Lecture 4. Entropy and Markov Chains

The most important numerical invariant related to the orbit growth in topological dynamical systems is *topological entropy.*[1] It represents the exponential growth rate of the number of orbit segments which are distinguishable with an arbitrarily high but finite precision. Of course, topological entropy is invariant by topological conjugacy. For measurable dynamical systems, an entropy can be defined using the invariant measure. It gives an indication of the amount of randomness or complexity of the system. The relation between measure–theoretical entropy and topological entropy is given by a variational principle.

4.1 Topological Entropy

We will follow the definition given by Rufus Bowen in [Bo, Chapter 4]. Let X be a compact metric space.

Definition 4.1 *Let* $S \subset X$, $n \in N$ *and* $\varepsilon > 0$. S *is a* (n, ε)–spanning set *if for every* $x \in X$ *there exists* $y \in S$ *such that* $d(f^j(x), f^j(y)) \leq \varepsilon$ *for all* $0 \leq j \leq n$.

It is immediate to check that the compactness of X implies the existence of finite spanning sets. Let $r(n, \varepsilon)$ be the least number of points in an (n, ε)–spanning set. If we bound the time of observation of our dynamical system by n and our precision in making measurements is ε we will see at most $r(n, \varepsilon)$ orbits.

Exercise 4.2 Show that if X admits a cover by m sets of diameter $\leq \varepsilon$ then $r(n, \varepsilon) \leq m^{n+1}$.

Definition 4.3 *The* topological entropy $h_{top}(f)$ *of* f *is given by*

$$h_{top}(f) = \lim_{\varepsilon \to 0} \limsup_{n \to +\infty} \frac{1}{n} \log r(n, \varepsilon) \ . \qquad [4.1]$$

In the previous definition one cannot replace lim sup with lim since there exist examples of maps for which the limit does not exist. However one can replace it with lim inf still obtaining the topological entropy (see [Mn1], Proposition 7.1, p. 237).

Exercise 4.4 Show that the topological entropy for any diffeomorphism of a compact manifold is always finite.

Exercise 4.5 Let $X = \{x \in \ell^2(\mathbb{N}),\, |x_i| < 2^{-i} \text{ for all } i \in \mathbb{N}\}$, $f((x_i)_{i\in\mathbb{N}}) = (2x_{i+1})_{i\in\mathbb{N}}$. Let $k \in \mathbb{N}$. Show that for this system $r(n, k^{-1}) > k^n$ thus $h_{top}(f) = \infty$.

[1] According to Roy Adler [BKS, p. 103] "topological entropy was first defined by C. Shannon [Sh] and called by him *noiseless channel capacity.*

Exercise 4.6 Show that the topological entropy of the p–adic map of Exercise 2.36 is $\log p$.

Remark 4.7 The topological entropy of a flow φ_t is defined as the topological entropy of the time–one diffeomorphism $f = \varphi_1$.

Exercise 4.8 Show that :
(i) the topological entropy of an isometry is zero ; if h is an isometry the topological entropy of f equals that of $h^{-1} \circ f \circ h$.
(ii) if f is a homeomorphism of a compact space X then $h_{top}(f) = h_{top}(f^{-1})$;
(iii) $h_{top}(f^m) = |m| h_{top}(f)$.

Exercise 4.9 Let X be a metric space and f a continuous endomorphism of X. We say that a set A is (n, ε)–separated if for all $x, y \in X$ there exists a $0 \le j \le n$ such that $d(f^j(x), f^j(y)) > \varepsilon$. We denote $s(n, \varepsilon)$ the maximal cardinality of an (n, ε)–separated set. Show that :
(i) $s(n, 2\varepsilon) \le r(n, \varepsilon) \le s(n, \varepsilon)$;
(ii) $h_{top}(f) = \lim_{\varepsilon \to 0} \limsup_{n \to +\infty} \frac{1}{n} \log s(n\varepsilon)$;
(iii) if X is a compact subset of $\mathbb{R}^l$ and f is Lipschitz with Lipschitz constant K then $h_{top}(f) \le l \log K$.

Proposition 4.10 *The topological entropy does not depend on the choice of the metric on X provided that the induced topology is the same. The topological entropy is invariant by topological conjugacy.*

Proof. We first show how the second statement is a consequence of the first. Let f, g be topologically conjugate via a homeomorphism h. Let d denote a fixed metric on X and d' denote the pullback of d via h : $d'(x_1, x_2) = d(h^{-1}(x_1), h^{-1}(x_2))$. Then h becomes an isometry so $h_{top}(f) = h_{top}(g)$ (see Exercise 4.8).

Let us now show the first part. Let d and d' be two different metrics on X which induce the same topology and let $r_d(n, \varepsilon)$ and $r_{d'}(n, \varepsilon)$ denote the minimal cardinality of a (n, ε)-spanning set in the two metrics. We will denote $h_{top,d}(f)$ and $h_{top,d'}(f)$ the corresponding topological entropies.

Let $\varepsilon > 0$ and consider the set D_ε of all pairs $(x_1, x_2) \in X \times X$ such that $d(x_1, x_2) \ge \varepsilon$. This is a compact subset of $X \times X$ thus d' takes a minimum $\delta'(\varepsilon) > 0$ on D_ε. Thus any $\delta'(\varepsilon)$–ball in the metric d' is contained in a ε–ball in the metric d. From this one gets $r_{d'}(n, \delta'(\varepsilon)) \ge r_d(n, \varepsilon)$ thus $h_{top,d'}(f) \ge h_{top,d}(f)$. Interchanging the role of the two metrics one obtains the opposite inequality. □

Exercise 4.11 Show that if g is a factor of f then $h_{top}(g) \le h_{top}(f)$.

An alternative but equivalent definition of topological entropy is obtained considering all possible open covers of X and their refinements obtained by iterating f.

Definition 4.12 *If α, β are open covers of X theire* join $\alpha \vee \beta$ *is the open cover by all sets of the form $A \cap B$, where $A \in \alpha$ and $B \in \beta$. An open cover β is a* refinement *of α, written $\alpha < \beta$, if every member of β is a subset of a member of α.*

Let α be an open cover of X and let $N(\alpha)$ be the number of sets in a finite subcover of α with smallest cardinality. We denote $f^{-1}\alpha$ the open cover consisting of all sets $f^{-1}(A)$ where $A \in \alpha$.

Exercise 4.13 If $\{a_n\}_{n\in\mathbb{N}}$ is a sequence of real numbers such that $a_{n+m} \le a_n + a_m$ for all n, m then $\lim_{n\to+\infty} a_n/n$ exists and equals $\inf_{n\in\mathbb{N}} a_n/n$. [Hint : $n = kp+m$, $\frac{a_n}{n} \le \frac{a_p}{p} + \frac{a_m}{kp}$.]

Theorem 4.14 *The topological entropy of f is given by*

$$h_{top}(f) = \sup_{\alpha} \lim_{n\to\infty} \frac{1}{n} \log N\left(\bigvee_{i=0}^{n-1} f^{-i}\alpha\right) . \qquad [4.2]$$

For its proof see [Wa, pp. 173-174].

4.2 Entropy and information. Metric entropy.

In order to define metric entropy and to make clear its analogy with the formula [4.2] of topological entropy we will preliminarly introduce some general considerations on the relationship between entropy and information (see [Khi]).

Suppose that one performs an experiment which we will denote α which has $m \in \mathbb{N}$ possible mutually esclusive outcomes $A_1, \ldots, A_m$ (e.g. throwing a coin $m = 2$ or a dice $m = 6$). Assume that each possible outcome A_i happens with a probability $p_i \in [0,1]$, $\sum_{i=1}^m p_i = 1$ (in an experimental situation the probability will be defined statistically). In a probability space $(X, \mathcal{A}, \mu)$ this corresponds to the following setting : α is a finite *partition* $X = A_1 \cup \ldots \cup A_m \mod(0)$, $A_i \in \mathcal{A}$, $\mu(A_i \cap A_j) = 0$, $\mu(A_i) = p_i$. We want to define a function (called *entropy*) which measures the *uncertainity* associated to a prediction of the result of the experiment (or, equivalently, which measures the amount of *information* which one can gain from performing the experiment).

Let $\Delta^{(m)}$ denote the standard m-simplex of $\mathbb{R}^m$,

$$\Delta^{(m)} = \{(x_1, \ldots, x_m) \in \mathbb{R}^m \mid x_i \in [0,1],\ \sum_{i=1}^m x_i = 1\}.$$

Definition 4.15 *A continuous function $H^{(m)} : \Delta^{(m)} \to [0, +\infty]$ is called an* entropy *if it has the following properties :*

(1) symmetry : $\forall\, i,j \in \{1,\dots,m\}$ $H^{(m)}(p_1,\dots,p_i,\dots,p_j,\dots,p_m) = H(p_1,\dots,p_j,\dots,p_i,\dots,p_m)$ *;*
(2) $H^{(m)}(1,0,\dots,0) = 0$ *;*
(3) $H^{(m)}(0,p_2,\dots,p_m) = H^{(m-1)}(p_2,\dots,p_m)$ $\forall\ m \geq 2,\ \forall\ (p_2,\dots,p_m) \in \Delta^{(m-1)}$ *;*
(4) $\forall\ (p_1,\dots,p_m) \in \Delta^{(m)}$ *one has* $H^{(m)}(p_1,\dots,p_m) \leq H^{(m)}\left(\frac{1}{m},\dots,\frac{1}{m}\right)$ *where equality is possible if and only if* $p_i = \frac{1}{m}$ *for all* $i = 1,\dots,m$ *;*
(5) Let $(\pi_{11},\dots,\pi_{1l},\pi_{21},\dots,\pi_{2l},\dots,\pi_{m1},\dots,\pi_{ml}) \in \Delta^{(ml)}$ *; for all* $(p_1,\dots,p_m) \in \Delta^{(m)}$ *one must have*

$$H^{(ml)}(\pi_{1l},\dots,\pi_{1l},\pi_{21},\dots,\pi_{ml}) = H^{(m)}(p_1,\dots,p_m) + \sum_{i=1}^{m} p_i H^{(l)}\left(\frac{\pi_{i1}}{p_i},\dots,\frac{\pi_{il}}{p_i}\right).$$

In the above definition :
(2) says that if some outcome is certain then the entropy is zero ;
(3) says that no information is gained from impossible outcomes (i.e. outcomes with probability zero) ;
(4) says that the maximal uncertainity of the outcome is obtained when the possible results have the same probabilitly ;
(5) describes the behaviour of entropy when independent distinct experiences are performed. Let β denote another experiment with possible outcomes $B_1,\dots,B_l$ (i.e. another partition of $(X,\mathcal{A},\mu)$). Let π_{ij} be the probablility of A_i *and* B_j. The conditional probability of B_j is prob$(B_j \mid A_i) = \frac{\pi_{ij}}{p_i}$ (i.e. $\mu(A_i \cap B_j)$). Clearly the uncertainity of the outcome of the experiment β once one has already performed α with outcome A_i is given by $H^{(l)}\left(\frac{\pi_{i1}}{p_i},\dots,\frac{\pi_{il}}{p_i}\right)$.

Theorem 4.16 *An entropy is necessarily a positive multiple of*

$$H(p_1,\dots,p_m) = -\sum_{i=1}^{m} p_i \log p_i\ . \qquad [4.3]$$

Here we adopt the convention $0\log 0 = 0$. The above theorem and its proof are taken from [Khi, pp. 10-13].

Proof. Let $K(m) = H\left(\frac{1}{m},\dots,\frac{1}{m}\right)$. By (3) and (4) K is increasing : $K(m) = H(0,1/m,\dots,1/m) \leq H(1/(m+1),\dots,1/(m+1)) = K(m+1)$. Let m and l be two positive integers. Applying (5) with $\pi_{ij} \equiv \frac{1}{ml}$, $p_i \equiv \frac{1}{m}$ gives

$$K(lm) = K(m) + \sum_{i=1}^{m} \frac{1}{m} K(l) = K(m) + K(l)$$

thus $K(l^m) = mK(l)$.

Given three integers r, n, l let m be such that $l^m \le r^n \le l^{m+1}$, i.e. $\frac{m}{n} \le \frac{\log r}{\log l} \le \frac{m}{n} + \frac{1}{n}$. Since

$$mK(l) = K(l^m) \le K(r^n) = nK(r) \le K(l^{m+1}) = (m+1)K(l)$$

one obtains $\frac{m}{n} \le \frac{K(r)}{K(l)} \le \frac{m}{n} + \frac{1}{n}$, i.e. $\left|\frac{K(r)}{K(l)} - \frac{\log r}{\log l}\right| \le \frac{1}{n}$. Thus $\frac{K(r)}{\log r} = \frac{K(l)}{\log l}$ and $K(m) = c \log m$, $c > 0$.

Let $(p_1, \ldots, p_m) \in \mathbb{Q}^m \cap \Delta^{(m)}$ and let s denote the least common multiple of their denominators. Then $p_i = \frac{r_i}{s}$ and $\sum_{i=1}^m r_i = s$.

In addition to the partition α with elements $A_1, \ldots, A_m$ and associated probabilities con $p_1, \ldots, p_m$ we also consider β with s outcomes $B_1, \ldots, B_s$ which we divide into m groups each of them containing $r_1, \ldots, r_m$ outcomes respectively. Let $\pi_{ij} = \frac{p_i}{r_i} = \frac{1}{s}$, $i = 1, \ldots, m$, $j = 1, \ldots, r_i$.

Given any outcome A_i of α the possible r_i outcomes of β are equally probable thus

$$H\left(\frac{\pi_{i1}}{p_i}, \ldots, \frac{\pi_{ir_i}}{p_i}\right) = c \log r_i \quad \text{and}$$

$$\sum_{i=1}^m p_i H\left(\frac{\pi_{i1}}{p_i}, \ldots, \frac{\pi_{ir_i}}{p_i}\right) = c \sum_{i=1}^m p_i \log r_i = c \sum_{i=1}^m p_i \log p_i + c \log s.$$

On the other hand $H(\pi_{i1}, \ldots, \pi_{mr_m}) = c \log s$ and by (5)

$$\begin{aligned} H(p_1, \ldots, p_m) &= H(\pi_{i1}, \ldots, \pi_{mr_m}) - \sum_{i=1}^m p_i H\left(\frac{\pi_{i1}}{p_i}, \ldots, \frac{\pi_{ir_i}}{p_i}\right) \\ &= -c \sum_{i=1}^m p_i \log p_i, \end{aligned}$$

thus [4.3] holds on a dense subset of $\Delta^{(m)}$. By continuity it must hold everywhere. □

The entropy H can be regarded as $-\frac{1}{N}\times$ the logarithm of the probability of a "typical" result of the experiment α repeated N times. Indeed, if N is large and α is repeated N times, by the law of large numbers one should observe each A_i approximately $p_i N$ times. Thus the probability of a "typical" outcome is $p_1^{p_1 N} p_2^{p_2 N} \ldots p_m^{p_m N}$.

We now want to extend the notion of entropy to measurable dynamical systems $(X, \mathcal{A}, \mu, f)$.

If α and β are two partitions of X, their *joint partition* $\alpha \vee \beta$ is $\{A \cap B\ ,\ A \in \alpha,\ B \in \beta\}$. Given n partitions $\alpha_1, \ldots, \alpha_n$ we will denote $\bigvee_{i=1}^n \alpha_i$ their joint partition. If f is measurable and $f^{-1}(A) \in \mathcal{A}$ for all $A \in \mathcal{A}$, and α is a partition, $f^{-1}\alpha$ is the partition defined by the subsets $\{f^{-1}A\,,\ A \in \alpha\}$. Finally a partition β is *finer* than α, denoted $\alpha < \beta$, if $\forall\ B \in \beta\ \exists\ A \in \alpha$ such that $B \subset A$.

The entropy $H(\alpha)$ of a partition $\alpha = \{A_1, \ldots, A_m\}$ is given by $H(\alpha) = -\sum_{i=1}^{m} \mu(A_i) \log \mu(A_i)$.

Definition 4.17 *Let $(X, \mathcal{A}, \mu, f)$ be a measurable dynamical system and α a partition. The* entropy of f w.r.t. the partition α *is*

$$h_\mu(f, \alpha) := \lim_{n\to\infty} \frac{1}{n} H\left(\bigvee_{i=0}^{n-1} f^{-i}\alpha\right) \tag{4.4}$$

The entropy of f *is*

$$h_\mu(f) := \sup\{h(S, \alpha),\ \alpha \text{ is a finite partition of } X\}. \tag{4.5}$$

Remark 4.18 Using the strict convexity of $x \log x$ on $\mathbb{R}_+$, one can prove the existence of the limit [4.4]. Indeed the sequence $\frac{1}{n} H\left(\bigvee_{i=0}^{n-1} f^{-i}\alpha\right)$ is non–negative monotonic non increasing. Thus $h_\mu(f, \alpha) \geq 0$ for all α.

Exercise 4.19 Show that if two measurable dynamical systems are isomorphic then they have the same entropy.

The above considerations show that the entropy of a partition α measures the amount of information obtained making a measurement by means of a device which distinguishes points of X with the resolution prescribed by $\{A_1, \ldots, A_m\} = \alpha$. If $x \in X$ and we consider the orbit of x up to time $n-1$ $x,\ fx,\ f^2x,\ \ldots,\ f^{n-1}x$, since α is a partition mod(0) of X the points $f^i x$, $0 \leq i \leq n-1$, belong (almost surely) to exactly one of the sets of α : $x_i \in A_{k_i}$ with $k_i \in \{1, \ldots, m\}$ for all $i = 0, \ldots, n-1$. $H\left(\vee_{i=0}^{n-1} f^{-i}\alpha\right)$ measures the information obtained from the knowledge of the distribution w.r.t. α of a segment of orbit of length n. Thus $\frac{1}{n} H\left(\bigvee_{i=0}^{n-1} f^{-i}\alpha\right)$ is the average amount of information per unit of time and $h_\mu(S, \alpha)$ is the amount of information (asymptotically) obtained at each iteration of the dynamical system from the knowledge of the distribution of the orbit of a point w.r.t. the partition α.

A more satisfactory formulation of this is given by the following theorem [Mn1].

Theorem 4.20 (Shannon-Breiman-McMillan) *Let $(X, \mathcal{A}, \mu, f)$ be an ergodic measurable dynamical system, α a finite partition of X. Given $x \in X$ let $\alpha^n(x)$ be the element of $\bigvee_{i=0}^{n-1} f^{-i}\alpha$ which contains x. For μ–a.e. $x \in X$ one has*

$$h_\mu(f, \alpha) = \lim_{n\to\infty} -\frac{1}{n} \log \mu(\alpha^n(x)). \tag{4.6}$$

Remark 4.21 The previous theorem admits the following interpretation : if a system is ergodic then there exists a non–negative number h such that $\forall\ \varepsilon > 0$ if α is a sufficiently fine partition of X then there exists a positive integer N such that for all $n \geq N$ there is a subset X_n of X with measure $\mu(X_n) > 1 - \varepsilon$ and made of approximately e^{nh} elements of $\bigvee_{i=0}^{n-1} S^{-i}\alpha$, each measuring about e^{-nh}.

Let X be a compact metric space and $\mathcal{A}$ be the Borel σ-algebra. Brin e Katok [M. Brin and A. Katok, Lecture Notes in Mathematics 1007 (1983) 30–38] gave a "topological version" of Shannon-Breiman-McMillan's Theorem. Let $B(x, \varepsilon)$ be the ball of center $x \in X$ and radius ε. Let $f : X \to X$ be continuous and preserving the probability measure $\mu : \mathcal{A} \to [0, 1]$. Let

$$B(x, \varepsilon, n) := \{y \in X \mid d(f^i x, f^i y) \leq \varepsilon \text{ forall } i = 0, \ldots, n-1\},$$

i.e. $B(x, \varepsilon, n)$ is the set of points $y \in X$ whose orbit stays at a distance at most ε from the orbit of x for at least $n - 1$ iterations. Then one has

Theorem 4.22 (Brin-Katok) *Assume that $(X, \mathcal{A}, \mu, f)$ is ergodic. For μ–a.e. $x \in X$ one has*

$$\sup_{\varepsilon>0} \limsup_{n\to\infty} -\frac{1}{n} \log \mu(B(x, \varepsilon, n)) = h_\mu(f). \quad [4.7]$$

When the entropy is positive some of the observables are not predictable.

A system is *chaotic* if it has *positive entropy*. Brin-Katok's Theorem together with Poincaré recurrence theorem show that the orbits of chaotic systems are subject to two apparently contrasting requirements. On one hand almost every orbit is recurrent. On the other hand the probability that two orbits stay close to each other for an inteval of time of length n decays exponentially with n. Since two initially close orbits must come infinitely many times close to their origin, if the entropy is positive they cannot be correlated. Tipically they will separate one from the other and return at different times n.

To this complexity of the motions one associates the notion of chaos and shows how it can be impossible to compute the values that an observable will assume from the knowledge of the past.

Remark 4.23 To compute the entropy one can use the following important result of Kolmogorov and Sinai : if α is a partition of X which generates the σ-algebra $\mathcal{A}$ the entropy of $(X, \mathcal{A}, \mu, f)$ is simply given by

$$h_\mu(f) = h_\mu(f, \alpha). \quad [4.8]$$

We recall that α *generates* $\mathcal{A}$ iff $\bigvee_{-\infty}^{+\infty} f^{-i}\alpha = \mathcal{A} \bmod(0)$ if f is invertible, $\bigvee_{i=0}^{\infty} f^{-i}\alpha = \mathcal{A} \bmod(0)$ if f is not invertible.

Exercise 4.24 Show that the entropy of the p–adic map is $\log p$.

Exercise 4.25 Interpret formula [4.2] in terms of information (so as its analogy with [4.4] is clear).

4.3 Shifts and Bernoulli schemes

Let $N \geq 2$, $\Sigma_N = \{1, \ldots N\}^{\mathbb{Z}}$. For $x = (x_i)_{i \in \mathbb{Z}}$, $y = (y_i)_{i \in \mathbb{Z}}$ we define their distance

$$d(x, y) = 2^{-a(x,y)} \quad \text{where } a(x, y) = \inf\{|n|\,,\ n \in \mathbb{Z}\,,\ x_n \neq y_n\}\ . \qquad [4.9]$$

Then (Σ_N, d) is a compact (ultra)–metric space. The *shift* $\sigma : \Sigma_N \to \Sigma_N$ is the bilipschitzian homeomorphism of Σ_N (the Lipschitz constant is N) defined by

$$\sigma((x_i)_{i \in \mathbb{Z}}) = (x_{i+1})_{i \in \mathbb{Z}}\ . \qquad [4.10]$$

Topological properties of the shift map :

- The phase space Σ_N is totally disconnected and has Hausdorff dimension 1.
- The homeomorphism σ is *expansive* : for all $x \neq y$ there exists n such that $d(\sigma^n(x), \sigma^n(y)) \geq 1$.
- The topological entropy of (Σ_N, σ) is $\log N$.

Let $(p_1, \ldots, p_N) \in \Delta^{(N)}$ and let ν be the probability measure on $\{1, \ldots N\}$ such that $\nu(\{i\}) = p_i$.

Definition 4.26 *The* Bernoulli scheme $BS(p_1, \ldots, p_N)$ *is the measurable dynamical system given by the shift map* $\sigma : \Sigma_N \to \Sigma_N$ *with the (product) probability measure* $\mu = \nu^{\mathbb{Z}}$ *on* Σ_N.

Exercise 4.27 Show that the σ–algebra of measurable subsets of Σ_N coincides with its Borel σ–algebra and its generated by *cylinders* : if $j_1, \ldots, j_k \in \{1, \ldots N\}$ and $i_1, \ldots, i_k \in \mathbb{Z}$ the corresponding cylinder is

$$C\begin{pmatrix} j_1, \ldots, j_k \\ i_1, \ldots, i_k \end{pmatrix} = \{x \in \Sigma_N \mid x_{i_1} = j_1,\ x_{i_2} = j_2, \ldots, x_{i_k} = j_k\}\ , \qquad [4.11]$$

Check that the measure of cylinders for the Bernoulli scheme $BS(p_1, \ldots, p_N)$ is

$$\mu\left(C\begin{pmatrix} j_1, \ldots, j_k \\ i_1, \ldots, i_k \end{pmatrix}\right) = p_{j_1} \cdots p_{j_k}\ , \qquad [4.12]$$

and that it is preserved by the shift map.

Proposition 4.27 *The Kolmogorov–Sinai entropy of the Bernoulli scheme* $BS(p_1, \ldots, p_N)$ *is* $-\sum_{i=1}^{N} p_i \log p_i$.

Proof. The partition α defined by the cylinders $\left\{C\binom{j}{0}\right\}_{j=1,\dots,N}$ generates the *sigma*-algebra $\mathcal{A}$. By Remark 4.22 we can thus use it to compute the entropy. Since

$$\alpha \cup \sigma^{-1}\alpha = \left\{C\begin{pmatrix} j_0 & j_1 \\ 0 & 1 \end{pmatrix}\right\}_{j_0,j_1=1,\dots,N}$$

$$\alpha \cup \sigma^{-1}\alpha \cup \sigma^{-2}\alpha = \left\{C\begin{pmatrix} j_0 & j_1 & j_2 \\ 0 & 1 & 2 \end{pmatrix}\right\}_{\substack{j_i=1,\dots,N,\\ i=0,1,2}}$$

and so on, and the corresponding entropies are

$$H(\alpha) = -\sum_{j=1}^{N} p_j \log p_j$$

$$\begin{aligned} H(\alpha \cup \sigma^{-1}\alpha) &= -\sum_{j_0=1}^{N}\sum_{j_1=1}^{N} p_{j_0}p_{j_1}\log p_{j_0}p_{j_1} = \\ &= -\sum_{j_0=1}^{N}(p_{j_0}\log p_{j_0}\sum_{j_1=1}^{N} p_{j_1} - \sum_{j_1=1}^{N}(p_{j_1}\log p_{j_1}\sum_{j_0=1}^{N} p_{j_0} = \\ &= -2\sum_{j=1}^{N} p_j \log p_j \end{aligned}$$

$$H(\alpha \cup \sigma^{-1}\alpha \cup \sigma^{-2}) = -\sum_{j_0,j_1,j_2} p_{j_0}p_{j_1}p_{j_2}\log p_{j_0}p_{j_1}p_{j_2} = -3\sum_{j=1}^{N}(p_j \log p_j$$

and so on. Thus $h_\mu(\sigma,\alpha) = -\sum_{j=1}^{N} p_j \log p_j$. □

Remark 4.28 Note that $h_\mu(\sigma) \le \log N$ for all $(p_1,\dots,p_N) \in \Delta^{(N)}$ with equality if and only if $p_i = 1/N$ for all i for which we get the *unique* invariant measure of the shift on N symbols which realizes the topological entropy.

Let us see how the shift and the shift–invariant compact subsets of Σ_N arise naturally in the context of *symbolic dynamics* (the following description is taken from the lectures of J.–C. Yoccoz at the 1992 ICTP School on Dynamical Systems).

Let (Y,d) be a compact metric space and f a homeomorphism of Y. Let $Y = Y_1 \cup \ldots \cup Y_N$, where the Y_i are compact. Given a point $y \in Y$ we define

$$\Sigma(f,y) = \{x \in \Sigma_N \,,\; f^i(y) \in Y_{x_i}\; \forall i \in \mathbb{Z}\}\ .$$

This is a nonempty compact subset of Σ_N. Moreover we define

$$\begin{aligned} \Sigma(f) &= \cup_{y\in Y}\Sigma(f,y) \\ &= \{x \in \Sigma_N \,,\; \cap_{i\in\mathbb{Z}} f^{-i}(Y_{x_i}) \ne \emptyset\}\ . \end{aligned}$$

Exercise 4.29 Show that $\Sigma(f)$ is also a compact subset of Σ_N, invariant under the shift. [Hint : $\Sigma(f, f(y)) = f(\Sigma(f, y))$.]

Assume that the map f is *expansive*, i.e. there exists $\varepsilon > 0$ such that for all $y_1 \neq y_2$ there exists an integer n such that $d(f^n(y_1), f^n(y_2)) > \varepsilon$, and choose the compacts Y_i above with $\mathrm{diam}(Y_i) < \varepsilon$.

Then by expansivity if $y_1 \neq y_2$ the sets $\Sigma(f, y_1)$ and $\Sigma(f, y_2)$ are disjoint and we can define a map $h : \Sigma(f) \to Y$ by the property $h^{-1}(y) = \Sigma(f, y)$.

Exercise 4.30 Show that h is surjective, continuous and $h \circ \sigma = f \circ h$, i.e. h is a semiconjugacy from the restriction of the shift σ to $\Sigma(f)$ to f.

Exercise 4.31 Show that the semiconjugacy above is indeed a topological conjugacy if and only if Y is totally disconnected (and f is expansive). [Hint : choose the compacts Y_i with $\mathrm{diam}(Y_i) < \varepsilon$ and disjoint.]

4.4 (Topological) Markov chains and Markov maps

The discussion at the end of the previous section shows the importance of the shift invariant compact subsets of Σ_N. Among these a very important subclass are the so–called *topological Markov chains* or *subshifts of finite type.*

Let $\Gamma \subset \{1, \dots N\}^2$ and let $\overrightarrow{\Gamma}$ be a connected directed graph on the vertices $\{1, \dots N\}$ with at most one arrow between two vertices : there is an arrow from i to j if and only if $(i, j) \in \Gamma$.

We denote $A = A_\Gamma$ the $N \times N$ matrix with entries $a_{ij} \in \{0, 1\}$ defined as follows :

$$a_{ij} = \begin{cases} 1 & \iff (i,j) \in \Gamma \iff \text{there is an arrow in } \overrightarrow{\Gamma} \text{ from } i \text{ to } j \\ 0 & \text{otherwise} \end{cases}$$

We moreover assume that for all $i \in \{1, \dots N\}$ there exist $j, k \in \{1, \dots N\}$ such that $a_{ij} = a_{ki} = 1$.

We associate to the matrix A (or, equivalently, to the directed graph $\overrightarrow{\Gamma}$) the subset $\Sigma_A \subset \Sigma_N$ defined as follows :

$$\Sigma_A = \{x \in \Sigma_N \, , \, (x_i, x_{i+1}) \in \Gamma \, \forall i \in \mathbb{Z}\} \, .$$

Exercise 4.32 Show that Σ_A is a compact shift invariant subset of Σ_N.

The restriction of the shift σ to Σ_A is denoted σ_A and is called the *topological Markov chain* (or *subshift of finite type*) associated to the matrix A (equivalently to the graph $\overrightarrow{\Gamma}$).

Exercise 4.33 Show that $\mathrm{card}\,(\mathrm{Fix}(\sigma_A^n)) = \mathrm{Tr}(A^n)$ for all $n \in \mathbb{N}$. Deduce from this that the Artin-Mazur zeta function

$$\zeta_A(t) = \exp\left(\sum_{n=0}^{\infty} \frac{1}{n} \mathrm{card}\,(\mathrm{Fix}(\sigma_A^n)) t^n\right)$$

is rational (indeed it is equal to $\det(I - tA)^{-1}$).

The matrix A is called *primitive* if there exists a positive integer m such that all the entries of A^m are strictly positive : $A^m = (a_{ij}^m)$ and $a_{ij}^m > 0$ for all i, j. Then it is easy to show that for all $n \geq m$ one also has $a_{ij}^n > 0$ for all i, j.

Exercise 4.34 Show that if A is primitive then σ_A is topologically transitive, and its periodic orbits are dense in Σ_A. Moreover σ_A is topologically mixing (...).

When the matrix is primitive one can apply the classical Perron–Frobenius theorem to compute the topological entropy of the associated subshift.

Theorem 4.35 (Perron–Frobenius, see [Gan]) If A is primitive then there exists an eigenvalue $\lambda_A > 0$ such that :
(i) $|\lambda_A| > \lambda$ for all eigenvalues $\lambda \neq \lambda_A$;
(ii) the left and right eigenvectors associated to λ_A are strictly positive and are unique up to constant multiples ;
(iii) λ_A is a simple root of the characteristic polynomial of A.

Exercise 4.35 Assume that A is primitive. Show that the topological entropy of σ_A is $\log \lambda_A$ (clearly $\lambda_A > 1$ since all the integers $a_{ij}^m > 0$).

Very much as the shift on N symbols preserves many invariant measures (the Bernoulli schemes on N symbols) a topological Markov chain preserves many invariant measures (which are called *Markov chains*).

Let $P = (P_{ij})$ be an $N \times N$ matrix such that
(i) $P_{ij} \geq 0$ for all i, j, and $P_{ij} > 0 \iff a_{ij} = 1$;
(ii) $\sum_{j=1}^{N} P_{ij} = 1$ for all $i = 1, \ldots, N$;
(iii) P^m has all its entries strictly positive.

Such a matrix is called a *stochastic matrix.* Applying Perron–Frobenius theorem to P we see that 1 is a simple eigenvalue of P and there exists a normalized eigenvector $p = (p_1, \ldots, p_N) \in \Delta^{(N)}$ such that $p_i > 0$ for all i and

$$\sum_{i=1}^{N} p_i P_{ij} = p_j \ , \quad \forall\, 1 \leq i \leq N \ .$$

We define a probability measure μ on Σ_A corresponding to P prescribing its value on the cylinders :

$$\mu\left(C\begin{pmatrix} j_0, \ldots, j_k \\ i, \ldots, i+k \end{pmatrix}\right) = p_{j_0} P_{j_0 j_1} \cdots P_{j_{k-1} j_k} \ ,$$

for all $i \in \mathbb{Z}$, $k \geq 0$ and $j_0, \ldots, j_k \in \{1, \ldots, N\}$. It is called the *Markov measure* associated to the stochastic matrix P.

Exercise 4.36 Prove that the subshift σ_A preserves the Markov measure μ.

The subshift of finite type σ_A with the preserved measure μ is called a *Markov chain.*

Exercise 4.37 Show that the Kolmogorov–Sinai entropy of $(\Sigma_A, \mathcal{A}, \sigma_A, \mu)$ is

$$h_\mu(\sigma_A) = - \sum_{i,j=1}^{N} p_i P_{ij} \log P_{ij} \ .$$

Check that $h_\mu(\sigma_A) \leq h_{top}(\sigma_A)$.

One can prove that there exists a stochastic matrix P such that the entropy of the associated Markov chain is equal to the topological entropy of σ_A. Moreover this measure is unique (Parry measure, see [Mn1]).

Remark 4.38 There is another point of view which can be useful in studying topological Markov chains and their invariant Markov measures. Call a sequence $x \in \Sigma_A$ a configuration of a one–dimensional *spin system* (or Potts system) with configuration space Σ_A. Then part of the classical stastistical mechanics of spin systems [Ru] is just the ergodic theory of the topological Markov chain (the shift–invariant measures being interpreted as translation–invariant measures).

Remark 4.39 An interesting application of the symbolic dynamics method described at the end of Section 3 is the theory of piecewise expanding Markov maps of the interval (Exercise 2.21). Let $Y = [0,1]$, $f : Y \to Y$ piecewise monotonic and $\mathcal{C}^2$, i.e. there exists a *finite* decomposition of the interval $[0,1]$ in N subintervals $I_i = [a_i, a_{i+1})$, $(a_1 = 0,\ a_{N+1} = 1)$ on which f si monotonic and of class $\mathcal{C}^2$ on their closure. On each of these subintervals an inverse branch f_i^{-1} of f is well–defined. Assume moreover

- *Markov property* $f(I_i) = I_{k_i} \cup I_{k_i+1} \cup \ldots \cup I_{k_i+n_i}$;
- *aperiodicity* there exists an integer m such that $f^m(\overline{I_i}) = Y$ for all $i = 1, \ldots, N$;
- *eventual expansivity* some iterate of f has its derivative bounded away from 1 in modulus.

After Section 3 the symbolic dynamics of these maps is just a topological Markov chain. Moreover one can prove that there exists a unique invariant ergodic measure absolutely continuous w.r.t. the Lebesgue measure with piecewise continuous density bounded away from 0 and ∞. With this measure the system is isomorphic to the Markov chain with the Parry measure : see [AF]. The existence of an absolutely continuous invariant measure can be proven also under weaker assumptions, see the classical [LY].