26 Lecture 26. Symbolic dynamics

26.1 Shift

Let us consider a finite set S whose elements we call symbols (thus S may be understood as a set of alphabets). A finite sequence of symbols is called a word. Make a set Σ consisting of some infinite sequences (both-sided or one-sided) $\omega = (\omega_n)_n$ ($\omega_n \in S$). Then we can make a shifted sequence $\omega' = \sigma\omega$ from ω as $\omega' = (\omega_{n+1})_n$:

$$\omega = \cdots \omega_{-k}\omega_{-k+1} \cdots \omega_{-1} \quad \omega_0 \quad \omega_1\omega_2 \cdots \omega_k\omega_{k+1} \cdots, \tag{26.1}$$

$$\sigma\omega = \omega' = \cdots \omega_{-k+1}\omega_{-k+2} \cdots \omega_0 \quad \omega_1 \quad \omega_2\omega_3 \cdots \omega_{k+1}\omega_{k+2} \cdots. \tag{26.2}$$

or in the one-sided case

$$\omega = \quad \omega_0 \quad \omega_1\omega_2 \cdots \omega_k\omega_{k+1} \cdots, \tag{26.3}$$

$$\sigma\omega = \omega' = \quad \omega_1 \quad \omega_2\omega_3 \cdots \omega_{k+1}\omega_{k+2} \cdots. \tag{26.4}$$

The operator σ is called a shift operator or shift.

26.2 Shift dynamical systems

$\Sigma \equiv S^{\mathbb{Z}}$ (resp. $\Sigma^+ = S^{\mathbb{N}}$) is the totality of both-side (resp. one-side) infinite symbol sequences on S. We can define shift on Σ and is called, as a dynamical system, the full shift on S.

A subset M of Σ is an invariant set if $\sigma M = M$. Then, we may define a restriction of σ to M (sometimes, it is written as σ_M), which is called a subshift.

For example, we can consider all the sequences $M \subset \{0, 1\}^{\mathbb{Z}}$ that never contains 11. Needless to say, the shift never changes the sequence structures, so we can define a shift dynamical system (σ, M). This is an example of a Markov subshift.

A finite sequence of symbols is called a word. Thus M above may be characterized as a sequence without word 11.

26.3 Topology or metric in symbol sequence space

We wish to compare different sequences. As can be seen from the illustration **22.4** if the length n words close to 0 are identical we should regard the sequences close. Thus, we introduce the following metric in Σ:

$$d(\omega, \omega') = \sum_n 2^{-|n|}\delta_{\omega_n\omega'_n}. \tag{26.5}$$

The full shift $(\sigma, \Sigma^{\mathbb{Z}})$ (or $(\sigma, \Sigma^{\mathbb{N}})$) has dense periodic orbits and topologically mixing **26.4**.

26.4 Topological mixing

Let $f : M \to M$ be a topological dynamical system (= C^0-endomorphism of a (smooth) manifold M). For any $U, V \subset M$ if $f^n(U) \cup V \neq \emptyset$ for any sufficiently large $n \in \mathbb{N}$ (i.e., there is a positive integer $N(U,V)$ and for all $n > N(U,V)$), we say this dynamical system is topologically mixing.

26.5 Cylinder set

A subset of sequence space M with a particular word at a particular position is called a cylinder set: for example

$$\{\omega \,|\, \omega_0 = \alpha_1, \omega_1 = \alpha_2, \cdots \omega_{10} = \alpha_{11}, \omega \in M\} \tag{26.6}$$

is a length (or rank) 11 cylinder set consisting of the totality of sequences in M such that

$$\cdots \omega_{-2}\omega_{-1}\alpha_1\alpha_2 \cdots \alpha_{10}\alpha_{11}\omega_{11}\omega_{12}\cdots, \tag{26.7}$$

where ω_k can be anything as long as $\omega \in M$.

26.6 Markov subshift

Let $S^\circ = n$ (the size of the alphabet). Let A be a $n \times n$ matrix whose elements are 0 or 1. Define

$$\Sigma_A = \{\omega \in \Sigma \,|\, A_{\omega_n, \omega_{n+1}} = 1, n \in \mathbb{Z}\}. \tag{26.8}$$

Σ_A^+ for one-sided systems may be analogously defined by replacing $\mathbb{Z}$ with $\mathbb{N}$. (σ, Σ_A) is called a Markov subshift whose structure matrix is A.

26.7 Fixed points of Markov subshift

If there is an orbit starting from symbol x that returns to itself after p shift, $(A^p)_{xx} > 0$. This number is actually the number of ways to go from x to itself in p steps.

The fixed points of σ^p has the following structure:[318]

$$(i_1 i_2 \cdots i_p i_1 i_2 \cdots i_p \cdots). \tag{26.9}$$

Therefore, the total number $N_{\text{Fix}}(\sigma^p)$ of fixed points of σ^p is $\text{Tr}\, A^p$:

$$N_{\text{Fix}}(\sigma^p) = \text{Tr}\, A^p. \tag{26.10}$$

Notice that this is closely related to the number of periodic orbits:

$$N_{\text{Fix}}(\sigma^n) = \sum_{(k,\tau):k\tau=n} \tau, \tag{26.11}$$

where τ is the period of a periodic orbit τ, and the summation is over all the pairs (k, τ).

26.8 Zeta function of dynamical system

The zeta function ζ_f for a dynamical system $f \in C^r(M)$ is generally defined as

$$\zeta_f(s) = \exp\left[\sum_{n=1}^{\infty} \frac{N_{\text{Fix}}(f^n)}{n} s^n\right]. \tag{26.12}$$

Using (26.11), we get

$$= \exp\left[\sum_{n=1}^{\infty} \frac{1}{n} \sum_{(k,\tau):k\tau=n} \tau s^n\right]. \tag{26.13}$$

Here, all the combinations (k, τ) appear once and only once; the totality of τ (if orbits are distinct, even if they have the same τ they must be counted separately) is system dependent, but if the system has one τ, $k\tau$ for all $k \in \mathbb{N}^+$ appear. Thus, for each τ the summation over k runs from 1 to ∞:

$$\begin{aligned} \zeta_f(s) &= \exp\left[\sum_{(k,\tau)} \frac{1}{k\tau} \tau s^{k\tau}\right] && (26.14) \\ &= \exp\left[\sum_{\tau} \sum_{k=1}^{\infty} \frac{1}{k} s^{k\tau}\right] = \exp\left[-\sum_{\tau} \log(1 - s^{\tau})\right] && (26.15) \\ &= \prod_{\tau} (1 - s^{\tau}). && (26.16) \end{aligned}$$

[318]Notice that (26.9) and its 'cyclic permutation', e.g., $(i_2 i_3 \cdots i_p i_1 i_2 i_3 \cdots i_p i_1 \cdots)$ are distinct points.

Here the summation (or product) over τ is over all the periods allowed to the system. This is why it is called the zeta function.

26.9 Zeta function for Markov subshift

Let us compute the zeta function for (σ, Σ_A) with the structure matrix A. We have an explicit formula for $N_{\mathrm{Fix}}(\sigma^n)$, so[319]

$$\zeta_\sigma(s) = \exp\left[\sum_{n=1}^{\infty} \frac{\mathrm{Tr}\, A^n}{n} s^n\right] = \exp[-\mathrm{Tr}\, \log(1 - sA)] = 1/\det(1 - sA). \qquad (26.17)$$

Thus, the inverse of the eigenvalue of A with the largest modulus gives the convergence radius ρ_A for the zeta function.

As we will see, $-\log \rho_A$ is the topological entropy of the dynamical system (the sup of the KS entropy). The invariant measure that realizes this sup value (thus, actually the max value) is the invariant measure of the Markov chain defined by the transition matrix A.

26.10 Topologically transitive Markov subshift

If there is $m \in \mathbb{N}$ such that $A^n > 0$ (all the elements positive), we say the resultant Markov subshift is topologically transitive: there is an orbit starting from any symbol to reach any symbols with a finite number of steps.

A topologically transitive Markov subshift is topologically mixing and periodic orbits are dense in Σ_A.[320]

The Perron-Frobenius theorem **26.11** tells us that $\rho_A = 1/\lambda_{\mathrm{PF}}$, where λ_{PF} is the Perron-Frobenius eigenvalue of A.

26.11 Perron-Frobenius theorem

Let A be a square matrix whose elements are all non-negative, and there is a positive integer n such that all the elements of A^n are positive. Then, there is a nondegenerate real positive eigenvalue λ such that

(i) $|\lambda_i| < \lambda$, where λ_i are eigenvalues of A other than λ,[321]

---

[319] Recall $\log\det(A) = \log(\prod \text{eigenvalue of } A) = \sum \log(\text{eigenvalue of } A) = \mathrm{Tr}\, \log A$.

[320] P1.9.9 KH p51

[321] That is, λ gives the spectral radius of A.

(ii) the elements of the eigenvector belonging to λ may be chosen all positive.
This special real eigenvalue giving the spectral radius is called the Perron-Frobenius eigenvalue.

For the 'shortest demonstration' see **26.13**.[322]

26.12 Preview of thermodynamic formalism

As you realize, thus dynamical systems are closely related to 1D spin systems. Periodic states correspond to ordered states in spin systems; thus, periodic dynamical systems correspond to spin systems with long-range interactions. Chaotic states corresponds to high-temperature states. Critical phenomena in spin systems correspond to Feigenbaum critical phenomena **8.7**.

Sinai introduced the thermodynamic formalism to understand dynamical systems, which systematize the ideas above. Invariant measures correspond to Gibbs measures. Observability of chaos may be characterized by a special temperature.

26.13 Proof of the Perron-Frobenius theorem[323]

Let us introduce the vectorial inequality notation for n-column vectors): $\boldsymbol{x} > 0$ (≥ 0) implies that all the components of $\boldsymbol{x}$ are positive (non-negative). Also let us write $\boldsymbol{x} \geq (>)$ $\boldsymbol{y}$, if $\boldsymbol{x} - \boldsymbol{y} \geq (>)\ 0$.

We use analogous symbols for $n \times n$ matrices A, $B, \cdots$ as well. $A > 0$: (≥ 0) implies that all the components of A are positive (non-negative). Also let us write $A \geq (>)\ B$, if $A - B \geq (>)\ 0$.

We consider $n \times n$ matrices $A \geq 0$ for which there is a positive integer m such that $A^m > 0$.

Let $\boldsymbol{x}$ be a vector such that $|\boldsymbol{x}| = 1$ and $\boldsymbol{x} \geq 0$. The largest ρ satisfying

$$A\boldsymbol{x} \geq \rho \boldsymbol{x} \tag{26.18}$$

is denoted by $\Lambda(\boldsymbol{x})$: $\Lambda(\boldsymbol{x}) = \max\{\rho \,|\, A\boldsymbol{x} \geq \rho\boldsymbol{x}\}$.

(o) Note that

$$\Lambda(\boldsymbol{x}) = \min_{i \text{ s.t. } x_i \neq 0} \frac{(A\boldsymbol{x})_i}{x_i}, \tag{26.19}$$

so it is continuous for $\boldsymbol{x} > 0$.

(o′) $A^m > 0$ implies $A^m \boldsymbol{x} > 0$ for $\boldsymbol{x} \geq 0$ ($\neq 0$).

[322] The newest version, taking account of H Tasaki's critical comments (April, 2018).

[323] A standard reference may be E. Seneta, *Non-negative matrices and Markov chains* (Springer, 1980). The proof here is an eclectic version due to many sources, including N. Iwahori, *Graphs and Stochastic Matrices* (Sangyo-tosho, 1974) and S. Sternberg: `http://www.math.harvard.edu/library/sternberg/slides/1180912pf.pdf`..

(i) Define a compact[324] set $U = \{\boldsymbol{x} \,|\, \boldsymbol{x} \geq 0, |\boldsymbol{x}| = 1\}$ and a continuous map Q on it as

$$Q\boldsymbol{x} \equiv \frac{A^m \boldsymbol{x}}{|A^m \boldsymbol{x}|} > 0. \tag{26.20}$$

Then, $\Lambda(Q\boldsymbol{x}) \geq \Lambda(\boldsymbol{x})$ on U.
[Demo] Let $A\boldsymbol{x} \geq \lambda \boldsymbol{x}$ for $\boldsymbol{x} \in U$. Then,

$$A^m A\boldsymbol{x} = AA^m \boldsymbol{x} \geq \lambda A^m \boldsymbol{x} \;\Rightarrow AQ\boldsymbol{x} \geq \lambda Q\boldsymbol{x}. \tag{26.21}$$

Therefore, $\Lambda(Q\boldsymbol{x}) \geq \Lambda(\boldsymbol{x})$ for any $\boldsymbol{x} \in U$.

(ii) There is a vector $\boldsymbol{z} \in U$ that maximizes $\Lambda(\boldsymbol{x})$.
[Demo] Let $C = Q(U)$. Notice that

$$C \subset \{\boldsymbol{x} \,|\, \boldsymbol{x} > 0, |\boldsymbol{x}| = 1\}, \tag{26.22}$$

because for any $\boldsymbol{x} \in U$ $\tilde{\boldsymbol{x}} = Q\boldsymbol{x}$ satisfies $|\tilde{\boldsymbol{x}}| = 1$ and $\tilde{\boldsymbol{x}} > 0$. Thus. $C \subset U$. Since Q is continuous on U, C is compact. $\boldsymbol{x} > 0$ for any $\boldsymbol{x} \in C$, so $\Lambda(\boldsymbol{x})$ is continuous and has a max value on C. For any $\boldsymbol{x} \in U$ (i) says $\Lambda(Q\boldsymbol{x}) \geq \Lambda(\boldsymbol{x})$, so Λ defined on U is maximum on C, since $C \subset U$.

(iii) Let us write $\lambda(A) = \Lambda(\boldsymbol{z})$. That is, $\lambda(B) = \max_{\boldsymbol{x} \in U}\{\Lambda(\boldsymbol{x}) \,|\, B\boldsymbol{x} \geq \Lambda(\boldsymbol{x})\boldsymbol{x}\}$ for any $n \times n$ matrix with some positive m such that $B^m > 0$. $\lambda(A)$ is an eigenvalue of A, and $\boldsymbol{z}$ belongs to its eigenspace: $A\boldsymbol{z} = \lambda(A)\boldsymbol{z}$.
[Demo] Even if not, we have $\boldsymbol{w} = A\boldsymbol{z} - \lambda(A)\boldsymbol{z} \geq 0$ (not equal to zero) by definition of $\lambda(A)$. Notice that for any vector $\boldsymbol{x} \geq 0$ $A^m \boldsymbol{x} > 0$, so unless $\boldsymbol{w} = 0$

$$A^m \boldsymbol{w} = AA^m \boldsymbol{z} - \lambda(A)A^m \boldsymbol{z} > 0. \tag{26.23}$$

This implies $\Lambda(A^m \boldsymbol{z}) > \lambda(A)$, but $\lambda(A)$ is the maximum of Λ, so this is a contradiction. Therefore, $\boldsymbol{w} = 0$. That is, $\boldsymbol{z}$ is an eigenvector belonging to λ.

(iv) $\boldsymbol{z} > 0$ because max of Λ is on C.

(v) λ is the spectral radius of A.
[Demo] Suppose $A\boldsymbol{y} = \lambda' \boldsymbol{y}$. Let $\boldsymbol{q}$ be the vector whose components are absolute values of $\boldsymbol{y}$: $q_i = |y_i|$. Then, $A\boldsymbol{q} \geq |\lambda'|\boldsymbol{q}$ (as seen from (26.19)). Therefore, $|\lambda'| \leq \lambda(A)$.

(vi) The absolute values of other eigenvalues are smaller than $\lambda(A)$. That is, no eigenvalues other than $\lambda(A)$ is on the spectral circle.
[Demo] Suppose λ' is an eigenvalue on the spectral circle but is not real positive. Let $\boldsymbol{q}$ be the vector whose components are absolute values of an eigenvector belonging to λ'. Since $A\boldsymbol{q} \geq |\lambda'|\boldsymbol{q} = \lambda(A)\boldsymbol{q}$, actually we must have $A\boldsymbol{q} = \lambda(A)\boldsymbol{q}$. Therefore, the absolute value of each component of the vector $A^m \boldsymbol{y} = \lambda'^m \boldsymbol{y}$ coincides with the corresponding component of $A^m \boldsymbol{q}$. This implies

$$\left| \sum_j (A^m)_{ij} y_j \right| = \sum_j (A^m)_{ij} |y_j| = \sum_j |(A^m)_{ij} y_j|. \tag{26.24}$$

[324]Since we deal with finite-dimensional vector spaces, 'closed' can always replace 'compact.'

All the components of A^m are real positive. Therefore, all the arguments of y_j are identical,[325] so $\boldsymbol{y}$ is parallel to $\boldsymbol{q}$. Hence, $\lambda' = \lambda(A)$.

(vii) $\lambda(A)$ is non-degenerate.
We show that $\lambda(A)$ is a non-degenerate root of the characteristic equation $\det(\lambda I - A)$. We use
(viia) For any matrix $B \geq 0$ ($\neq 0$) $\det(\lambda I - B) > 0$ for real $\lambda > \lambda(B)$ (= the spectral radius).
(viib) Let $A_{(i)}$ be the matrix with the ith row and ith column removed from A. Then,[326]

$$\frac{d}{d\lambda}\det(\lambda I - A) = \sum_i \det(\lambda I - A_{(i)}). \tag{26.25}$$

Let $A_{[i]}$ be the matrix with all the elements in ith row and ith column of A being replaced with 0. Suppose $A_{[i]}\boldsymbol{z} = \sigma\boldsymbol{z}$ ($\boldsymbol{z} \neq 0$). Then, for all i

$$A|\boldsymbol{z}| \geq A_{[i]}|\boldsymbol{z}| \geq |\sigma||\boldsymbol{z}|. \tag{26.26}$$

That is $\lambda(A) \geq |\sigma|$. If $\lambda(A) = |\sigma|$, then $A|\boldsymbol{z}| = |\sigma||\boldsymbol{z}|$, so $|\boldsymbol{z}| > 0$ and $(A - A_{[i]})|\boldsymbol{z}| = 0$, but this is impossible. Therefore, the spectral radius of $A_{[i]}$ is smaller than $\lambda(A)$ for any i. Since $A_{[i]}$ and $A_{(i)}$ have the same spectral circle, (viia) implies all the terms in (26.25) are positive for $\lambda = \lambda(A)$, so (viib) implies $\lambda(A)$ is non-degenerate.

---

[325] $a, b \neq 0$ and $|a + b| = |a| + |b|$ imply the real positivity of a/b. This means inductively that $|\sum a_i| = \sum |a_i|$ implies all a_i must have the same argument.

[326] Let $X = n \times n$ diagonal matrix with the diagonal $x_1, \cdots, x_n$. Then, expanding the determinant with respect to the ith row, obviously,

$$\frac{d}{dx_i}\det(X - A) = \det(X_{(i)} - A_{(i)}),$$

where $X_{(i)} = (n-1) \times (n-1)$ diagonal matrix with the diagonal $x_1, \cdots, (x_i), \cdots, x_n$ (x_i omitted). The rest is due to the chain rule.