Lecture 12

6.3 Measures with maximal entropy

Recall we have proved the variational principle. Let (X,T) be a TDS, then

$$h_{top}(T) = \sup\{h_\mu(T) : \mu \in M(X,T)\}.$$

Definition 6.1. *Say $\mu \in M(X,T)$ is a measure with maximal entropy if*

$$h_{top}(T) = h_\mu(T).$$

Proposition 6.2. *Let (X,T) be a TDS. Suppose that the entropy map*

$$\mu \mapsto h_\mu(T)$$

is upper-semi-continuous on $M(X,T)$. Then there exists at least one measure in $M(X,T)$ with maximal entropy.

Proof. By the variational principle, there exists a sequence $(\mu_n) \subset M(X,T)$ such that

$$h_{\mu_n}(T) \to h_{top}(T).$$

By compactness, there is a subsequence (μ_{n_k}) of (μ_n) such that $\mu_{n_k} \to \mu \in M(X,T)$. Hence

$$h_\mu(T) \geq \lim_{k\to\infty} h_{\mu_{n_k}}(T) = h_{top}(T).$$

□

Proposition 6.3. *Let (X,T) be a subshift over $\{1,\cdots,k\}$. Then the entropy map is upper-semi-continuous.*

Proof. Let $\mu \in M(X,T)$. Recall that

$$h_\mu(T) = \lim_{n\to\infty} \frac{1}{n} H_\mu(\bigvee_{i=0}^{n-1} T^{-i}\xi) = \inf_n \frac{1}{n} H_\mu(\bigvee_{i=0}^{n-1} T^{-i}\xi),$$

where $\xi = \{[i] : 1 \leq i \leq k\}$. Hence for any $\epsilon > 0$, there exists $n \in \mathbb{N}$ such that

$$\frac{1}{n} H_\mu(\bigvee_{i=0}^{n-1} T^{-i}\xi) \leq h_\mu(T) + \epsilon.$$

Write $\xi_n = \bigvee_{i=0}^{n-1} T^{-i}\xi$, recall it consists of closed and open sets. Suppose $\mu_m \in M(X,T)$ with $\mu_m \to \mu$, then

$$\lim_{m\to\infty} \frac{1}{n} H_{\mu_m}(\xi_n) = \frac{1}{n} H_\mu(\xi_n) \leq h_\mu(T) + \epsilon.$$

Since $h_{\mu_m}(T) \leq \frac{1}{n} H_{\mu_m}(\xi_n)$, we have $\overline{\lim}_{m\to\infty} h_{\mu_m}(T) \leq h_\mu(T) + \epsilon$. Since $\epsilon > 0$ is arbitrary, we complete the proof. □

Example 1. (Full shift over $\{1, \cdots, k\}$). Let $\Sigma = \{1, \cdots, k\}^{\mathbb{N}}$, let $\sigma : \Sigma \to \Sigma$ be the left shift. Recall $h_{top}(\sigma) = \log k$. Let μ be the $\{\frac{1}{k}, \cdots, \frac{1}{k}\}$-product measure on Σ. Recall $h_\mu(\sigma) = \log k$. Hence μ is a measure with maximal entropy. In fact, μ is the unique measure to attain the maximal entropy, we will show this in the following more general example.

Example 2. (Subshift of finite type over $\{1, \cdots, k\}$). Let A be a $k \times k$ 0-1 matrix. Assume A is irreducible (that is there exists $l \in \mathbb{N}$, such that $\sum_{i=1}^{l} A^i > 0$). Define

$$\Sigma_A = \{(x_i)_{i=1}^\infty \in \Sigma : A_{x_i x_{i+1}} = 1 \text{ for all } i \geq 1\}.$$

Let σ be the left shift over Σ_A. Recall that

$$h_{top}(\sigma) = \log \lambda,$$

where λ is the largest positive eigenvalue of A, which exists by Perron–Frobenius theorem.

Let $(u_1, \cdots, u_k)$ be a strictly positive left eigenvector of A corresponding to λ, let $(v_1, \cdots, v_k)^T$ be a strictly positive right eigenvector of A of λ. Suppose that $\sum_{i=1}^{k} u_i v_i = 1$. Define $\vec{p} = \{u_1 v_1, \cdots, u_k v_k\}$. Define a $k \times k$ matrix $P = (p_{ij})_{k \times k}$ by

$$p_{ij} = \frac{A_{ij} v_j}{\lambda v_i}.$$

Observe that

(i) P is a stochastic matrix.

(ii) $\vec{p}P = \vec{p}$.

To see (i), for each $i \in \{1, \cdots, k\}$,

$$\sum_{j=1}^{k} p_{ij} = \sum_{j=1}^{k} \frac{A_{ij} v_j}{\lambda v_i} = \frac{1}{\lambda v_i} \sum_{j=1}^{k} A_{ij} v_j = 1.$$

To see (ii), for each $j \in \{1, \cdots, k\}$,

$$(\vec{p}P)_j = \sum_i p_i P_{ij} = \sum_i u_i v_i \frac{A_{ij} v_j}{\lambda v_i} = \frac{v_j}{\lambda} \sum_i u_i A_{ij} = v_j u_j = p_j.$$

Let μ be the $(\vec{p}, P)$-Markov measure, that is

$$\mu([i_1 i_2 \cdots i_n]) = p_{i_1} p_{i_1 i_2} \cdots p_{i_{n-1} i_n},$$

for any admissible word $[i_1 i_2 \cdots i_n]$.

Recall that

$$h_\mu(\sigma) = \sum_{i,j} -p_i p_{ij} \log p_{ij}.$$

By the definitions of $\vec{p}$ and P,

$$\begin{aligned}
h_\mu(\sigma) &= \sum_{i,j} -p_i p_{ij} \log p_{ij} \\
&= \sum_{i,j:A_{ij}=1} -u_i v_i \frac{v_j}{\lambda v_i} \log \frac{v_j}{\lambda v_i} \\
&= -\frac{1}{\lambda} \sum_{i,j:A_{ij}=1} u_i v_j (\log v_j - \log v_i - \log \lambda) \\
&= -\frac{1}{\lambda}(\sum_{i,j} u_i A_{ij} v_j \log v_j - \sum_{i,j} u_i A_{ij} v_j \log v_i - \log \lambda \sum_{i,j} u_i A_{ij} v_j) \\
&= \log \lambda.
\end{aligned}$$

Hence μ is a measure of maximal entropy.

We can see that μ attains the maximal entropy in another way. Notice that for any admissible word $[i_1 i_2 \cdots i_n]$,

$$\begin{aligned}
\mu([i_1 i_2 \cdots i_n]) &= p_{i_1} p_{i_1 i_2} \cdots p_{i_{n-1} i_n} \\
&= u_{i_1} v_{i_1} \cdot \frac{v_{i_2}}{\lambda v_{i_1}} \cdots \frac{v_{i_n}}{\lambda v_{i_{n-1}}} \\
&= u_{i_1} v_{i_n} \lambda^{-(n-1)}.
\end{aligned}$$

Hence there is some constant $c > 0$, such that

$$c^{-1}\lambda^{-n} \le \mu([i_1 i_2 \cdots i_n]) \le c\lambda^{-n},$$

for any admissible word $[i_1 i_2 \cdots i_n]$. In general, we call this property the Gibbs property.

Recall we have Shannon-McMillan-Breiman theorem,

$$\lim_{n\to\infty} -\frac{\log \mu(\xi_n(x))}{n} = h_\mu(\sigma), \text{ for } \mu\text{-a.e. } x,$$

where $\xi_n(x)$ is the admissible cylinder where x lies in. By Gibbs property, the limit on the right hand side equals $\log \lambda$ for every x, hence we see again $h_\mu(\sigma) = \log \lambda$.

The measure μ constructed above is called the Parry measure, which was first discovered by William Parry in 1964, he also showed that μ is the unique measure that attains the maximal entropy.

Lemma 6.5. *Let $p_1, \cdots, p_m > 0$ with $\sum_{i=1}^m p_i = s$. Let $a_1, \cdots, a_m \in \mathbb{R}$. Then*

$$\sum_{i=1}^m (p_i a_i - p_i \log p_i) \le s(\log(\sum_{i=1}^m e^{a_i}) + \log\frac{1}{s}).$$

Proof. Let $p_i = sq_i$, then $(q_1, \cdots, q_m)$ is a probability vector. Hence

$$\begin{aligned}
\sum_{i=1}^{m}(p_i a_i - p_i \log p_i) &= \sum_{i=1}^{m}(sq_i a_i - sq_i \log sq_i) \\
&= \sum_{i=1}^{m}(sq_i a_i - sq_i \log s - sq_i \log q_i) \\
&= s\sum_{i=1}^{m}(q_i a_i - q_i \log q_i) - s\log s \\
&= s\sum_{i=1}^{m} q_i \log \frac{e^{a_i}}{q_i} - s\log s \\
&\le s\log(\sum_{i=1}^{m} q_i \cdot \frac{e^{a_i}}{q_i}) - s\log s \\
&= s(\log(\sum_{i=1}^{m} e^{a_i}) - \log s).
\end{aligned}$$

□

Lemma 6.6. *Let μ, η be two probability measures on Σ_A. Suppose $\mu \perp \eta$. Then*

$$\lim_{n\to\infty} \sum_{I\in\xi_n} \eta(I)\log\mu(I) - \eta(I)\log\eta(I) = -\infty.$$

Proof. Since $\mu \perp \eta$, there exists $E \subset \Sigma_A$ Borel with $\mu(E) = 0$ and $\eta(E) = 1$. Given $\epsilon > 0$, there are compact sets $K_1 \subseteq E$, $K_2 \subseteq X \backslash E$, such that $\eta(E\backslash K_1) < \epsilon$ and $\mu((X \setminus E) \setminus K_2) < \epsilon$.

Let n be so large that $\mathrm{diam}(\xi_n) < \frac{1}{2}\mathrm{dist}(K_1, K_2)$, then for any $I \in \xi_n$, I intersects at most one of K_1 and K_2. Hence

$$\begin{aligned}
&\sum_{I\in\xi_n} (\eta(I)\log\mu(I) - \eta(I)\log\eta(I)) \\
&= \sum_{\substack{I\in\xi_n \\ I\cap K_1\neq\emptyset}} (\eta(I)\log\mu(I) - \eta(I)\log\eta(I)) \\
&+ \sum_{\substack{I\in\xi_n \\ I\cap K_1=\emptyset}} (\eta(I)\log\mu(I) - \eta(I)\log\eta(I)).
\end{aligned}$$

Notice that

$$1 - \epsilon = \eta(E) - \epsilon < \eta(K_1) \le \sum_{\substack{I\in\xi_n \\ I\cap K_1\neq\emptyset}} \eta(I) \le 1,$$

and

$$0 = \mu(K_1) \le \sum_{\substack{I\in\xi_n \\ I\cap K_1\neq\emptyset}} \mu(I) \le \mu((X \setminus E) \setminus K_2) < \epsilon.$$

Applying the above lemma,

$$\begin{aligned}
&\sum_{\substack{I\in\xi_n\\ I\cap K_1\neq\emptyset}} (\eta(I)\log\mu(I) - \eta(I)\log\eta(I))\\
&\leq \Big(\sum_{\substack{I\in\xi_n\\ I\cap K_1\neq\emptyset}} \eta(I)\Big)\Big[\log\Big(\sum_{\substack{I\in\xi_n\\ I\cap K_1\neq\emptyset}} \mu(I)\Big) - \log\Big(\sum_{\substack{I\in\xi_n\\ I\cap K_1\neq\emptyset}} \eta(I)\Big)\Big]\\
&\leq \log\frac{\epsilon}{1-\epsilon}.
\end{aligned}$$

Also we have estimate

$$\begin{aligned}
&\sum_{\substack{I\in\xi_n\\ I\cap K_1=\emptyset}} (\eta(I)\log\mu(I) - \eta(I)\log\eta(I))\\
&\leq \Big(\sum_{\substack{I\in\xi_n\\ I\cap K_1=\emptyset}} \eta(I)\Big)\Big[\log\Big(\sum_{\substack{I\in\xi_n\\ I\cap K_1=\emptyset}} \mu(I)\Big) - \log\Big(\sum_{\substack{I\in\xi_n\\ I\cap K_1=\emptyset}} \eta(I)\Big)\Big]\\
&\leq \max_{0\leq s\leq 1}(-s\log s).
\end{aligned}$$

Combining these estimates together, we have

$$\sum_{I\in\xi_n}(\eta(I)\log\mu(I) - \eta(I)\log\eta(I)) \leq \log\frac{\epsilon}{1-\epsilon} + \max_{0\leq s\leq 1}(-s\log s).$$

Since the right hand side tends to $-\infty$ as $\epsilon\to 0$, we complete the proof. □

We will need the following property of ergodic measures.

Proposition 6.4. *Let $\mu,\eta\in M(X,T)$ be two ergodic measures. If $\mu\neq\eta$, then $\mu\perp\eta$.*

Proof. By Lebesgue decomposition theorem, there exist two unique probability measures μ_1 and μ_2 and a unique number $r\in[0,1]$, such that

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

where $\mu_1 \ll \eta$ and $\mu_2\perp\eta$.

We first show that $\mu_1,\mu_2\in M(X,T)$. Notice that

$$\mu = \mu\circ T^{-1} = r\mu_1\circ T^{-1} + (1-r)\mu_2\circ T^{-1},$$

and

$$\mu_1\circ T^{-1} \ll \eta\circ T^{-1} = \eta,\ \ \mu_2\circ T^{-1}\perp\eta\circ T^{-1} = \eta.$$

By uniqueness of the decomposition, we have $\mu_1\circ T^{-1} = \mu_1$ and $\mu_2\circ T^{-1} = \mu_2$, namely $\mu_1,\mu_2\in M(X,T)$.

Next we show we must have $r = 0$, which shows $\mu \perp \eta$. Since μ is ergodic, μ is an extreme point of $M(X, T)$, we have $r = 0$ or $r = 1$. If $r = 1$, we have $\mu \ll \eta$. In this situation, we consider decomposition

$$\eta = p\eta_1 + (1-p)\eta_2, \text{ with } \eta_1 \ll \mu, \eta_2 \perp \mu, p \in [0,1].$$

Again we have $p = 0$ or $p = 1$. If $p = 0$, we have $\eta \perp \mu$ and $\mu \ll \eta$, which forces $\mu = 0$, a contradiction. If $p = 1$, we have $\eta \ll \mu$ and $\mu \ll \eta$, which leads to $\mu = \eta$, also a contradiction. Hence we have $r = 0$ and $\mu \perp \eta$. □

Now we can give the proof that the Parry measure is the unique measure that attains the maximal entropy.

Proof of μ is the unique measure with maximal entropy. Let μ be the Parry measure on Σ_A. Notice that μ is ergodic since A is irreducible. Recall we have

$$c^{-1}\lambda^{-n} \leq \mu(I) \leq c\lambda^{-n},$$

for any admissible word $I \in \xi_n$.

Now assume that η is another ergodic measure with entropy $\log\lambda$. By proposition above, we have $\mu \perp \eta$. Since

$$\log\lambda = \inf_n \frac{1}{n} H_\eta(\xi_n) = \inf_n \frac{1}{n} \sum_{I \in \xi_n} -\eta(I)\log\eta(I),$$

we have for any n,

$$\sum_{I \in \xi_n} -\eta(I)\log\eta(I) \geq n\log\lambda.$$

By the Gibbs property of μ,

$$\sum_{I \in \xi_n} \eta(I)\log\mu(I) \geq \log(c^{-1}\lambda^{-n}) = -n\log\lambda - \log c.$$

Taking the summation,

$$\sum_{I \in \xi_n} (\eta(I)\log\mu(I) - \eta(I)\log\eta(I)) \geq -\log c,$$

contradicting Lemma 6.6. The proof is completed.

□