§9. The Birkhoff Ergodic Theorem

Given any function $f : X \to X$, any orbit $f : x_0 \mapsto x_1 \mapsto \cdots$, and any real valued function $\varphi : X \to$, we can try to form the limit

$$A(x_0) \;=\; \lim_{n\to\infty} \frac{1}{n}\Big(\varphi(x_0)+\varphi(x_1)+\cdots+\varphi(x_{n-1})\Big) \;.$$

If this limit exists, it is called the *time average* of φ over the forward orbit of x_0. I will use the notations

$$S_0^n(\varphi) \;=\; S_0^n(\varphi\,,\,x_0) \;=\; \varphi(x_0)+\varphi(x_1)+\cdots\varphi(x_{n-1})\;,$$

and

$$A(x) \;=\; A_\varphi(x) \;=\; \lim_{n\to\infty} S_0^n(x)/n$$

when this limit exists.

Now suppose that $(X, \mathcal{A}, \mu)$ is a finite measure space, and that $\varphi : (X\,,\,\mathcal{A}) \to (\quad, \mathcal{B}\ )$ is an integrable function. Then the *space average* of φ is defined simply to be the ratio $(\int \varphi\, d\mu)/\mu(X)$ of the integral $\int_X \varphi(x)\, d\mu(x)$ to the total measure $\mu(X)$.

A basic question in dynamics is the problem of understanding when space averages are equal to time averages. The object of this exposition is to prove the following fundamental result, commonly known as the *Birkhoff Ergodic Theorem*, although it applies equally well to non-ergodic transformations.

Theorem 9.1 (Birkhoff). *If $\mu(X)$ is finite, then for any integrable function $\varphi : X \to$ the* time average

$$A_\varphi(x_0) \;=\; \lim_{n\to\infty} \frac{\varphi(x_0)+\varphi(x_1)+\cdots+\varphi(x_{n-1})}{n}$$

is defined for all x_0 outside of a set N_φ of measure $\mu(N_\varphi)=0$. Furthermore A_φ is integrable, with $A_\varphi \circ f = A_\varphi$ wherever it is defined, and with

$$\int_X A_\varphi \;=\; \int_X \varphi\;.$$

§9A. The Proof. The argument will be based on the following subsidiary result. For this statement, we no longer need the hypothesis that $\mu(X) < \infty$, although we continue to assume that μ is σ-finite.

Theorem 9.2. Existence of "Positive" Orbits. *Let $f : X \to X$ be measure preserving, and let $\varphi : X \to$ be integrable with $\int_X \varphi > 0$. Then there exists an orbit $x_0 \mapsto x_1 \mapsto \cdots$ which satisfies the inequality*

$$\varphi(x_0) \;+\; \varphi(x_1) \;+\; \cdots \;+\; \varphi(x_{n-1}) \;>\; 0$$

for every $n \ge 1$.

The hypothesis that $\mu(X) < \infty$ is essential for 9.1, but unneeded for 9.2. As an example, for the translation $f(x) = x+1$ on the real line, with $\varphi(x) = e^{-x^2} cos(x)$. Theorem 9.2 applies perfectly well, but the last equation in Theorem 9.1 fails, since $A_\varphi(x)$ is identically zero.

The proof of of 9.2 will be by contradiction. Suppose to the contrary that the following is satisfied:

Hypothesis A. For every $x_0 \in X$ there exists an integer $n \geq 1$ so that

$$\varphi(x_0) + \varphi(x_1) + \cdots + \varphi(x_{n-1}) \ \leq \ 0 \ .$$

Then we must show that[1] $\int_X \varphi \leq 0$. Following a method introduced by [Katznelson and Weiss], we first consider the special case where the following stronger condition is satisfied.

Hypothesis B. There exists a constant $k \geq 1$ so that, for every $x_0 \in X$ there exists an integer $1 \leq n \leq k$ with $\varphi(x_0) + \varphi(x_1) + \cdots + \varphi(x_{n-1}) \ \leq \ 0$.

The proof in this special case proceeds as follows. Fixing some orbit $x_0 \mapsto x_1 \mapsto \cdots$, we will first prove the inequality

$$\sum_{0 \leq j < N} \varphi(x_j) \quad \leq \quad \sum_{N-k \leq j < N} |\varphi(x_j)| \ , \tag{9:1}$$

where N can be any positive integer. It will be convenient to introduce the notation

$$S_p^q(\varphi) \ = \ \varphi(x_p) + \varphi(x_{p+1}) + \varphi(x_{p+2}) + \cdots + \varphi(x_{q-1}) \ = \ \sum_{p \leq j < q} \varphi(x_j) \ .$$

Then Hypothesis B guarantees that for each integer $p \geq 0$ we can find an integer q with $p < q \leq p+k$, and with $S_p^q(\varphi) \leq 0$. Hence, starting with $p_0 = 0$, we can inductively construct a sequence of integers $0 = p_0 < p_1 < p_2 < \cdots$ with $p_{i+1} \leq p_i + k$, and with $S_{p_i}^{p_{i+1}}(\varphi) \leq 0$. Summing this last inequality for $0 \leq i < j$, it follows also that

$$S_0^{p_j}(\varphi) \ \leq \ 0 \ .$$

Now given an arbitrarily large integer N , we can choose p_j so that $N - k \leq p_j \leq N$. It follows that

$$S_0^N(\varphi) \ = \ S_0^{p_j}(\varphi) + S_{p_j}^N(\varphi) \ \leq \ 0 + S_{p_j}^N(|\varphi|) \ \leq \ S_{N-k}^N(|\varphi|) \ ,$$

which proves (9 : 1). Now considering both sides of (9 : 1) as functions of x_0 and integrating over $x_0 \in X$, since the integral of $\varphi \circ f^{\circ j}$ is equal to the integral of φ by (8 : 1), we obtain

$$N \int_X \varphi \ \leq \ k \int_X |\varphi| \ .$$

Since k is fixed and N can be arbitrarily large, this proves that $\int_X \varphi \leq 0$, as required.

To prove that $\int_X \varphi \leq 0$ assuming only Hypothesis A, we consider the sequence of measurable real valued functions φ_k where

$$\varphi_k(x_0) \ = \ \begin{cases} \varphi(x_0) & \text{if there exists } \ 1 \leq n \leq k \ \text{ with } \ S_0^n(\varphi) \leq 0 \,, \\ 0 & \text{otherwise.} \end{cases}$$

Then φ_k is integrable since $|\varphi_k| \leq |\varphi|$. It is easy to check that $\varphi_1 \leq \varphi_2 \leq \varphi_3 \leq \cdots \leq \varphi$; hence $S_0^n(\varphi_k) \leq S_0^n(\varphi)$. Therefore each φ_k satisfies Hypothesis B, which implies that $\int_X \varphi_k \leq 0$. On the other hand, it follows from Hypothesis A that φ is equal to the pointwise limit of the φ_k . Therefore $\int \varphi = \lim_{k\to\infty} \int \varphi_k$ by Lebesgue's dominated convergence theorem, and it follows that $\int_X \varphi \leq 0$ also. This proves 9.2. □

[1] Changing signs, we obtain a completely equivalent statement commonly known as the "Maximal Ergodic Theorem": If for every choice of x_0 there is an $n \geq 1$ with $S_0^n(\varphi\,,\,x_0) \geq 0$, then $\int_X \varphi \geq 0$.

Proof of the Birkhoff Theorem 9.1. Although time averages may not always exist, we can certain form the *upper* and *lower* time averages

$$A^+(x) \;=\; \limsup_{n\to\infty}\; S_0^n(\varphi\,,\,x)/n\;, \qquad A^-(x) \;=\; \liminf_{n\to\infty}\; S_0^n(\varphi\,,\,x)/n\;,$$

where

$$-\infty \;\le\; A^-(x) \;\le\; A^+(x) \;\le\; +\infty\;.$$

It is not hard to check that both A^+ and A^- are measurable functions from X to $[-\infty\,,\,\infty]$, and that both are f-invariant, $A^\pm(f(x)) = A^\pm(x)$.

First consider the special case where both A^+ and A^- are bounded functions. Then they are certainly integrable since $\mu(X)$ is finite. We will show that

$$\int A^+ \;\le\; \int \varphi\;. \tag{9 : 2}$$

For otherwise, if $\int A^+ > \int \varphi$, then we could choose $\epsilon > 0$ so that

$$\int A^+ \;>\; \int(\varphi+\epsilon)\;.$$

(Here we again make use of the hypothesis $\mu(X) < \infty$.) Hence by 9.2 we could find an orbit $x_0 \mapsto x_1 \mapsto \cdots$ with

$$S_0^n(A^+) \;>\; S_0^n(\varphi+\epsilon) \qquad \text{for every} \qquad n > 0\;.$$

Since A^+ is constant on orbits, the left side of this inequality equals $n\,A^+(x_0)$. Now, dividing by n and taking the lim sup as $n\to\infty$, this yields

$$A^+(x_0) \;\ge\; A^+(x_0) + \epsilon\;,$$

which is impossible.

Combining (9 : 2) with an analogous statement for the lower time average, we see that

$$\int A^+ \;\le\; \int \varphi \;\le\; \int A^- \;\le\; \int A^+\;.$$

Hence all three integrals are equal, and it follows easily that $A^-(x) = A^+(x)$ except on a set of measure zero. Thus the time limit $A(x)$ is well defined for almost all x. This completes the proof when the $|A^\pm|$ are bounded. In particular, it completes the proof when the function $|\varphi|$ itself is bounded.

Proof when the $A^\pm$ are not bounded. For each positive integer n, let X_n be the set of points for which

$$-n \;\le\; A^-(x) \;\le\; A^+(x) \;\le\; n\;.$$

This is a measurable f-invariant set. Hence we can apply the argument above to conclude that the limit $A(x)$ exists for almost all $x \in X_n$, and that $\int_{X_n} A = \int_{X_n} \varphi$. It follows easily that corresponding statements are true for the union X_∞ of the various X_n, that is, for the set of all x satisfying

$$-\infty \;<\; A^-(x) \;\le\; A^+(x) \;<\; +\infty\;.$$

Thus, to complete the proof of 9.1, we need only check that the functions $A^\pm$ take finite values, except on a set of measure zero. For example let N be the invariant set consisting of points x for which $A^+(x) = +\infty$. If $\mu(N) > 0$, then we could choose a finite constant

c so that $\int_N c > \int_N \varphi$. Hence there would exist an orbit $x_0 \mapsto x_1 \mapsto \cdots$ in N so that

$$n\,c \;=\; S_0^n(c) \;>\; S_0^n(\varphi) \qquad \text{for every} \qquad n \;\geq\; 1\ .$$

Dividing by n and taking the lim sup as $n \to \infty$, this would yield $c \;\geq\; A^+(x_0)$, which contradicts the hypothesis that $A^+(x_0) = \infty$. This contradiction completes the proof of 9.1. □

§**9B. Some Consequences.** The most noteworthy consequence of Theorem 9.1 is the following. Let $(X, \mathcal{A}, \mu)$ be a measure space with $0 < \mu(X) < \infty$. Recall that a measurable transformation f is defined to be ***ergodic*** if there is no splitting $X = X_1 \cup X_2$ into two disjoint f-invariant subsets of strictly positive measure.

Corollary 9.3. Ergodic Theorem. *A measure preserving transformation* $f : X \to X$ *is ergodic if and only if, for every integrable function* $\varphi : X \to$ *, the time average* $A_\varphi(x)$ *is equal to the space average* $\int_X \varphi/\mu(X)$ *for all points* x *outside of some subset* N_φ *of measure* $\mu(N_\varphi) = 0$ *.*

Proof. If f is ergodic, then for each real number c the set

$$Y_c \;=\; \{x \in X\ ;\ A_\varphi^+(x) \;<\; c\}$$

is fully invariant, and hence must have either full measure or measure zero. Since the function $c \mapsto \mu(Y_c)$ is clearly monotone, it is a step function, with a jump at one uniquely defined value $c_0 \in [-\infty, \infty]$. In other words, this dichotomy determines a Dedekind cut, with

$$\mu(Y_c) \;=\; \begin{cases} 0 & \text{for} \quad c < c_0\,, \\ \mu(X) & \text{for} \quad c > c_0\,. \end{cases}$$

Let N_φ^+ be the set of all points for which $A_\varphi^+(x)$ is different from c_0 . It follows easily that N_φ^+ is a set of measure zero, with $A_\varphi^+(x) = c_0$ for all $x \notin N_\varphi^+$. Finally, we apply Birkhoff's Theorem 9.1 to conclude that the space average $A_\varphi(x)$ is defined and equal to $A_\varphi^+(x)$ almost everywhere, and that its value c_0 almost everywhere must be equal to the space average $\int_X \varphi/\mu(X)$.

Conversely, suppose that f is *not* ergodic. Choosing a partition $X = X_1 \cup X_2$ into disjoint invariant subsets of positive measure, set φ equal to the characteristic function $\mathbf{1}_{X_1}$ which takes the value $+1$ on X_1 and the value 0 on B . Then every time average $A_\varphi(x)$ is equal to either zero or one, but the space average is equal to $\mu(X_1)/\mu(X) \neq 0\,,\,1$. □

Remark 9.4. The set N_φ of exceptional points definitely depends on the choice of integrable function φ . As an example, for any specified $z \in X$ we can choose $\varphi(x)$ to take the value $+1$ on the forward orbit of z and the value zero elsewhere. If every point of X has measure zero, then it is easy to check that the exception set N_φ consists of all points whose orbit eventually lands on the orbit of z , In particular, N_φ contains the arbitrarily specified point z .

Remark 9.5. As an immediate consequence of 9.3, we see that an f-invariant ergodic probability measure μ is uniquely determined by its measure class $[\mu]$. To compute the measure $\mu(S)$, given only f and the measure class $[\mu]$, we proceed as follows. Let φ be the characteristic function $\mathbf{1}_S$. If there is such a invariant ergodic probability measure μ in the given measure class, then the time average $A_\phi(x)$ must take some constant value for $[\mu]$-almost all points of X . This constant value is the required measure $\mu(S) = \int_X \varphi\, d\mu$.

Here is a closely related statement. Define two measures μ and ν to be *mutually singular* if we can find a measurable set S which has full measure for μ but measure zero for ν. (Thus $[\mu]$-almost every point belongs to S, but $[\nu]$-almost every point belongs to the complementary set $X \smallsetminus S$.)

Corollary 9.6. *If μ and ν are two distinct invariant and ergodic probability measures for f, then μ and ν must be mutually singular.*

Proof. Choose φ so that $\int \varphi\, d\mu \neq \int \varphi\, d\nu$. Then the time average $A_\varphi(x)$ takes one value for $[\mu]$-almost every point, but takes a different value for $[\nu]$-almost every point. □

In the special case where f is an automorphism of the measure space, there is an important addendum to 9.1. In this case, we can form both a forward time average using f, and a backward time average using f^{-1}. Given any x_0, let $\{x_n\}$ be the full orbit, consisting of the points $x_n = f^{\circ n}(x_0)$ where n ranges over all integers, both positive and negative. We continue to assume that f is measure preserving, that $0 < \mu(X) < \infty$, and that φ is integrable, but we do not assume ergodicity.

Corollary 9.7. The Invertible Case. *If the inverse function $f^{-1} : X \to X$ is also defined and measurable, then the forward time average*

$$A_\varphi^f(x_0) = \lim_{n\to\infty} \sum_{0\le j<n} \varphi(x_j))/n$$

is defined and equal to the backward time average

$$A_\varphi^{f^{-1}}(x) = \lim_{n\to\infty} \sum_{0\le j<n} \varphi(x_{-j})/n\ ,$$

except on a set of measure zero.

Proof. For example let N be the the invariant set consisting of points x where $A_\varphi^f(x) > A_\varphi^{f^{-1}}(x)$. Since

$$\int_N A_\varphi^f(x) \ = \ \int_N \varphi \ = \ \int_N A_\varphi^{f^{-1}}(x)$$

by the Birkhoff Theorem, it follows that N must be a set of measure zero. □.

JWM, 9-8-01