Lecture 3

So far we have learned two fundamental recurrence theorems in dynamic systems, namely the Birkhoff recurrence theorem in a TDS and the Poincaré recurrence theorem in a measure-preserving system (MPS for short). However they do not give quantitative information about the behavior of orbits, for example, the frequency of a orbit returnning to a given set. In this lecture, we are going to present a method to study the statistical behavior of orbits and prove some ergodic theorems.

3 Statistical behavior of orbits and ergodic theorems

3.1 Statistical behavior of orbits

Let (X,T) be a TDS. For $B \in \mathscr{B}(X)$ and $x \in X$, set $F_B(T,x,n) = \sharp\{0 \leq j \leq n-1 : T^j x \in B\}$, let $F_B(T,x) = \lim\limits_{n\to\infty} \frac{F_B(T,x,n)}{n}$ provided the limit exists. The term $F_B(T,x)$ is called the asymptotic density of the distribution of the iterates over B and $X \setminus B$.

Let χ_B denote the characteristic function of B, that is $\chi_B(x) = 1$ if $x \in B$, $\chi_B(x) = 0$ if $x \notin B$. Then

$$F_B(T,x,n) = \sum_{k=0}^{n-1} \chi_B(T^k x) \text{ and } F_B(T,x) = \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \chi_B(T^k x).$$

The above expression of $F_B(T,x)$ is naturally interpreted as the average of χ_B at the orbit of x. Rather than dealing with characteristic functions, it is more reasonable to start from studying the average of continuous functions.

Let $C(X)$ be the space of real-valued continuous functions endowed with the uniform topology (that is the topology induced by sup-norm). For $\varphi \in C(X)$ and $x \in X$, let $I_x(\varphi) = \lim\limits_{n\to\infty} \frac{1}{n} \sum\limits_{k=0}^{n-1} \varphi(T^k x)$ if the limit exists. $I_x(\varphi)$ is called the Birkhoff average (or the time average) of φ at x.

Fix $x \in X$. If we assume that $I_x(\varphi)$ exists for any $\varphi \in C(X)$, it's easy to see $I_x : C(X) \longrightarrow \mathbb{R}$ enjoys the following properties:

(i) (Linearity) $I_x(\alpha\varphi+\beta\psi) = \alpha I_x(\varphi)+\beta I_x(\psi)$ for any $\alpha, \beta \in \mathbb{R}, \varphi, \psi \in C(X)$.

(ii) (Boundness) $|I_x(\varphi)| \leq \sup_{y\in X} |\varphi(y)|$, for any $\varphi \in C(X)$.

(iii) (Positivity) $I_x(\varphi) \geq 0$ if $\varphi \geq 0$, $I_x(1) = 1$.

(iv) (Invariant under T) $I_x(\varphi \circ T) = I_x(\varphi)$ for any $\varphi \in C(X)$.

(i)-(iii) indicate that I_x is a positive bounded linear functional on $C(X)$, hence by the Riesz representation theorem, there exists a unique Borel probability measure μ_x on X such that

$$I_x(\varphi) = \int \varphi d\mu_x, \text{ for any } \varphi \in C(X).$$

Applying (iv), $\int \varphi d\mu_x = \int \varphi \circ T d\mu_x$ for any $\varphi \in C(X)$, which implies that μ_x is T-invariant, that is $\mu_x(T^{-1}B) = \mu_x(B)$ for all $B \in \mathscr{B}(X)$.

Notice that the above argument is based on the assumption that $I_x(\varphi)$ exists for all $\varphi \in C(X)$, so it's natural to consider the following questions.

Questions: (1) Are there points $x \in X$ such that $I_x(\varphi)$ exists for all $\varphi \in C(X)$?

(2) If μ is a T-invariant measure, does there exist $x \in X$ such that $I_x(\varphi) = \int \varphi d\mu$ for all $\varphi \in C(X)$?

The answers to the above questions are both positive. It bases on two fundamental theorems, one in TDS, one in ergodic theory.

3.2 Existence of invariant measures

Theorem 3.1 (Krylov–Bogolyubov)**.** *For any TDS (X, T), there exists at least one T-invariant Borel probability measure.*

Proof. Fix a $y \in X$. Let $\{\varphi_i\}_{i=1}^{\infty}$ be a countable subset dense in $C(X)$. Notice that for each $m \in \mathbb{N}_+$, the sequence $\frac{1}{n}\sum_{k=0}^{n-1} \varphi_m(T^k y)$ is bounded by $\|\varphi_m\|_\infty$, hence it has a convergent subsequence. Then by the diagonal process, one can find a subsequence $\{n_k\}_{k=1}^{\infty} \subset \mathbb{N}_+$ such that

$$\lim_{k\to\infty} \frac{1}{n_k} \sum_{j=0}^{n_k-1} \varphi_m(T^j y) =: J(\varphi_m) \tag{3.1}$$

exists for all φ_m. We claim

$$\lim_{k\to\infty} \frac{1}{n_k} \sum_{j=0}^{n_k-1} \varphi(T^j y) =: J(\varphi) \text{ exists for all } \varphi \in C(X). \tag{3.2}$$

To see (3.2), let $\varphi \in C(X)$ and $\epsilon > 0$, choose φ_m such that

$$\sup_{x\in X} |\varphi_m(x) - \varphi(x)| < \epsilon,$$

then

$$\frac{1}{n_k} \sum_{j=0}^{n_k-1} \varphi(T^j y) = \frac{1}{n_k} \sum_{j=0}^{n_k-1} \varphi_m(T^j y) + \frac{1}{n_k} \sum_{j=0}^{n_k-1} (\varphi(T^j y) - \varphi_m(T^j y)).$$

Notice that on the right hand side of the above equality, the first term converges to $J(\varphi_m)$, the second term is bounded by ϵ in absolute value, hence all limit points of the left hand side differ only by ϵ in absolute value, letting $\epsilon \to 0$, we see (3.2) holds.

Now consider $J : C(X) \longrightarrow \mathbb{R}$. Just as I_x in subsection 3.1, J satisfies conditions: (1) linearity, (2) boundness, (3) positivity, (4) $J(\varphi) = J(\varphi \circ T)$. By the Riesz representation theorem, there exists a Borel probability measure μ on X, such that for all $\varphi \in C(X)$, $J(\varphi) = \int \varphi d\mu$, moreover $\int \varphi d\mu = \int \varphi \circ T d\mu$, which implies $\mu = \mu \circ T^{-1}$. □

3.3 Birkhoff ergodic theorem

Theorem 3.2 (Birkhoff Ergodic Theorem (1931)). *Let $(X, \mathscr{B}, \mu, T)$ be a MPS. Let $f \in L^1(\mu)$ (i.e. $f : X \longrightarrow \mathbb{C}$ measurable and $\int |f| d\mu < \infty$). Then*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(T^k x) =: f^*(x) \text{ exists } \mu\text{-a.e. } x \in X,$$

furthermore

$$f^* \in L^1(\mu) \text{ and } \int_A f^* d\mu = \int_A f d\mu \text{ for any } A \in \mathscr{B} \text{ with } T^{-1}A = A.$$

The proof of this theorem is based on the following.

Theorem 3.3 (Maximal Ergodic Theorem). *Let $(X, \mathscr{B}, \mu, T)$ be a MPS. Let $f \in L^1_{\mathbb{R}}(\mu)$. Set $f_0 = 0$ and $f_n = \sum\limits_{k=0}^{n-1} f(T^k x)$ for $n \geq 1$. For $N \in \mathbb{N}$, define $F_N(x) = \max\limits_{0 \leq n \leq N} f_n(x)$. Then*

$$\int_{\{x: F_N(x) > 0\}} f d\mu \geq 0.$$

Proof. Fix $N \geq 1$. Notice that

$$\begin{aligned} F_N(Tx) + f(x) &= \max_{0 \leq n \leq N} f_n(Tx) + f(x) \\ &= \max\{0, f(Tx), f(Tx) + f(T^2x), \cdots, f(Tx) + \cdots + f(T^N x)\} + f(x) \\ &= \max_{1 \leq n \leq N+1} f_n(x) \geq \max_{1 \leq n \leq N} f_n(x). \end{aligned}$$

Let $A = \{x : F_N(x) > 0\}$. Since $F_N(x) = \max\limits_{0 \leq n \leq N} f_n(x) = \max\{0, \max\limits_{1 \leq n \leq N} f_n(x)\}$, we have $F_N(x) = \max\limits_{1 \leq n \leq N} f_n(x)$ on A, hence $f(x) \geq F_N(x) - F_N(Tx)$ on A. Since F_N is nonnegative and $\int_X g(Tx) d\mu = \int_X g(x) d\mu$ for any $g \in L^1(\mu)$, we have

$$\begin{aligned} \int_A f d\mu &\geq \int_A F_N(x) d\mu - \int_A F_N(Tx) d\mu \\ &= \int_X F_N(x) d\mu - \int_A F_N(Tx) d\mu \\ &= \int_X F_N(Tx) d\mu - \int_A F_N(Tx) d\mu \geq 0. \end{aligned}$$

$\square$

Corollary 3.3.1. *Let* $(X, \mathscr{B}, \mu, T)$ *be a MPS. Let* $g \in L^1_{\mathbb{R}}(\mu)$ *and* $\alpha \in \mathbb{R}$*. Set* $B_\alpha = \{x : \sup\limits_{n \geq 1} \frac{1}{n} \sum\limits_{k=0}^{n-1} g(T^k x) > \alpha\}$*. Then for any* $A \in \mathscr{B}$ *with* $T^{-1}A = A$*, we have*

$$\int_{A \cap B_\alpha} g d\mu \geq \alpha \mu(A \cap B_\alpha).$$

Proof. First consider the case that $A = X$. Define $f = g - \alpha$ and F_N as in the above theorem. Then $B_\alpha = \bigcup\limits_{N \geq 1} \{x : F_N(x) > 0\}$ and $\{x : F_N(x) > 0\} \uparrow B_\alpha$. For each N, by the above theorem, $\int_{\{x : F_N(x) > 0\}} f d\mu \geq 0$, apply the dominated convergence theorem, $\int_{B_\alpha} f d\mu \geq 0$, therefore $\int_{B_\alpha} g d\mu \geq \alpha \mu(B_\alpha)$.

In general case that $A \neq X$, since $A = T^{-1}A$, we can consider the subsystem $(A, \mathscr{B}(A), \mu|_A, T|_A)$, where $\mathscr{B}(A)$ is the sub-σ-algebra of $\mathscr{B}$ when restricted to A, more precisely, $\mathscr{B}(A) = A \cap \mathscr{B} := \{A \cap B : B \in \mathscr{B}\}$, $\mu|_A$ is defined by $\mu|_A(B) = \frac{\mu(B)}{\mu(A)}$ for $B \in \mathscr{B}(A)$ (the case $\mu(A) = 0$ is trivial). To apply the previous result to the new system, replace B_α by $B_\alpha \cap A$ and μ by $\mu|_A$, then $\int_{A \cap B_\alpha} g d\mu|_A \geq \alpha \mu|_A(A \cap B_\alpha)$, that is $\frac{1}{\mu(A)} \int_{A \cap B_\alpha} g d\mu \geq \frac{\alpha \mu(A \cap B_\alpha)}{\mu(A)}$, therefore $\int_{A \cap B_\alpha} g d\mu \geq \alpha \mu(A \cap B_\alpha)$.

$\square$