \in \mathcal{B}$,

$$\int \chi_B \, d(T_*\mu) = \int \chi_B \circ T \, d\mu.$$

Thus the result also holds for simple functions. If $f \in C(X,\mathbb{R})$ is such that $f \geq 0$, we can choose an increasing sequence of simple functions f_n converging to f pointwise. We have

$$\int f_n \, d(T_*\mu) = \int f_n \circ T \, d\mu$$

and, applying the Monotone Convergence Theorem to each side, we obtain

$$\int f \, d(T_*\mu) = \int f \circ T \, d\mu.$$

The result extends to an arbitrary $f \in L^1$ by considering positive and negative parts. □

Recall that a measure μ is said to be T-invariant if $\mu(T^{-1}B) = \mu(B)$ for all $B \in \mathcal{B}$. Hence μ is T-invariant if and only if $T_*\mu = \mu$. Write

$$M(X,T) = \{\mu \in M(X) \mid T_*\mu = \mu\}$$

to denote the space of all T-invariant Borel probability measures.

The following result gives a useful criterion for checking whether a measure is T-invariant.

Lemma 6.5
Let $T : X \to X$ be a continuous mapping of a compact metric space. The following are equivalent:

(i) $\mu \in M(X,T)$;

(ii) for all $f \in C(X,\mathbb{R})$

$$\int f \circ T \, d\mu = \int f \, d\mu. \tag{6.1}$$

Proof. We prove (i) implies (ii). Suppose $\mu \in M(X,T)$. Then, by Lemma 6.4, for any $f \in C(X,\mathbb{R})$ we have

$$\int f \circ T \, d\mu = \int f \, d(T_*\mu) = \int f \, d\mu.$$

Conversely, Lemma 6.4 allows us to write (6.1) as: $\mu(f) = (T_*\mu)(f)$ for all $f \in C(X,\mathbb{R})$. Hence μ and $T_*\mu$ determine the same linear functional on $C(X,\mathbb{R})$. By uniqueness in the Riesz Representation theorem we have $T_*\mu = \mu$. □

§6.5 Existence of invariant measures

Given a continuous mapping $T : X \to X$ of a compact metric space, it is natural to ask whether invariant measures necessarily exist, i.e., whether $M(X,T) \neq \emptyset$. The next result shows that this is the case.

Theorem 6.6
Let $T : X \to X$ be a continuous mapping of a compact metric space. Then there exists at least one T-invariant probability measure.

Proof. Let $\nu \in M(X)$ be a probability measure (for example, we could take ν to be a Dirac measure). Define the sequence $\mu_n \in M(X)$ by

$$\mu_n = \frac{1}{n}\sum_{j=0}^{n-1} T_*^j \nu,$$

so that, for $B \in \mathcal{B}$,

$$\mu_n(B) = \frac{1}{n}(\nu(B) + \nu(T^{-1}B) + \cdots + \nu(T^{-(n-1)}B)).$$

Since $M(X)$ is weak* compact, some subsequence μ_{n_k} converges, as $k \to \infty$, to a measure $\mu \in M(X)$. We shall show that $\mu \in M(X,T)$. By Lemma 6.5, this is equivalent to showing that

$$\int f\, d\mu = \int f \circ T\, d\mu \quad \forall f \in C(X).$$

To see this, note that

$$\begin{aligned}
\left| \int f \circ T\, d\mu - \int f\, d\mu \right| &= \lim_{k\to\infty} \left| \int f \circ T\, d\mu_{n_k} - \int f\, d\mu_{n_k} \right| \\
&= \lim_{k\to\infty} \left| \frac{1}{n_k} \int \sum_{j=0}^{n_k-1} (f \circ T^{j+1} - f \circ T^j)\, d\nu \right| \\
&= \lim_{k\to\infty} \left| \frac{1}{n_k} \int (f \circ T^{n_k} - f)\, d\nu \right| \\
&\leq \lim_{k\to\infty} \frac{2\|f\|_\infty}{n_k} = 0.
\end{aligned}$$

Therefore, $\mu \in M(X,T)$, as required. $\square$

We will need the following additional properties of $M(X,T)$.

Theorem 6.7
Let $T : X \to X$ be a continuous mapping of a compact metric space. Then $M(X,T)$ is weak compact and convex subset of $M(X)$.*

Proof. The fact that $M(X,T)$ is convex is clear from the definition.

To see that $M(X,T)$ is weak* compact it is sufficient to show that it is a weak* closed subset of the weak* compact $M(X)$. This follows easily from the definitions. Suppose that $\mu_n \in M(X,T)$ is such that $\mu_n \rightharpoonup \mu \in M(X)$. We need to show that $\mu \in M(X,T)$. To see this, observe that for any $f \in C(X,\mathbb{R})$ we have that

$$\int f \circ T\, d\mu = \lim_{n\to\infty} \int f \circ T\, d\mu_n = \lim_{n\to\infty} \int f\, d\mu_n = \int f\, d\mu.$$

$\square$

§6.6 Ergodic measures for continuous transformations

§6.6.1 Extremal points for convex sets

A point in a convex set is called an *extremal point* if it cannot be written as a non-trivial convex combination of (other) elements of the set. More precisely, μ is an extremal point of $M(X,T)$ if whenever

$$\mu = \alpha\mu_1 + (1-\alpha)\mu_2,$$

with $\mu_1, \mu_2 \in M(X,T)$, $0 < \alpha < 1$ then we have $\mu_1 = \mu_2 = \mu$.

Remarks.

(i) Let Y be the unit square

$$Y = \{(x, y) \mid 0 \le x \le 1,\ 0 \le y \le 1\} \subset \mathbb{R}^2.$$

Then the extremal points of Y are the corners $(0,0), (0,1), (1,0), (1,1)$.

(ii) Let Y be the (closed) unit disc

$$Y = \{(x, y) : x^2 + y^2 \le 1\} \subset \mathbb{R}^2.$$

Then the set of extremal points of Y is precisely the unit circle $\{(x, y) \mid x^2 + y^2 = 1\}$.

Definition. If Y is a convex set then we denote by $\mathrm{Ext}(Y)$ the set of all extremal points of Y.

We shall also need the following definition.

Definition. Let Z be a subset of a topological vector space V. The *closed convex hull* of Z is the smallest closed convex set that contains Z and is denoted by $\mathrm{Cov}(Z)$. Equivalently

$$\mathrm{Cov}(Z) = \{t_1 v_1 + \cdots + t_k v_k \mid t_j \in [0, 1], \sum_{j=1}^{k} t_j = 1, v_j \in Z\}.$$

The following theorem tells us that the set of extremal points is a large set.

Theorem 6.8 (Krein-Milman)
Let X be a topological vector space on which X^* separates points. Let $K \subset X$ be a non-empty compact convex subset. Then K is the closed convex hull of its extremal points: $K = \mathrm{Cov}(\mathrm{Ext}(K))$.

§6.6.2 Existence of ergodic measures

The next result will allow us to show that ergodic measures for continuous transformations on compact metric spaces always exist.

Theorem 6.9
The following are equivalent:

(i) the T-invariant probability measure μ is ergodic;

(ii) μ is an extremal point of $M(X, T)$.

Proof. We prove (ii) $\Rightarrow$ (i): if μ is extremal then it is ergodic. In fact, we shall prove the contrapositive. Suppose that μ is not ergodic; we show that μ is not extremal. As μ is not ergodic, there exists $B \in \mathcal{B}$ such that $T^{-1}B = B$ and $0 < \mu(B) < 1$.

Define probability measures μ_1 and μ_2 on X by

$$\mu_1(A) = \frac{\mu(A \cap B)}{\mu(B)}, \quad \mu_2(A) = \frac{\mu(A \cap (X \setminus B))}{\mu(X \setminus B)}.$$

(The condition $0 < \mu(B) < 1$ ensures that the denominators are not equal to zero.) Clearly, $\mu_1 \neq \mu_2$, since $\mu_1(B) = 1$ while $\mu_2(B) = 0$.

Since $T^{-1}B = B$, we also have $T^{-1}(X \setminus B) = X \setminus B$. Thus we have

$$\begin{aligned}
\mu_1(T^{-1}A) &= \frac{\mu(T^{-1}A \cap B)}{\mu(B)} \\
&= \frac{\mu(T^{-1}A \cap T^{-1}B)}{\mu(B)} \\
&= \frac{\mu(T^{-1}(A \cap B))}{\mu(B)} \\
&= \frac{\mu(A \cap B)}{\mu(B)} \\
&= \mu_1(A)
\end{aligned}$$

and (by the same argument)

$$\mu_2(T^{-1}A) = \frac{\mu(T^{-1}A \cap (X \setminus B))}{\mu(X \setminus B)} = \mu_2(A),$$

i.e., μ_1 and μ_2 are both in $M(X,T)$.

However, we may write μ as the non-trivial (since $0 < \mu(B) < 1$) convex combination

$$\mu = \mu(B)\mu_1 + (1 - \mu(B))\mu_2,$$

so that μ is not extremal.

We prove (i) $\Rightarrow$ (ii). Suppose that μ is ergodic and that $\mu = \alpha\mu_1 + (1 - \alpha)\mu_2$, with $\mu_1, \mu_2 \in M(X,T)$ and $0 < \alpha < 1$. We shall show that $\mu_1 = \mu$ (so that $\mu_2 = \mu$, also). This will show that μ is extremal.

If $\mu(A) = 0$ then $\mu_1(A) = 0$, so $\mu_1 \ll \mu$. Therefore the Radon-Nikodym derivative $d\mu_1/d\mu \geq 0$ exists. One can easily deduce from the statement of the Radon-Nikodym Theorem that $\mu_1 = \mu$ if and only if $d\mu_1/d\mu = 1$ μ-a.e. We shall show that this is indeed the case by showing that the sets where, respectively, $d\mu_1/d\mu < 1$ and $d\mu_1/d\mu > 1$ both have μ-measure zero.

Let

$$B = \left\{ x \in X : \frac{d\mu_1}{d\mu}(x) < 1 \right\}.$$

Now

$$\mu_1(B) = \int_B \frac{d\mu_1}{d\mu}\, d\mu = \int_{B\cap T^{-1}B} \frac{d\mu_1}{d\mu}\, d\mu + \int_{B\setminus T^{-1}B} \frac{d\mu_1}{d\mu}\, d\mu \tag{6.2}$$

and

$$\mu_1(T^{-1}B) = \int_{T^{-1}B} \frac{d\mu_1}{d\mu}\, d\mu = \int_{B\cap T^{-1}B} \frac{d\mu_1}{d\mu}\, d\mu + \int_{T^{-1}B\setminus B} \frac{d\mu_1}{d\mu}\, d\mu. \tag{6.3}$$

As $\mu_1 \in M(X,T)$, we have that $\mu_1(B) = \mu_1(T^{-1}B)$. Hence comparing the last summand in both (6.2) and (6.3) we obtain

$$\int_{B\setminus T^{-1}B} \frac{d\mu_1}{d\mu}\, d\mu = \int_{T^{-1}B\setminus B} \frac{d\mu_1}{d\mu}\, d\mu. \tag{6.4}$$

In fact, these integrals are taken over sets of the same μ-measure:

$$\begin{aligned}
\mu(T^{-1}B \setminus B) &= \mu(T^{-1}B) - \mu(T^{-1}B \cap B) \\
&= \mu(B) - \mu(T^{-1}B \cap B) \\
&= \mu(B \setminus T^{-1}B).
\end{aligned}$$

Note that on the left-hand side of (6.4), the integrand $d\mu_1/d\mu < 1$. However, on the right-hand side of (6.4), the integrand $d\mu_1/d\mu \geq 1$. Thus we mus have that $\mu(B\setminus T^{-1}B) = \mu(T^{-1}B\setminus B) = 0$, which is to say that $\mu(T^{-1}B \triangle B) = 0$, i.e. $T^{-1}B = B$ μ-a.e. Therefore, since μ is ergodic, we have that $\mu(B) = 0$ or $\mu(B) = 1$.

We can rule out the possibility that $\mu(B) = 1$ by observing that if $\mu(B) = 1$ then

$$1 = \mu_1(X) = \int_X \frac{d\mu_1}{d\mu}\, d\mu = \int_B \frac{d\mu_1}{d\mu}\, d\mu < \mu(B) = 1,$$

a contradiction. Therefore $\mu(B) = 0$.

If we define

$$C = \left\{ x \in X \mid \frac{d\mu_1}{d\mu}(x) > 1 \right\}$$

then repeating essentially the same argument gives $\mu(C) = 0$.

Hence

$$\begin{aligned}
\mu\left\{ x \in X \mid \frac{d\mu_1}{d\mu}(x) = 1 \right\} &= \mu(X \setminus (B \cup C)) \\
&= \mu(X) - \mu(B) - \mu(C) = 1,
\end{aligned}$$

i.e., $d\mu_1/d\mu = 1$ μ-a.e. Therefore $\mu_1 = \mu$, as required. $\square$

Theorem 6.10
Let $T : X \to X$ be a continuous mapping of a compact metric space. Then there exists at least one ergodic measure in $M(X,T)$.

Proof. We know that $M(X,T)$ is a convex, weak* compact subset of $M(X)$. By the Krein-Milman Theorem, $M(X,T)$ is equal to its convex hull. As $M(X,T)$ is non-empty, this means that there exist extremal points. By Theorem 6.9 these are precisely the ergodic measures. □

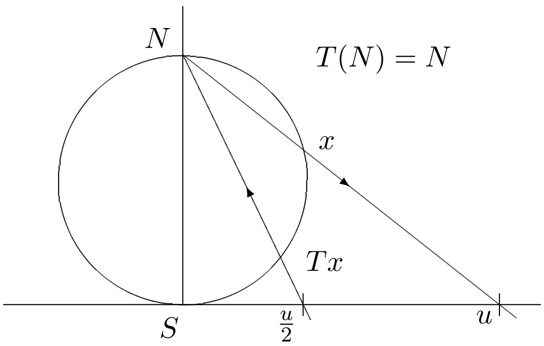
§6.7 An example: the North-South map

For many dynamical systems there exist uncountably many different ergodic measures. This is the case for the doubling map, shifts of finite type, toral automorphisms, etc. Here we give an example of a dynamical system $T : X \to X$ in which one can construct $M(X,T)$ and $E(X,T)$ explicitly.

Let $X \subset \mathbb{R}^2$ denote the circle of radius 1 centred at $(0,1)$ in $\mathbb{R}^2$. Call $(0,2)$ the North Pole (N) and $(0,0)$ the South Pole (S) of X. Define a map $\phi : X \setminus \{N\} \to \mathbb{R} \times \{0\}$ by drawing a straight line through N and x and denoting by $\phi(x)$ the unique point on the x-axis that this line crosses (this is just stereographic projection of the circle). Define $T : X \to X$ by

$$T(x) = \begin{cases} \phi^{-1}\left(\frac{1}{2}\phi(x)\right) & \text{if } x \in X \setminus \{N\}, \\ N & \text{if } x = N. \end{cases}$$

Hence $T(N) = N$, $T(S) = S$ and if $x \neq N, S$ then $T^n(x) \to S$ as $n \to \infty$.



We call T the *north-south map.*

Clearly both N and S are fixed points for T. Hence δ_N and δ_S (the Dirac delta measures at N, S, respectively) are T-invariant. It is easy to see that both δ_N and δ_S are ergodic.

Now let $\mu \in M(X,T)$ be an invariant measure. We claim that μ assigns zero measure to the set $X \setminus \{N,S\}$. Let $x \in X$ be any point in the right semi-circle (for example, take $x = (1,1) \in \mathbb{R}^2$) and consider the arc I of semi-circle from x to $T(x)$. Then $\cup_{n=-\infty}^{\infty} T^{-n} I$ is a disjoint union of arcs of semi-circle and, moreover, is equal to the entire right semi-circle. Now

$$\mu\left(\bigcup_{n=-\infty}^{\infty} T^{-n}I\right) = \sum_{n=-\infty}^{\infty} \mu(T^{-n}I) = \sum_{n=-\infty}^{\infty} \mu(I)$$

and the only way for this to be finite is if $\mu(I) = 0$. Hence μ assigns zero measure to the entire right semi-circle. Similarly, μ assigns zero measure to the left semi-circle.

Hence μ is concentrated on the two points N, S, and so must be a convex combination of the Dirac delta measures δ_N and δ_S. Hence

$$M(X,T) = \{\alpha\delta_N + (1-\alpha)\delta_S \mid \alpha \in [0,1]\}$$

and the ergodic measures are the extremal points of $M(X,T)$, namely δ_N, δ_S.

§6.8 Unique Ergodicity

We conclude this lecture by looking at the case where $T : X \to X$ has a *unique* invariant probability measure.

Definition. Let $T : X \to X$ be a continuous transformation of a compact metric space X. If there is a *unique* T-invariant probability measure then we say that T is *uniquely ergodic*.

Remark You might wonder why such T are not instead called 'uniquely invariant'. Recall that the extremal points of $M(X,T)$ are precisely the ergodic measures. If $M(X,T)$ consists of just one measure then that measure is extremal, and so must be ergodic.

Unique ergodicity implies the following strong convergence result.

Theorem 6.11 (Oxtoby's ergodic theorem)
Let X be a compact metric space and let $T : X \to X$ be a continuous transformation. The following are equivalent:

(i) T is uniquely ergodic;

(ii) for each $f \in C(X,\mathbb{R})$ there exists a constant $c(f)$ such that

$$\frac{1}{n}\sum_{j=0}^{n-1} f(T^j x) \to c(f),$$

uniformly for $x \in X$, as $n \to \infty$.

Remark Our intuition behind Oxtoby's Ergodic Theorem is as follows. By Birkhoff's Ergodic Theorem, we know that

$$\frac{1}{n}\sum_{j=0}^{n-1} f(T^j x) \tag{6.5}$$

converges for μ-almost all $x \in X$, where μ is an ergodic measure. If μ_1 and μ_2 are distinct ergodic measures then they must be mutually singular (this is a simple exercise). Thus, for example, the set of zero μ_1-measure for which (6.5) fails to converge to $\int f \, d\mu_1$ contains a set of full μ_2-measure for which it converges to $\int f \, d\mu_2$ (it also contains sets of full measure for other ergodic measures for which (6.5) converges to the integral of f, along with the set of points for which (6.5) fails to converge at all. If T is uniquely ergodic, then the only possibility for a limit for (6.5) as $n \to 0$ is $\int f \, d\mu$ where μ is the unique invariant measure; and hence one should expect this to hold everywhere, assuming (6.5) converges.

Proof. (ii) $\Rightarrow$ (i): Suppose that μ, ν are T-invariant probability measures; we shall show that $\mu = \nu$.

Integrating the expression in (ii), we obtain

$$\begin{aligned}
\int f \, d\mu &= \lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} \int f \circ T^j \, d\mu \\
&= \int \lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} f \circ T^j \, d\mu \\
&= \int c(f) \, d\mu = c(f),
\end{aligned}$$

and, by the same argument

$$\int f \, d\nu = c(f).$$

Therefore

$$\int f \, d\mu = \int f \, d\nu \quad \forall f \in C(X)$$

and so $\mu = \nu$ (by the Riesz Representation Theorem).

(i) $\Rightarrow$ (ii): Let $M(X,T) = \{\mu\}$. If (ii) is true, then, by the Dominated Convergence Theorem, we must necessarily have $c(f) = \int f \, d\mu$. Suppose that (ii) is false. Then we can find $f \in C(X)$ and sequences $n_k \in \mathbb{N}$ and $x_k \in X$ such that

$$\lim_{k\to\infty} \frac{1}{n_k} \sum_{j=0}^{n_k-1} f(T^j x_k) \neq \int f \, d\mu.$$

For each $k \geq 1$, define a measure $\nu_k \in M(X)$ by

$$\nu_k = \frac{1}{n_k} \sum_{j=0}^{n_k - 1} T_*^j \delta_{x_k},$$

so that

$$\int f \, d\nu_k = \frac{1}{n_k} \sum_{j=0}^{n_k - 1} f(T^j x_k).$$

By the following the proof of Theorem 6.6, it is easy to see that ν_k has a subsequence which converges weak* to a measure $\nu \in M(X,T)$. In particular, we have

$$\int f \, d\nu = \lim_{k \to \infty} \int f \, d\nu_k \neq \int f \, d\mu.$$

Therefore, $\nu \neq \mu$, contradicting unique ergodicity. □

§6.8.1 Example: The Irrational Rotation

Let $X = \mathbb{R}/\mathbb{Z}$, $T : X \to X : x \mapsto x + \alpha \bmod 1$, α irrational. Then T is uniquely ergodic (and $\mu =$ Lebesgue measure is the unique invariant probability measure).

Proof. Let m be an arbitrary T-invariant probability measure; we shall show that $m = \mu$.

Write $e_k(x) = e^{2\pi i k x}$. Then

$$\begin{aligned} \int e_k(x) \, dm &= \int e_k(Tx) \, dm \\ &= \int e_k(x + \alpha) \, dm \\ &= e^{2\pi i k \alpha} \int e_k(x) \, dm. \end{aligned}$$

Since α is irrational, if $k \neq 0$ then $e^{2\pi i k \alpha} \neq 1$ and so

$$\int e_k \, dm = 0. \tag{6.6}$$

Let $f \in C(X)$ have Fourier series $\sum_{k=-\infty}^{\infty} a_k e_k$, so that $a_0 = \int f \, d\mu$. For $n \geq 1$, we let σ_n denote the average of the first n partial sums. Then $\sigma_n \to f$ uniformly as $n \to \infty$. Hence

$$\lim_{n \to \infty} \int \sigma_n \, dm = \int f \, dm.$$

However using (6.6), we may calculate that

$$\int \sigma_n \, dm = a_0 = \int f \, d\mu.$$

Thus we have that $\int f \, dm = \int f \, d\mu$, for every $f \in C(X)$, and so $m = \mu$. □

§6.9 Appendix: Radon-Nikodym derivatives

In the above we used Radon-Nikodym derivatives. Here we give a brief statement of the required results.

Definition. Let μ be a measure on $(X, \mathcal{B})$. We say that a measure ν is *absolutely continuous with respect to* μ and write $\nu \ll \mu$ if $\nu(B) = 0$ whenever $\mu(B) = 0$, $B \in \mathcal{B}$.

Remark Thus ν is absolutely continuous with respect to μ if sets of μ-measure zero also have ν-measure zero (but there may be more sets of ν-measure zero). For example, let $f \in L^1(X, \mathcal{B}, \mu)$ be non-negative and define a measure ν by

$$\nu(B) = \int_B f \, d\mu.$$

Then $\nu \ll \mu$.

The following theorem says that, essentially, all absolutely continuous measures occur in this way.

Theorem 6.12 (Radon-Nikodym)
Let $(X, \mathcal{B}, \mu)$ be a probability space. Let ν be a measure defined on $\mathcal{B}$ and suppose that $\nu \ll \mu$. Then there is a non-negative measurable function f such that

$$\nu(B) = \int_B f \, d\mu, \quad \text{for all } B \in \mathcal{B}.$$

Moreover, f is unique in the sense that if g is a measurable function with the same property then $f = g$ μ-a.e.

If $\nu \ll \mu$ then it is customary to write $d\nu/d\mu$ for the function given by the Radon-Nikodym theorem, that is

$$\nu(B) = \int_B \frac{d\nu}{d\mu} \, d\mu.$$

The following (easily-proved) relations indicate why this notation is used.

(i) If $\nu \ll \mu$ and f is a μ-integrable function then

$$\int f \, d\nu = \int f \frac{d\nu}{d\mu} \, d\mu.$$

(ii) If $\nu_1, \nu_2 \ll \mu$ then

$$\frac{d(\nu_1 + \nu_2)}{d\mu} = \frac{d\nu_1}{d\mu} + \frac{d\nu_2}{d\mu}.$$

(iii) If $\lambda \ll \nu \ll \mu$ then

$$\frac{d\lambda}{d\mu} = \frac{d\lambda}{d\nu} \frac{d\nu}{d\mu}.$$

§6.10 Extension to arbitrary measure spaces

Throughout this lecture, the setting has been that of a continuous transformation of a compact metric space equipped with the Borel σ-algebra. In previous lectures, the setting has been that of a measurable transformation of a probability space $(X, \mathcal{B}, \mu)$. One can often embed the latter in the former, as follows.

Recall that the Borel σ-algebra on $[0, 1]$ is the smallest σ-algebra that contains the open sets. There is another, larger, σ-algebra that is often of use, namely the Lebesgue σ-algebra. Recall that a set B is *Lebesgue measurable* if for all sets $A \subset [0, 1]$ we have

$$\lambda^*(A) = \lambda^*(A \cap B) + \lambda^*(A \cap B^c)$$

where λ^* denotes the Lebesgue outer measure. The collection of all Lebesgue measurable sets forms a σ-algebra which contains, but is strictly larger than, the Borel σ-algebra. Lebesgue outer measure restricts to a measure, Lebesgue measure, on the Lebesgue σ-algebra. Any Lebesgue measurable set is almost everywhere equal to a Borel set.

A finite measure space $(X, \mathcal{B}, \mu)$ is said to be a *Lebesgue measure space* if there is a bi-measurable measure-preserving bijection ϕ from $(X, \mathcal{B}, \mu)$ to the unit interval $[0, 1]$ equipped with the Lebesgue σ-algebra and Lebesgue measure, together with a countable number of point masses. (Note that ϕ being measurable does not imply that ϕ^{-1} is measurable, and this is why we have to say 'bi-measurable'.) It is worth noting that one has to go to a great deal of pathological effort to construct a non-Lebesgue measure space.

Proposition 6.13
Let $(X, \mathcal{B}, \mu)$ be a Lebesgue measure space and let $T : X \to X$ be a measure-preserving transformation. Then there exists a compact metric space X', a continuous transformation $T' : X' \to X'$, a measure μ' on the completion $\bar{\mathcal{B}}(X)$ of the Borel σ-algebra of X', and a measure-preserving bimeasurable isomorphism $V : (X, \mathcal{B}, \mu) \to (X', \bar{\mathcal{B}}(X), \mu')$ between T and T'.

Hence, up to measurable isomorphism, there is often no loss in assuming that a measure-preserving transformation of a probability space is a continuous transformation of a compact metric space, and the invariant measure is a probability measure on the completion of the Borel σ-algebra.

§6.11 References

A particularly readable account of the material in this lecture can be found in:

P. Walters, *An introduction to ergodic theory*, Springer, Berlin, 1982.

The connections between the space of ergodic measures and the extremal points go much deeper than the material presented above using ideas from convexity theory. The standard reference on convex sets is

R. Phelps, *Lectures on Choquet's Theorem*, van Nostrand, New York, 1966.

Precise statements, together with proofs, about embedding Lebesgue measurable spaces in compact metric spaces can be found in

H. Furstenberg, *Disjointness in ergodic theory, minimal sets, and a problem in Diophantine approximation*, Math. Systems Theory **1** (1967), 1–49.

§6.12 Exercises

Exercise 6.1
Give an example to show that $\mu_n \rightharpoonup \mu$ does not necessarily imply that $\mu_n(B) \to \mu(B)$ for all $B \in \mathcal{B}$.

Exercise 6.2
Show that the map $\delta : X \to M(X) : x \mapsto \delta_x$ is continuous in the weak topology. Show that the map $T_* : M(X) \to M(X)$ is continuous in the weak* topology. (Hint: This is really just unravelling the underlying definitions.)*

Exercise 6.3
Let X be a compact metric space. For $\mu \in M(X)$ define

$$\|\mu\| = \sup_{f \in C(X), \|f\|_\infty \leq 1} \left| \int f \, d\mu \right|.$$

We say that μ_n converges strongly to μ if $\|\mu_n - \mu\| \to 0$ as $n \to \infty$. The topology this determines is called the strong topology (or the operator topology).

(i) Show that if $\mu_n \to \mu$ strongly then $\mu_n \rightharpoonup \mu$ in the weak topology.*

(ii) Show that $X \hookrightarrow M(X) : x \mapsto \delta_x$ is not *continuous in the strong topology.*

(iii) Prove that $\|\delta_x - \delta_y\| = 2$ if $x \neq y$. (You may use Urysohn's Lemma: Let A and B be disjoint closed subsets of a metric space X. Then there is a continuous function $f \in C(X, \mathbb{R})$ such that $0 \leq f \leq 1$ on X while $f \equiv 0$ on A and $f \equiv 1$ on B.)

Hence prove that $M(X)$ is not compact in the strong topology when X is infinite.

Exercise 6.4
Let $T : X \to X$ be a continuous transformation of a compact metric space X.

(i) Suppose that μ is an ergodic measure for T. Prove that there exists a set $Y \in \mathcal{B}$ with $\mu(Y) = 1$ such that

$$\frac{1}{n}\sum_{j=0}^{n-1} f(T^j x) \to \int f \, d\mu$$

for all $x \in Y$, for all $f \in C(X, \mathbb{R})$. (That is, the set of full measure for which convergence holds in the Birkhoff Ergodic Theorem can be chosen to work simultaneously for all continuous functions.)

(ii) Let $\mu \in M(X, T)$ be a T-invariant measure. Prove that μ is ergodic if and only if

$$\frac{1}{n}\sum_{j=0}^{n-1} \delta_{T^j x} \rightharpoonup \mu \text{ for } \mu - \text{a.e } x \in X.$$

(iii) Hence give another proof that the only ergodic measures for the North-South map are δ_N, δ_S.

Exercise 6.5
Let $T : X \to X$ be a continuous transformation of a compact metric space X. Let $B = \{g \circ T - g \mid g \in C(X, \mathbb{R})\}$ (this is often called the space of coboundaries of T) and let $\bar{B} \subset C(X, \mathbb{R})$ denote its closure in the uniform topology. Prove that T is uniquely ergodic if and only if $C(X, \mathbb{R}) = \bar{B} \oplus \mathbb{R}$ (here $\mathbb{R}$ denotes the space of constant functions). (Hint: use the Hahn-Banach Theorem and the Hahn Decomposition Theorem.)