NOTES FOR MATH 625, FALL 2018

NICOLAI HAYDN

Contents

Date: May 6, 2019.

1. Introduction

Given a map $T:\Omega\to\Omega$ on a space Ω which could be a manifold or simply a measure space. We look at longterm behaviour of orbits $\{T^jx:j\}$ for generic points $x\in\Omega$. For instance, if $A\subset\Omega$ is a subset and we can observe the frequency with which a point returns to A. That is, if we put $N_n(x)=|\{j:0\le j\le n-1, T^j\in A\}|$ for the number of hits in A the point x takes along the orbit segment of length n then we would like to find the limit $D(x)=\lim_{n\to\infty}\frac{N_n(x)}{n}$ if it exists.

1.1. **Example.** Let us consider the irrational rotation on the unit interval (or circle). Let $\alpha\in(0,1)\setminus\mathbb{Q}$ be irrational and define $R_\alpha:[0,1)\to[0,1)$ (here we use $\Omega=[0,1)$ which mod 1 is the circle $\mathbb{T}^1$) by

$$R_\alpha x = x+\alpha \mod 1.$$

Its iterates are $R_\alpha^j = x+j\alpha \mod 1$. We will later find out that for $\mathscr{L}^1$-functions f one has $\frac{1}{n}\sum_{j=0}^{n-1}f(R_\alpha^jx)\to\int f\,dx$ as $n\to\infty$ almost surely (pointwise ergodic theorem). In particular we can choose $f=\chi_{(a,b)}$, the characteristic function of the interval $A=(a,b)$ $(0\le a<b\le 1)$. Since $N_n(x)=\sum_{j=0}^{n-1}\chi_{(a,b)}R_\alpha^jx$ we obtain $D(x)=b-a$ almost surely.

The following interesting application is due to Arnold. Let $\Gamma=\{2^j:j=0,1,2,\dots\}$ be the powers of 2, i.e.

$$\Gamma=\{1,2,4,8,16,32,64,128,256,512,1024,2048,4096,8192,16384,32768,65536,131072\dots\}.$$

We look at their first digit (in base 10 expansion) and obtain the sequence

$$1,2,4,8,1,3,6,1,2,5,1,2,4,8,1,3,6,1\dots$$

and want to know what is the frequency of a digit occurring in first position. Let $k\in\{1,2,\dots,9\}$ be one of the ten digits and

$$N_k(n)=|\{j:0\le j\le n-1,\text{the decimal expansion of }2^j\text{ begins with the first digit }k\}|.$$

What is the limit of $\frac{N_k(n)}{n}$ as n goes to infinity? If the decimal expansion of 2^j begins with k then

$$2^j=ka_1a_2a_3\cdots a_m=k\cdot 10^m+a_1\cdot 10^{m-1}+a_2\cdot 10^{m-2}+\cdots a_{m-1}\cdot 10+a_m$$

for some m where $a_i\in\{0,1,\dots,9\}$. Thus

$$2^j=k\cdot 10^m+r_j$$

where the remainder $r_j\ge 0$ has at most m digits and is therefore less than 10^m. Therefore $k\cdot 10^m\le 2^j<(k+1)10^m$ and taking logarithms to the base 10 yields

$$\log_{10}k+m\le j\log_{10}2<\log_{10}(k+1)+m$$

and if we put $\alpha=\log_{10}2$ (which is an irrational number), then we get

$$\log_{10}k\le R_\alpha^j(0)<\log_{10}(k+1)$$

as $j\log_{10}2=R_\alpha^j(0)$ and R_α is the irrational rotation by α on the unit interval. Thus

$$N_k(n)=\sum_{j=0}^{n-1}\chi_{[\log_{10}k,\log_{10}(k+1))}(R_\alpha^j(0))$$

and by the result mentioned above

$$D_k = \lim_{n\to\infty} \frac{N_k(n)}{n} = \log_{10}(k+1) - \log_{10} k = \log_{10} \frac{k+1}{k}.$$

In particular $D_1 = \log_{10} 2, D_2 = \log_{10} \frac{3}{2}, \dots, D_9 = \log_{10} \frac{10}{9}$ and of course $\sum_{k=0}^{9} D_k = 1$. Naturally this game can be played with any base d expansion for which $\log_d 2$ is irrational (don't use the binary expansion $d = 2$).

1.2. Poincaré recurrence theorem.

Theorem 1. *(Poincaré recurrence theorem) Let $T : \Omega \to \Omega$ and μ be a T-invariant probability measures. For $U \subset \Omega$ put $\tau_U(x) = \min\{k \geq 1 : T^k x \in U\}$ for the return time of $x \in \Omega$ (we have $\tau_U(x) = \infty$ if the forward orbit of x never intersects U). If $\mu(U) > 0$, then $\tau_U(x) < \infty$ for almost every $x \in U$.*

Proof. Let $U \subset \Omega$ have positive measure and put $U_n = \bigcup_{j=n}^{\infty} T^{-j} U$ for the set of points $x \in \Omega$ that enter U at least once after time n. Obviously $U_0 \supset U_1 \supset U_2 \supset \cdots$. We also have $U_n = T^{-1} U_{n+1}$ which implies by invariance of the measure that $\mu(U_n) = \mu(T^{-1} U_{n+1}) = \mu(U_{n+1})$ and consequently $\mu(U_0) = \mu(U_n)\ \forall\ n$. Now $W = \bigcap_{n=1}^{\infty} U_n = \{x \in \Omega \text{ enters } U \text{ infinitely often}\}$ and $V = W \cap U = \{x \in U \text{ enters } U \text{ infinitely often}\}$. Since $\mu(U_0) = \mu(U_n)$ we obtain that $\mu(W) = \mu(U_0)$ and since $U \subset U_0$ we conclude that $\mu(V) = \mu(U)$. ∎

The recurrence statement is not true if the measure is infinite. As an example one can take the Lebesgue measure on $\mathbb{R}$ and the map $T : \mathbb{R} \to \mathbb{R}$ given by $Tx = x + 1$. No set of positive measure is recurrent.

2. Ergodic theorems

Let (Ω, T, μ) be a dynamical system that consists of a space Ω, a map $T : \Omega \to \Omega$ and a T-invariant probability measure μ on Ω (μ is *T-invariant* if $\mu(U) = \mu(T^{-1}U)$ for all measurable $U \subset \Omega$).

For a real valued (or complex valued) function f on Ω we write $f^n = \sum_{j=0}^{n-1} f \circ T^j$ (nth ergodic sum) and put $\frac{1}{n} f^n$ for the time average along orbit segments of lengths n. If n goes to infinity then we obtain the time average of f over an orbit. Ergodic theorems are concerned with the existence of the limit and its value. Ergodic theorems relate these time averages to spacial average $\int_\Omega f(x)\, d\mu(x)$. We first prove the von Neuman Mean Ergodic Theorem which was published in 1932.

Theorem 2. *(Mean Ergodic Theorem) Let $S = \{g \in \mathscr{L}^2 : g \circ T = g\}$ and P the projection from $\mathscr{L}^2$ to S. Then for all $f \in \mathscr{L}^2$ one has*

$$\frac{1}{n} f^n \to Pf$$

as $n \to \infty$ in $\mathscr{L}^2$.

Proof. Let us define the *Koopman operator* $U : \mathscr{L}^2 \to \mathscr{L}^2$ by $Uf = f \circ T$. If T is invertible then so it U. Let us note that U is an isometry on $\mathscr{L}^2$. This is because by

T-invariance of μ one has

$$\begin{aligned}(Uf, Ug) &= \int_\Omega (Uf)(x)(\bar{Ug})(x)\, d\mu(x)\\ &= \int_\Omega (f\bar{g})\circ T(x)\, d\mu(x)\\ &= \int_\Omega (f\bar{g})(x)\, d\mu(x)\\ &= (f, g).\end{aligned}$$

This identity also implies that $(f, g) = (Uf, Ug) = (U^*Uf, g)$ for all $f, g \in \mathscr{L}^2$. Hence U^*U is the identity operator on $\mathscr{L}^2$.

Let us now put $W = \{Ug - g : g \in \mathscr{L}^2\}$. We claim that $W^\perp = S$. To show that $S \subset W^\perp$ let $f \in \mathscr{L}^2$ and $g \in W$. That is $Uf = f$ and there exists an $h \in \mathscr{L}^2$ so that $g = Uh - h$. Then

$$(f, g) = f, Uh - h) = (f, Uh) - (f, h) = (Uf, Uh) - (f, h) = (f, h) - (f, h) = 0$$

and therefore $f \in W^\perp$. Since $f \in S$ was arbitrary, we obtain $S \subset W^\perp$. To get the inclusion $W^\perp \subset S$ let $f \in W^\perp$ which means $(f, g) = 0$ for all $g \in W$. Since $g = Uh - h$ for some $h \in \mathscr{L}^2$ we get $(f, Uh-h) = 0$ implies that $(f, Uh) = (f, h)$ and $(U^*f, h) = (f, h)$. Since h was arbitrary we get $U^*f = f$. With this we now get

$$\begin{aligned}\|Uf - f\|^2 &= (Uf - f, Uf - f)\\ &= (Uf, Uf) - (f, f) - (Uf, f) - (f, Uf)\\ &= 2\|f\|^2 - (f, U^*f) - (U^*f, f)\\ &= 2\|f\|^2 - 2(f, f) = 0.\end{aligned}$$

Hence $f \in S$.

Now if $f \in S$, then $\frac{1}{n}f^n = \frac{1}{n}\sum_{j=0}^{n-1} U^j f = \frac{1}{n}nf = f = Pf$ as $U^j f = f$. If $g = Uh - h \in W$, then $\sum_{j=0}^{n-1} U^j g = \sum_{j=1}^{n-1}(U^{j+1}h - U^j h) = U^n h - h$ which implies that $\frac{1}{n}\sum_{j=0}^{n-1} U^j g = \frac{1}{n}(U^n h - h) \to 0 = Pg$ as $n \to \infty$

We have $\mathscr{L}^2 = S \oplus \bar{W}$. Let $g \in \bar{W}$, then there exists a sequence $g_i \in W$ which in $\mathscr{L}^2$ converges to g. Let $\varepsilon > 0$ and i so that $\|g - g_i\|_2 < \varepsilon$. Then

$$\frac{1}{n}\|\sum_{j=0}^{n-1} U^j g\|_2 \le \frac{1}{n}\sum_{j=0}^{n-1}\|U^j(g - g_i)\|_2 + \frac{1}{n}\|\sum_{j=0}^{n-1} U^j g_i\|_2 < 2\varepsilon$$

as U is an isometry and we can choose n large enough so that the second sum is less than ε. Since $\varepsilon > 0$ was arbitrary we get that $\frac{1}{n}\sum_{j=0}^{n-1} U^j g \to 0$ in $\mathscr{L}^2$ as $n \to \infty$. For arbitrary $F \in \mathscr{L}^2$ we write $F = f + g$ where $f \in S$ and $g \in \bar{W}$. Then $\frac{1}{n}F^n = \frac{1}{n}f^n + \frac{1}{n}g^n \to f$ as $f = PF$. ∎

Corollary 3. *(Mean Ergodic Theorem in $\mathscr{L}^1$) Let $f \in \mathscr{L}^1$, then $\frac{1}{n}f^n \to f^*$ in $\mathscr{L}^1$ as $n \to \infty$, where $f^* \in \mathscr{L}^1$ is a T-invariant function.*

Proof. We use the fact that bounded $\mathscr{L}^1$ functions are dense in $\mathscr{L}^1$ and are in $\mathscr{L}^2$. In other words, the set $\mathscr{L}^1_b = \{g \in \mathscr{L}^1 : \|g\|_\infty < \infty\}$ is dense in $\mathscr{L}^1$ and also lies in $\mathscr{L}^2$. If

$g \in \mathscr{L}^1_b$, then by the Mean Ergodic Theorem $\frac{1}{n}g^n \to g^*$, where $g^* \in S \subset \mathscr{L}^2$ ($g^* = Pg$). Also, as $\|g^n\|_\infty \le n\|g\|_\infty$, we get $\|\frac{1}{n}g^n\|_\infty \le \|g\|_\infty$ which implies that $\|g^*\|_\infty \le \|g\|_\infty < \infty$. Since $\|\cdot\|_1 \le \|\cdot\|_2$ we get that $\frac{1}{n}g^n$ converges in $\mathscr{L}^1$ to g^*.

Now let $f \in \mathscr{L}^1$ and $\varepsilon > 0$, then there exists a $g \in \mathscr{L}^1_b$ so that $\|f - g\|_1 < \varepsilon$ and $\|\frac{1}{n}g^n - g^*\|_1 < \varepsilon$ for all n big enough. As $\|\frac{1}{n}f^n - \frac{1}{n}g^n\|_1 < \varepsilon$ for all n, we obtain $\|\frac{1}{n}f^n - g^*\|_1 < 2\varepsilon$ for all n big enough. Eliminating g^* from the estimates yields $\|\frac{1}{n}f^n - \frac{1}{m}f^m\|_1 < 4\varepsilon$ for all n, m big enough. This means that $\{\frac{1}{n}f^n : n = 1, 2, \dots\}$ is a Cauchy sequence which has a limit f^* in $\mathscr{L}^1$. Obviously f^* is T-invariant. ∎

The Mean Ergodic Theorem is strengthened considerably by the Pointwise Ergodic Theorem which is due to Birkhoff and was published in 1931 although it was preceded by von Neuman's MET which appeared later.

Theorem 4. *(Pointwise Ergodic Theorem) For $f \in \mathscr{L}^1$, then the limit*

$$\lim_{n\to\infty} \frac{1}{n} f^n(x) = f^*(x)$$

exists almost everywhere, where $f^* \in \mathscr{L}^1$ is T-invariant and satisfies $\int_\Omega f^*\,d\mu = \int_\Omega f\,d\mu$.

The main ingredient in the proof of the PET is the following result:

Theorem 5. *(Maximal Ergodic Theorem, Garsia 1965) Put $E(f) = \{x \in \Omega : \sup_{n\ge1} f^n(x) > 0\}$. Then $\int_{E(f)} f\,d\mu \ge 0$.*

Corollary 6. *If $E_\alpha(g) = \{x \in \Omega : \sup_{n\ge1} \frac{1}{n}g^n(x) > \alpha\}$, then $\int_{E_\alpha(g)} g\,d\mu \ge \alpha\mu(E_\alpha(f))$.*

Proof. We put $f = g - \alpha$ which implies that $f^n = g^n - n\alpha$ and thus gives us $E(f) = E_\alpha(g)$ and therefore by the Maximal Ergodic Theorem

$$0 \le \int_{E(f)} f\,d\mu = \int_{E_\alpha(g)} (g-\alpha)\,d\mu = \int_{E_\alpha(g)} g\,d\mu - \alpha\mu(E_\alpha(g)).$$

∎

Corollary 7. *If $A \subset E(f)$ is T-invariant, i.e. $T^{-1}A = A$, then $\int_A f\,d\mu \ge 0$.*

Proof. Let χ_A be the characteristic function of A. As A is T-invariant $\chi_A \circ T^j = \chi_A$ and therefore

$$(f\chi_A)^n = \sum_{j=0}^{n-1} (\chi_A f) \circ T^j = \chi_A f^n$$

which leads to

$$\begin{aligned} E(f\chi_A) &= \{x \in \Omega : \sup_{n\ge1} (f\chi_A)^n(x) > 0\} \\ &= \{x \in \Omega : \chi_A(x) \sup_{n\ge1} f^n(x) > 0\} \\ &= E(f) \cap A = A \end{aligned}$$

as $A \subset E(f)$. Thus by the Maximal Ergodic Theorem

$$\int_A f\,d\mu = \int_{E(f\chi_A)} f\chi_A\,d\mu \ge 0.$$

∎

Combining the last two corollaries yields the following result.

Corollary 8. *If $A \subset E_\alpha(f)$ is T-invariant, then $\int_A f\,d\mu \ge \alpha\mu(A)$.*

Proof of the Pointwise Ergodic Theorem. Put

$$\begin{aligned} f^+(x) &= \limsup_{n\to\infty} \frac{1}{n} f^n(x) \\ f^-(x) &= \liminf_{n\to\infty} \frac{1}{n} f^n(x). \end{aligned}$$

Obviously $f^-(x) \le f^+(x)$ and the two functions are T-invariant. We have to show that $f^-(x) = f^+(x)$ almost surely, i.e. we have to show that the set $E = \{x \in \Omega : f^-(x) < f^+(x)\}$ has zero μ-measure. Let $\alpha < \beta$ and put $E_{\alpha,\beta} = \{x \in \Omega : f^-(x) < \alpha, f^+(x) > \beta\}$. Since the functions f^-, f^+ are T-invariant, the set $E_{\alpha,\beta}$ too is T-invariant. We will prove that $\mu(E_{\alpha,\beta}) = 0$. Put $E^+_\beta = \{x \in \Omega; f^+(x) > \beta\}$. One has $E_{\alpha,\beta} \subset E^+_\beta \subset E(f-\beta)$ since if $x \in E^+_\beta$ then $\frac{1}{k} f^k(x) > \beta$ for some $k \in \mathbb{N}$. Hence $f^k(x) - k\beta > 0$ which implies that $x \in E(f-\beta)$. Thus by the above corollaries

$$\int_{E_{\alpha,\beta}} (f-\beta)\,d\mu \ge 0$$

which implies

$$\int_{E_{\alpha,\beta}} f\,d\mu \ge \beta\mu(E_{\alpha,\beta}).$$

Similarly one shows that $\int_{E_{\alpha,\beta}} \le \alpha\mu(E_{\alpha,\beta})$. One therefore gets that

$$\beta\mu(E_{\alpha,\beta}) \le \int_{E_{\alpha,\beta}} f\,d\mu \le \alpha\mu(E_{\alpha,\beta})$$

which for $\alpha < \beta$ can only be satisfied if $\mu(E_{\alpha,\beta}) = 0$. To represent E as a countable union, one restricts to rational values for α and β. Thus, since

$$E = \bigcup_{\alpha<\beta;\alpha,\beta\in\mathbb{Q}} E_{\alpha,\beta}$$

we get that $\mu(E) = 0$. Thus the limit exists. From the Mean Ergodic Theorem we know that $\frac{1}{n} f^n$ converges in $\mathscr{L}^1$ to the limit f^*. Since $\int \frac{1}{n} f^n\,d\mu = \int f\,d\mu$ we get that $\int f^*\,d\mu = \int f\,d\mu$. ∎

Proof of the Maximal Ergodic Theorem. Put $F_N(x) = \max_{0\le k\le N} f^k(x)$ where we use the convention that $f^0 = 0$. Thus $F_N \ge 0$ and forms an increasing sequence: $\cdots \le F_N \le F_{N+1} \le \dots$. The sets $E_N = \{x \in \Omega : F_N(x) > 0\}$ form a nested sequence which gives $E = \bigcup_N E_N$. Note that $F_N(x) = 0$ for $x \notin E_N$. Now

$$F_N(Tx) + f(x) = f(x) + \max_{0\le k\le N} f^k(Tx) \ge f(x) + f^k(Tx) = f^{k+1}(x)$$

for any $k = 0, \dots, N$. Thus $f(x) \ge f^k(x) - F_N(Tx)$ for $k = 1, 2, \dots, N+1$, and for $x \in E_N$ one has $F_N(x) = \max_{1\le k\le N} f^k(x)$ which then gives $f(x) \ge F_N(x) - F_N(Tx)$.

Hence, since $F_N = 0$ on $\Omega \setminus E_N$ and is otherwise non-negative we get

$$\begin{aligned}\int_{E_N} f\,d\mu &\geq \int_{E_N} F_N\,d\mu - \int_{E_N} F_N \circ T\,d\mu \\ &= \int_{\Omega} F_N\,d\mu - \int_{E_N} F_N \circ T\,d\mu \\ &\geq \int_{\Omega} F_N\,d\mu - \int_{\Omega} F_N \circ T\,d\mu = 0\end{aligned}$$

by invariance of the measure. By the Dominated Convergence Theorem one now gets $\int_{E_N} f\,d\mu \to \int_E f\,d\mu \geq 0$. ∎

Theorem 9. *(Mean Ergodic Theorem in $\mathscr{L}^p$) Let $p \in [1, \infty)$, then if $f \in \mathscr{L}^p$ there exists an $f^* \in \mathscr{L}^p$, T-invariant, so that $\frac{1}{n} f^n \to f^*$ in $\mathscr{L}^p$.*

Proof. Let $f \in \mathscr{L}^p$ and $\varepsilon > 0$. We approximate f by a bounded function $g \in \mathscr{L}^p$ so that $\|f - g\|_p < \varepsilon$. Then
(i) $\|\frac{1}{n} f^n - \frac{1}{n} g^n\|_p \leq \|f - g\|_p < \varepsilon$ for all n.
(ii) $g^* = \lim_n \frac{1}{n} g^n$ exists by the Pointwise Ergodic Theorem and is moreover a bounded function ($\|g^*\|_\infty = \|g\|_\infty < \infty$).
(iii) $f^* = \lim_n \frac{1}{n} f^n$ exists because one can approximate for every $k \in \mathbb{N}$ one has

$$f_k(x) = \begin{cases} f(x) & \text{if} \quad |f(x)| < k \\ 0 & \text{if} \quad |f(x)| \geq k \end{cases}$$

now let $k \to \infty$.
(iv) $\|f^* - g^*\|_p \leq \|f - g\|_p < \varepsilon$, since by Fatou's lemma

$$\liminf_{n\to\infty} \int \left| \frac{1}{n}(f^n - g^n) \right|^p d\mu \leq \int \liminf_{n\to\infty} \frac{1}{n} |f^n - g^n|^p \, d\mu = \int |f^* - g^*|^p \, d\mu < \varepsilon^p.$$

Hence $\|\frac{1}{n} f^n - f^*\|_p < 3\varepsilon$ for n big enough. ∎

3. Ergodicity

Here we want to focus on 'primitive' measures, which are probability measures that are minimal in the sense that they don't have genuine invariant subsets.

Definition 10. *Let $(\Omega, \mathscr{B}, \mu)$ be a probability space ($\mathscr{B}$ a σ-algebra) and $T : \Omega \to \Omega$ a measure preserving map. We say μ is* ergodic *if for all T-invariant $A \in \mathscr{B}$ $\mu(A)$ is either 0 or 1.*

Consequences:

(I) μ ergodic $\iff$ if $\mu(B \triangle T^{-1}B) = 0$ then $\mu(B) = 0, 1$.

Proof. "$\Rightarrow$": Assume $\mu(B \triangle T^{-1}B) = 0$ and put $C_n = \bigcup_{k=n}^{\infty} T^{-k}B$. Then $C_n \supset C_{n+1} \supset \cdots$ and $T^{-1}C_n = C_{n+1}$. Hence $C = \bigcap_n C_n$ is T-invariant, i.e. $T^{-1}C = C$. Now note that

$$B \triangle C_n = B \triangle \bigcup_{j=n}^{\infty} T^{-j}B \subset \bigcup_{j=n}^{\infty} (B \triangle T^{-j}B)$$

and since $B\triangle T^{-j}B\subset\bigcup_{k=0}^{j-1}(T^{-k}B\triangle T^{-k-1}B)\subset\bigcup_{k=0}^{j-1}T^{-k}(B\triangle T^{-1}B)$ we conclude that $\mu(B\triangle T^{-j}B)=0$. Consequently $\mu(B\triangle C_n)=0$ which implies $\mu(B\triangle C)=0$. Since C is T-invariant we get by ergodicity that $\mu(C)=0,1$ and thus $\mu(B)=0,1$.

"$\Leftarrow$": If C is T-invariant then put $B=C$. Clearly $\mu(B\triangle C)=0$. ∎

(II) μ ergodic $\iff$ every T-invariant function f is a constant (almost everywhere).

Proof. "$\Rightarrow$": Assume μ is ergodic and let f be invariant, i.e. $f\circ T=f$. For any real α put $E_\alpha=\{x\in\Omega: f(x)<\alpha\}$. Clearly E_α is a T-invariant subset of Ω and since μ is ergodic its measure is either 0 or 1. Thus f is constant almost surely.

"$\Leftarrow$": Let $A\in\mathscr{B}$ be T-invariant and χ_A its characteristic function. Then $\chi_A\circ T=\chi_A$. By assumption χ_A is constant, which means χ_A is either 0 or 1 almost surely. Hence μ is ergodic. ∎

(III) μ is ergodic $\iff$ For all $f\in\mathscr{L}^1$, then $\frac{1}{n}f^n(x)\to\mu(f)$ almost surely.

Proof. "$\Rightarrow$": If $f\in\mathscr{L}^1$ then by the Pointwise Ergodic Theorem $\frac{1}{n}f^n\to f^*$ where $f*$ is T-invariant and satisfies $\mu(f^*)=\mu(f)$. It now follows from (II) f^* is a constant and therefore equal to $\mu(f)$.

"$\Leftarrow$": Let $U\subset\Omega$ be a T-invariant set and put $f=\chi_U$. Then $f^n=\chi_U^n=n\chi_U=f$ which implies that $\lim_{n\to\infty}\frac{1}{n}\chi_U^n=\chi_U=\mu(\chi_U)=\mu(U)$ almost surely by the Pointwise ergodic theorem. Thus χ_U is almost surely either equal to 0 or equal to 1. Hence $\mu(U)=0,1$ and since U was an arbitrary invariant set we conclude that μ is ergodic. ∎

(IV) μ ergodic $\iff$ for all $U,V\in\mathscr{B}$ one has $\frac{1}{n}\sum_{j=0}^{n-1}\mu(U\cap T^{-j}V)\to\mu(U)\mu(V)$.

Proof. "$\Rightarrow$": Assume μ is ergodic and let $U,V\in\mathscr{B}$. Put $f=\chi_U$ and use the pointwise ergodic theorem: $\frac{1}{n}\chi_U^n\to\int\chi_U\,d\mu=\mu(U)$ almost surely. Thus $\frac{1}{n}\chi_U^n\chi_V\to\mu(U)\chi_V$. Integration yields (by the Dominated Convergence Theorem)

$$\frac{1}{n}\int\chi_U^n\chi_V\,d\mu\to\int\mu(U)\chi_V\,d\mu=\mu(U)\mu(V).$$

"$\Leftarrow$": Assume that $\frac{1}{n}\sum_{j=0}^{n-1}\mu(U\cap T^{-j}V)\to\mu(U)\mu(V)$ and let U be a T-invariant subset. With $V=U$ we obtain

$$\mu(U)=\frac{1}{n}\sum_{j=0}^{n-1}\mu(U\cap T^{-j}U)\to\mu(U)^2$$

and therefore $\mu(U)=0,1$ which means μ is ergodic. ∎

(V) μ ergodic $\iff$ for all $A\in\mathscr{B}$ with $\mu(A)>0$ one has $\bigcup_{j=0}^\infty T^{-j}A=\Omega$ (up to nullsets).

Proof. "$\Rightarrow$": Assume μ is ergodic, let $A\subset\Omega$, $\mu(A)>0$, and put $U=\bigcup_{j=0}^\infty T^{-j}A$. Clearly $T^{-1}U\subset U$ and $\mu(T^{-1}U)=\mu(U)$ as μ is T-invariant. Hence $T^{-1}U=U$ (up to nullsets) and therefore $\mu(U)=1$ by ergodicity of μ as $\mu(U)\ge\mu(A)>0$.

"$\Leftarrow$": Let $A\subset\Omega$ be T-invariant. If $\mu(A)>0$ then by assumption $U=\bigcup_{j=0}^\infty T^{-j}A$ has full measure. But since $T^{-j}A=A$ we have $U=A$ and therefore $\mu(A)=1$. ∎

(VI) μ ergodic $\iff$ For all $U,V\in\mathscr{B}$ with $\mu(U),\mu(V)>0$ there exists a j so that $\mu(U\cap T^{-j}V)>0$.

Proof. "$\Rightarrow$": If μ is ergodic then by (V) $\bigcup_{j=0}^\infty T^{-j}V=\Omega$ if $\mu(V)>0$. Thus $U\subset\bigcup_{j=0}^\infty T^{-j}V$ which implies that $\mu(U\cap T^{-j}V)>0$ for some j provided $\mu(U)>0$.

"$\Leftarrow$": Let $A \subset \Omega$ be T-invariant, i.e. $T^{-1}A = A$, $T^{-1}A^c = A^c$. With $U = A, V = A^c$ we get by assumption $\mu(U \cap T^{-j}V) = \mu(A \cap T^{-j}A^c) > 0$ for some j. This is impossible if $\mu(A), \mu(A^c) > 0$. ∎

4. Examples

4.1. **Bernoulli shift.** Let $\mathcal{A} = \{1, 2, \dots, M\}$ be an alphabet and put $\Sigma = \mathcal{A}^{\mathbb{N}_0} = \{\vec{x} = (x_0 x_1 x_2 \dots) : x_j \in \mathcal{A}\}$ for the set of infinite sequences composed from the alphabet $\mathcal{A}$. The map $\sigma : \Sigma \to \Sigma$ is given by $(\sigma\vec{x})_j = (\vec{x})_{j+1}$ and called the *shift transformation*. The topology is generated by the following metric: Let $\vartheta \in (0, 1)$ and put $d(\vec{x}, \vec{y}) = \vartheta^{n(\vec{x},\vec{y})}$, where $n(\vec{x}, \vec{y}) = \min\{|j| : x_j \neq y_j\}$. A basis for the topology (and the Borel σ-algebra) consists of *cylinder sets* $U(x_1 x_2 \dots x_n) = \{\vec{y} \in \Sigma : y_1 \dots y_n = x_1 \dots x_n\}$ and their shifts. Notice that balls in Σ are open-closed and that every $\vec{y} \in B_\varepsilon(\vec{x})$ lies in the centre of the ball, i.e. $B_\varepsilon(\vec{y}) = B_\varepsilon(\vec{x})\ \forall\, \vec{y} \in B_\varepsilon(\vec{x})$.

Let $\vec{p} = (p_1, p_2, \dots, p_M)$ be a positive probability vector, i.e. $p_j > 0$ and $\sum_j p_j = 1$. Then we have a probability measure μ on Σ which on cylinder sets is given by

$$\mu(U(x_1 x_2 \cdots x_n)) = p_{x_1} p_{x_2} \cdots p_{x_n}.$$

It is clear that μ is σ-invariant since

$$\sigma^{-1}U(x_0 \dots x_{n-1}) = \bigcup_{a=1}^{M} U(a x_0 \dots x_{n-1})$$

(disjoint union of $(n+1)$-cylinder sets. Hence

$$\mu(\sigma^{-1}U(x_0 \dots x_{n-1})) = \sum_{a=1}^{M} \mu(U(a x_0 \dots x_{n-1})) = \sum_{a=1}^{M} p_a p_{x_0} \cdots p_{x_{n-1}} = \mu(U(x_0 \dots x_{n-1}))$$

as $\sum_a p_a = 1$.

Lemma 11. *The Bernoulli measure μ is ergodic.*

Proof. By (V) it is enough to show that for $V, W \subset \Sigma$ (both with positive measure) one has $\mu(V \cap \sigma^{-j}W) > 0$ for some j. This is shown for generators of the σ-algebra. Assume $V = U(x_1 \dots x_n)$, $W = U(y_1 \dots y_m)$ are cylinder sets. Then

$$V \cap T^{-j}W = \bigcup_{z_1 \cdots z_{j-n}} U(x_1 \cdots x_n z_1 \cdots z_{j-n} y_1 \cdots y_m)$$

is a disjoint union where the union is over all $z_1 \cdots z_{j-n} \in \mathcal{A}^{j-n}$. Thus for $j \geq n$

$$\begin{aligned}
\mu(V \cap T^{-j}W) &= \sum_{z_1 \dots z_{j-n}} p_{x_1} \cdots p_{x_n} p_{z_1} \cdots p_{z_{j-n}} p_{y_1} \cdots p_{y_m} \\
&= p_{x_1} \cdots p_{x_n} p_{y_1} \cdots p_{y_m} \\
&= \mu(V)\mu(W) > 0
\end{aligned}$$

as $\sum_{z_k} p_{z_k} = 1$. Hence μ is ergodic. In fact we have proven a much stronger result here. ∎

4.2. **Irrational rotation.** Let $\Omega = [0,1)$ and $\alpha \in (0,1)$ an irrational number and define the map $T : \Omega \to \Omega$ by $Tx = x + \alpha \mod 1$. Obviously T is invertible. Let λ be the Lebesgue measure on $[0,1)$. Since $\frac{dT}{dx} = 1$ one has that T preserves λ, i.e. $T^*\lambda = \lambda$.

Lemma 12. *λ is ergodic.*

Proof. We use the fact that λ is ergodic iff every T-invariant function $f \in \mathscr{L}^1$ is constant almost everywhere. Since $\mathscr{L}^2$ is dense in $\mathscr{L}^1$ it is sufficient to prove it for $f \in \mathscr{L}^2$. If $f \in \mathcal{L}^1$ is T-invariant then so are the approximating functions f_n which are defined by putting $f_n(x) = f(x)$ if $|f(x)| \le n$ and equal to 0 if $|f(x)| > n$. This follows from the T-invariance of the sets $\{x : |f(x)| \le n\}$ are T-invariant. We write f as a Fourierseries: $f(x) = \sum_{n=-\infty}^{\infty} a_n e^{2\pi i n x}$, where $a_n = \int_0^1 f(t)e^{-2\pi i n t}\,dt$ are the Fourier coefficients. Since

$$f(Tx) = \sum_{n=-\infty}^{\infty} a_n e^{2\pi i n (x+\alpha \bmod 1)} = \sum_{n=-\infty}^{\infty} a_n e^{2\pi i n\alpha} e^{2\pi i n x}$$

we get by uniqueness of the Fourier expansion that $a_n = a_n e^{2\pi i n\alpha}$ for every $n \in \mathbb{Z}$. Since α is irrational, $e^{2\pi i n\alpha} \ne 1$ for all $n \ne 0$ and thus $a_n = 0$ for all $n \ne 0$. Hence $f(x) = a_0$ is a constant and therefore λ is ergodic. ∎

4.3. **Affine expanding maps on the interval.** Let $\Omega = [0,1)$ and $d \ge 2$ an integer (degree). Then we define the map T on Ω by $Tx = dx \mod 1$. Clearly T is not invertible and every point x has d preimages. In fact $T^{-1}x = \{\frac{x}{d} + \frac{j}{d} : j = 0, 1, 2, \dots, d-1\}$. Again let λ be the Lebesgue measure.

Lemma 13. *λ is T-invariant.*

Proof. We show that $\lambda(T^{-1}I) = \lambda(I)$ for intervals $I = [a,b)$, $0 \le a < b < 1$. We have that $T^{-1}I = \dot{\bigcup}_{j=0}^{j-1} I_j$ (disjoint union), where $I_j = [\frac{a}{d} + \frac{j}{d}, \frac{b}{d} + \frac{j+1}{d})$. Thus

$$\lambda(T^{-1}) = \sum_{j=0}^{d-1} \lambda(I_j) = \sum_{j=0}^{d-1} \left(\left(\frac{b}{d} + \frac{j+1}{d} \right) - \left(\frac{a}{d} + \frac{j}{d} \right) \right) = b - a = \lambda(I).$$

Since this is true for any $0 \le a < b < 1$ one sees that λ is T-invariant. ∎

Lemma 14. *λ is ergodic*

Proof. As above we show that any T-invariant $f \in \mathscr{L}^2$ must be a constant. We use the Fourier expansion $f(x) = \sum_{n=-\infty}^{\infty} a_n e^{2\pi i n x}$ where the coefficients a_n satisfy Parseval's identity $\sum_n |a_n|^2 = \int_0^1 |f(t)|^2\,dt < \infty$. We have

$$f(Tx) = \sum_{n=-\infty}^{\infty} a_n e^{2\pi i n (dx \bmod 1)} = \sum_{n=-\infty}^{\infty} a_n e^{2\pi i n d x}$$

and comparing coefficients with $f(x) = \sum_{n=-\infty}^{\infty} a_n e^{2\pi i n x}$ we see that $a_n = a_{dn}$. Iterating yields $a_n = a_{dn} = a_{d^2 n} = a_{d^3 n} = \cdots$. By Parseval's identity we conclude that $a_n = 0$ for all $n \ne 0$ and hence $f(x) = a_0$ is a constant. Since $\mathscr{L}^2$ is dense in $\mathscr{L}^1$ we obtain the same result for $f \in \mathscr{L}^1$ and therefore λ is ergodic. ∎

We use this result to show a well-known result on the distribution of digits in a base d expansion of real numbers ($d \geq 2$). For $x \in [0,1)$ let

$$x = 0.a_1a_2a_3\cdots = \sum_{j=1}^{\infty} \frac{a_j}{d^j}$$

be its base d expansion. We are interested in the distribution of the digits $a_j \in \{0,1,2,\dots,d-1\}$.

Theorem 15. *(Borel's law on the normality of numbers) For λ-almost every $x \in [0,1)$ the density*

$$D_k(x) = \lim_{n\to\infty} \frac{|\{j : 0 \leq j < n, a_j = k\}|}{n}$$

exists and equals $\frac{1}{d}$ for all $k \in \{0,1,\dots,d-1\}$.

Proof. Let $k \in \{0,1,\dots,d-1\}$ and put

$$\chi_k(x) = \begin{cases} 1 & \text{if} \qquad x \in \left[\frac{k}{d}, \frac{k+1}{d}\right) \\ 0 & \text{otherwise} \end{cases} .$$

One has $\chi_k(x) = 1$ if $a_1 = k$ and 0 otherwise and similarly $\chi_k(T_jx) = 1$ iff $a_j = 1$, where T is the affine stretching map from the previous example and a_j are the digits in the base d expansion of x. Thus

$$|\{j : 0 \leq j < n, a_j = k\}| = \sum_{j=0}^{n-1} \chi_k(T^jx)$$

and the limit $D_k(x) = \lim_{n\to\infty} \frac{1}{n}\chi_k^n(x)$ exists almost surely by the pointwise ergodic theorem and equals $\int_0^1 \chi_k(x)\, d\lambda(x) = \frac{1}{d}$. ∎

4.4. **Gauss map.** Again we use the unit interval and put $\Omega = (0,1]$. The *Gauss map* T on Ω is defined by $Tx = \frac{1}{x} \bmod 1$ and is related to the continued fraction expansion of real numbers. Any $x \in \Omega$ can be written as a continued fraction expansion

$$x = \cfrac{1}{a_1 + \frac{1}{a_2 + \frac{1}{a_3 + \cdots}}}$$

where the integers $a_j \in \mathbb{N}$ are uniquely determined by x. We also write $x = [a_1, a_2, a_3, \dots]$ and note that the sequence is finite if and only if x is a rational number. The rational numbers

$$\frac{p_n}{q_n} = [a_1, a_2, \dots, a_n]$$

which are the truncated continued fraction expansions are the *approximants* of x and satisfy the recursion formulas

$$\begin{aligned} p_n &= a_n p_{n-1} + p_{n-2} \\ q_n &= a_n q_{n-1} + q_{n-2} \end{aligned}$$

with the initial values $p_0 = 0, p_1 = 1, q_0 = 1, q_1 = a_1$. The golden mean has the continued fraction expansion $\frac{1}{2}(\sqrt{5} - 1) = [1,1,1,\dots]$ and the sequence $\{p_n\}$ are the Fibonacci numbers.

If we denote by $n(x)$ for $x \in \Omega$ the function with values in $\mathbb{N}$ that is given by $Tx = \frac{1}{x} - n(x)$ then we can solve for x and obtain $x = \frac{1}{Tx+n(x)}$. Iteration yields

$$x = \frac{1}{n(x) + \frac{1}{n(Tx)+\frac{1}{n(T^2x)+\frac{1}{n(T^3x)+\dots}}}}$$

which implies that $a_j = n(T^{j-1}x)$. The *Gauss measure* μ on $(0,1]$ is the probability measure μ which has the density $\frac{1}{\log 2}\frac{1}{1+x}$, i.e.

$$\mu([a,b]) = \frac{1}{\log 2}\int_a^b \frac{dx}{1+x} = \frac{1}{\log 2}\log\frac{1+b}{1+a}$$

for $0 < a < b \le 1$.

Lemma 16. *The Gauss measure is invariant under the Gauss map.*

Proof. Clearly T is not invertible, in fact

$$T^{-1}x = \left\{\frac{1}{x+j} : j = 1, 2, \dots\right\}$$

since $y \in T^{-1}x$ means that $x = Ty = \frac{1}{y} - j$ for some $j \in \mathbb{N}$ and therefore $y = \frac{1}{x+j}$. Hence, if $I = (a,b]$, $0 \le a < b \le 1$, is an interval, then $T^{-1}I = \dot{\bigcup}_{j=0}^{\infty} I_j$, where $I_j = \left[\frac{1}{b+j}, \frac{1}{a+j}\right)$ (notice that T reverses orientation). Therefore

$$\begin{aligned}
\mu(T^{-1}I) &= \sum_{j=1}^{\infty}\mu(I_j)\\
&= \sum_{j=1}^{\infty}\frac{1}{\log 2}\log\frac{1+\frac{1}{a+j}}{1+\frac{1}{b+j}}\\
&= \sum_{j=1}^{\infty}\frac{1}{\log 2}\log\frac{a+j+1}{b+j+1}\frac{b+j}{a+j}\\
&= \frac{1}{\log 2}\sum_{j=1}^{\infty}\left(\log\frac{a+j+1}{b+j+1} - \log\frac{a+j}{b+j}\right)\\
&= \frac{1}{\log 2}\log\frac{b+1}{a+1} = \mu(I).
\end{aligned}$$

∎

Lemma 17. *The Gauss measure μ is ergodic.*

Proof. We use that ergodicity is equivalent to the fact when for every $U, V \subset \Omega$ of positive measures one has $\mu(U \cap T^{-j}V) > 0$ for some $j \in \mathbb{N}$. It is enough to consider the case when U, V are intervals since the Borel σ-algebra is generated by intervals.

Denote by ψ_j the inverse branches of T, that is $\psi_j(x) = \frac{1}{x+j}$ where we also put $\Delta_j = \psi_j(\Omega) = \left[\frac{1}{j+1}, \frac{1}{j}\right)$ for its range. Iterating the inverse maps yields inverse branches of higher powers of T. We denote $\psi_{a_1a_2\ldots a_n} = \psi_{a_n} \circ \cdots \circ \psi_{a_2} \circ \psi_{a_1}$ for the inverse branches of T^n where $a_j \in \mathbb{N}$. Its range is $\Delta_{a_1a_2\ldots a_n} = \psi_{a_1a_2\ldots a_n}(\Omega)$ and $(0,1] = \bigcup_{a_1a_2\ldots a_n} \Delta_{a_1a_2\ldots a_n}$ is a disjoint union. Since T is an expanding map we conclude that

$$D_n = \sup_{(a_1,a_2,\ldots,a_n)\in\mathbb{N}^n} \operatorname{diam}(\Delta_{a_1a_2\ldots a_n}) \to 0.$$

Assume U, V are intervals and let n be large enough so that $D_n < \frac{1}{2}\min\{\operatorname{diam}(U), \operatorname{diam}(V)\}$. Then there exist n-words $b_1b_2\cdots b_n, a_1a_2\cdots a_n \in \mathbb{N}^n$ so that $\Delta_{b_1b_2\ldots b_n} \subset V$ and $\Delta_{a_1a_2\ldots a_n} \subset U$. Since $\psi'_{b_1b_2\cdots b_n} > 0$ we conclude that $\mu(\psi_{b_1b_2\cdots b_n}(V)) > 0$ ($\mu(V) > 0$ by assumption) and thus since

$$\Delta_{a_1a_2\ldots a_n} = \Delta_{a_1a_2\ldots a_n} \cap \psi_{b_1b_2\cdots b_n}(\Delta_{b_1\ldots b_n}) \subset U \cap \psi_{b_1b_2\cdots b_n}(V) \subset U \cap T^{-n}V$$

we obtain $\mu(U \cap T^{-n}V) > 0$. This proves that μ is ergodic. ∎

Theorem 18. *Let $x \in (0,1]$ and $[a_1(x), a_2(x), a_3(x), \ldots]$ its continued fraction expansion. Then for every $k \in \mathbb{N}$ the limit*

$$D_k(x) = \lim_{n\to\infty} \frac{|\{j : 1 \le j \le n, a_j(x) = k\}|}{n} = \frac{1}{\log 2} \log \frac{(1+k)^2}{k(k+2)}$$

exists μ-almost everywhere.

Proof. We use the BET and ergodicity of μ. If we again denote by $n(x)$ the integer part of $\frac{1}{x}$, i.e. $Tx = \frac{1}{x} - n(x)$, then for $x \in [\frac{1}{k+1}, \frac{1}{k})$ one has $n(x) = a_1(x) = k$. Similarly $a_j(x) = n(T^{j-1}x)$. We now put

$$\chi_k(x) = \begin{cases} 1 & \text{if} \qquad x \in [\frac{1}{k+1}, \frac{1}{k}) \\ 0 & \text{otherwise} \end{cases}$$

and get

$$|\{j : 1 \le j \le n, a_j(x) = k\}| = \sum_{j=1}^{n} \chi_k(T^{j-1}x) = \chi_k^n(x).$$

By the pointwise ergodic theorem we get for the limit:

$$\begin{aligned} D_k(x) &= \lim_{n\to\infty} \frac{1}{n}\chi_k^n(x) \\ &= \int_0^1 \chi_k(x)\, d\mu(x) \\ &= \frac{1}{\log 2} \int_{\frac{1}{k+1}}^{\frac{1}{k}} \frac{dx}{1+x} \\ &= \frac{1}{\log 2} \log \frac{1+\frac{1}{k}}{1+\frac{1}{k+1}} \\ &= \frac{1}{\log 2} \log \frac{(k+1)^2}{k(k+2)}. \end{aligned}$$

The asymptotic is $D_k(x) \sim \frac{1}{\log 2}\frac{1}{k^2+2k}$. ∎

Lemma 19. *The limit*

$$\lim_{n\to\infty}(a_1(x)a_2(x)\cdots a_n(x))^{\frac{1}{n}}=\prod_{j=1}^\infty\left(\frac{(j+1)^2}{j(j+2)}\right)^{\frac{\log j}{\log 2}}$$

exists for Lebesgue almost every x.

Proof. Define the function $f:(0,1)\to\mathbb{R}^+$ by putting $f(x)=\log j$ for $x\in\left(\frac{1}{j+1},\frac{1}{j}\right)$, $j=1,2,\dots$. Clearly $f\in\mathscr{L}^1(\mu)$ because $\int_0^1 f(x)\,dx=\sum_j\left(\frac{1}{j}-\frac{1}{j+1}\right)\log j\le\sum_j\frac{\log j}{j^2}<\infty$. Then by the pointwise ergodic theorem

$$\frac{1}{n}\sum_{j=1}^n\log a_j(x)=\frac{1}{n}\sum_{j=1}^n f(T^{j-1}x)=\frac{1}{n}f^n(x)\longrightarrow\mu(f)$$

as $n\to\infty$ since μ is ergodic. For the value of the integral on the RHS we obtain

$$\begin{aligned}\mu(f)&=\sum_j\mu((\frac{1}{j+1},\frac{1}{j}))\\&=\sum_j\frac{\log j}{\log 2}\int_{\frac{1}{j+1}}^{\frac{1}{j}}\frac{dx}{1+x}\\&=\sum_j\frac{\log j}{\log 2}\log\frac{1+\frac{1}{j}}{1+\frac{1}{j+1}}\\&=\sum_j\frac{\log j}{\log 2}\frac{(j+1)^2}{j(j+2)}.\end{aligned}$$

Exponentiation yields the statement in the lemma. ∎

4.5. **Subshift of finite type (SFT).** Let $\mathcal{A}=\{1,2,\dots,M\}$ be a finite alphabet and A a $0,1$-valued $M\times M$-matrix. Then we put

$$\Sigma=\left\{\vec{x}\in\mathcal{A}^{\mathbb{Z}}:A_{x_i,x_{i+1}}=1\;\forall\;i\right\}$$

The map is the left shift $\sigma:\Sigma\to\Sigma$ as above. The metric is the same as above for the Bernoulli shift, that is $d(\vec{x},\vec{y})=\vartheta^{n(\vec{x},\vec{y})}$ $(\vartheta<1)$ where $n(\vec{x},\vec{y})$ is the smallest $|j|$ for which $x_j\ne y_j$.

For instance the two element alphabet $\mathcal{A}=\{0,1\}$ and the transition matrix $A=\begin{pmatrix}1&1\\1&0\end{pmatrix}$ give rise to the subshift Σ that consists of all doubly infinite sequences composed of 0s and 1s so that never any two 1s are consecutive since the transition matrix A allows for the 2-words $00,01,10$ but disallows 11.

Let P be a stochastic matrix, that is $P\mathbb{1}=\mathbb{1}$ $(\mathbb{1}=(1,1,\dots,1))$ and $\vec{p}P=\vec{p}$ for a left eigenvector $\vec{p}=(p_1,\dots,p_M)$ which is positive and satisfies $\sum_j p_j=1$. This defines a σ-invariant measure on Σ by putting on cylindersets

$$\mu(U(x_1x_2\dots x_n))=p_{x_1}P_{x_1x_2}P_{x_2x_3}\cdots P_{x_{n-1}x_n}.$$

Lemma 20. *The limit* $Q = \lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} P^k$ *exists and moreover* Q *is a stochastic matrix. Also* $PQ = QP = Q = Q^2$.

Proof. If we put

$$\chi_i(x) = \begin{cases} 1 & \text{if} \qquad x \in U(i) \\ 0 & \text{otherwise} \end{cases}$$

then $\mu(\chi_i) = \mu(U(i)) = p_i$. By the pointwise ergodic theorem $\lim_{n\to\infty} \frac{1}{n}\chi_i^n(x) = \tilde{\chi}_i(x)$ μ-almost surely where $\tilde{\chi}_i$ is σ-invariant and satisfies $\mu(\tilde{\chi}_i) = \mu(\chi_i)$. Integrating $\lim_n \frac{1}{n}\chi_i^n\chi_j = \tilde{\chi}_i\chi_j$ yields

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1}\int_\Sigma \chi_i \circ T^k \ \chi_j \, d\mu = \int_\Sigma \tilde{\chi}_i\chi_j \, d\mu.$$

Since $\int_\Sigma \chi_i \circ T^k \ \chi_j \, d\mu = \mu(U(j) \cap T^{-k}U(i)) = p_j P_{ji}^k$ we obtain $\lim_{n\to\infty} \frac{1}{n} p_j \sum_{k=0}^{n-1} P_{ji}^k = \mu(\tilde{\chi}_i\chi_j)$ Put

$$Q_{ji} = \frac{1}{p_j}\mu(\tilde{\chi}_i\chi_j) = \lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1} P_{ji}^k.$$

This proves the first claim of the lemma. It is clear that Q is stochastic matrix with left eigenvector $\vec{p}$. It also follows easily that $PQ = QP = Q = Q^2$. ∎

Lemma 21. *Let* P *be irreducible and aperiodic (i.e.* $P^n > 0$ *for all* n *large enough). Then all rows of* Q *are identical. In fact* $Q_{ij} = p_j$.

Proof. Let n be large enough so that P^n is positive. Since $QP = Q$ by the last lemma, we also have $QP^n = Q$ which means that $Q_{ij} = \sum_{k=1}^M Q_{ik}(P^n)_{kj}$. Since Q is stochastic, $\sum_k Q_{ik} = 1$ and therefore at least one of the $Q_{ik}, k = 1, \ldots, M$, is non-zero. Since $P^n > 0$ we conclude that $Q_{ij} > 0$ for all i, j. Thus $Q > 0$. To see that all rows are equal let us put $q_j = \max_i Q_{ij}$ be the maximum in the j-th column. Assume there is a $Q_{ij} < q_j$, then we get as $Q = Q^2$:

$$Q_{ij} = \sum_k Q_{ik}Q_{kj} < q_j \sum_k Q_{kj} = q_j$$

since one of the Q_{ik} is less than q_j and the weights Q_{kj}, k add up to 1 (Q_{ij} is a convex combination of $\{Q_{kj} : k\}$). Since i is arbitrary we get that $Q_{ij} < q_j$ for all i which contradicts the definition of q_j. Hence $Q_{ij} = q_j$ for all i, j.

To get the values of the entries of Q we use that $\vec{p}$ is a left eigenvector of Q to the eigenvalue 1. To wit $\vec{p}Q = \vec{p}$ and therefore for every j:

$$p_j = \sum_i p_i Q_{ij} = \sum_i p_i q_j = q_j$$

as $\sum_i p_i = 1$. Thus $Q_{ij} = p_j$ for all i, j. ∎

Lemma 22. μ *is ergodic.*

Proof. We will show that for cylindersets V, W one has

$$\frac{1}{N}\sum_{j=0}^{N-1} \mu(W \cap \sigma^{-j}V) \to \mu(W)\mu(V)$$

as N goes to infinity. Let $W=U(x_1\cdots x_n), V=U(y_1\cdots y_m)$, then if $j\geq n$ one has

$$W\cap\sigma^{-j}V=\bigcup_{z_1\cdots z_{j-n}}U(x_1\cdots x_nz_1\cdots z_{j-n}y_1\cdots y_m)$$

where the union is over all words $z_1\cdots z_{j-n}$ in Σ of length $j-n$ which allow for the transitions $A_{x_nz_1}=A_{z_{j-n}y_m}=1$. Thus

$$\begin{aligned}\mu(W\cap\sigma^{-j}V) &= \sum_{z_1\cdots z_{j-n}}U(x_1\cdots x_nz_1\cdots z_{j-n}y_1\cdots y_m)\\ &= p_{x_1}P_{x_1x_2}\cdots P_{x_{n-1}x_n}(P^{j-n})_{x_ny_1}P_{y_1y_2}\cdots P_{y_{m-1}y_m}\end{aligned}$$

and consequently

$$\frac{1}{N}\sum_{j=n}^{N-1}\mu(W\cap\sigma^{-j}V)=p_{x_1}P_{x_1x_2}\cdots P_{x_{n-1}x_n}\left(\frac{1}{N}\sum_{j=n}^{N-1}P^{j-n}\right)_{x_ny_1}P_{y_1y_2}\cdots P_{y_{m-1}y_m}$$

which converges to $\mu(W)\mu(V)$ as $N\to\infty$ since the entry in brackets converges by the previous two lemmas to p_{y_1} and in the averaging n terms will not affect the limit. ∎

4.6. **Linear toral automorphisms.** Here we put $\Omega=\mathbb{T}^n=\mathbb{R}^n/\mathbb{Z}^n$. Let A be an $n\times n$-matrix with integer entries. Then A induces a continuous map on Ω which is invertible if $\det A=\pm1$.

For instance the matrix $A=\begin{pmatrix}2&1\\1&1\end{pmatrix}$ induces a homeomorphism on $\mathbb{T}^2$. A has eigenvalues $\lambda_\pm=\frac{1}{2}(3\pm\sqrt{5})$, where $0<\lambda_-<1<\lambda_+$. The eigenvectors $\vec{v}_\pm=\begin{pmatrix}1\\\frac{1\pm\sqrt{5}}{2}\end{pmatrix}$ span the expanding manifold and contracting manifolds at the fixed point 0:

$$\begin{aligned}W^u(0) &= \{x\in\mathbb{T}^2:|T^jx|\to0 \text{ for } j\to\infty\}\\ W^s(0) &= \{x\in\mathbb{T}^2:|T^{-j}x|\to0 \text{ for } j\to\infty\}\end{aligned}$$

There exists a finite partition $\mathscr{P}=\{R_1,R_2,\dots,R_M\}$ of Ω whose elements have boundaries that consist of unions of pieces from W^s,W^u. To be more precise $\mathscr{P}$ satisfies the *Markov condition* which is given by:

$$\begin{aligned}T(\partial R_j\cap W^s) &\subset \partial R\cap W^s\\ T(\partial R_j\cap W^u) &\supset \partial R\cap W^u\end{aligned}$$

Then one can define a transition matrix on the alphabet $\mathcal{A}=\{1,\dots,M\}$ by putting

$$A_{ij}=\begin{cases}1 & \text{if } \ \mathrm{int}TR_i\cap\mathrm{int}R_j\neq\emptyset\\ 0 & \text{otherwise}\end{cases}.$$

The shiftspace Σ is then defined as above and has the left shift map $\sigma:\Sigma\to\Sigma$. Then T and σ are semiconjugate, i.e. $\pi\circ\sigma=T\circ\pi$ where $\pi:\Sigma\to\mathbb{T}^2$ is the map defined by

$$\pi(\vec{x})=\bigcap_{j=-\infty}^{+\infty}T^{-j}R_{x_j}$$

for all $\vec{x}\in\Sigma$.

5. Return times and the induced map

5.1. **Kac's theorem and the induced map.** For $U \subset \Omega$ we define the return time $\tau_U(x) = \min\{j \geq 1 : T^j x \in U\}$. By the Poincaré recurrence theorem $\tau_U(x) < \infty$ for almost every $x \in U$ for any finite T-invariant measure μ on Ω. Poincaré's theorem doesn't tell us anything about how big τ_U is. The next result gives us the expected value of τ_U on U which in particular implies that τ_U is integrable on U (assuming $\mu(U) > 0$).

Theorem 23. *(Kac 1947) If μ is an ergodic T-invariant probability measure on Ω then for any $U \subset \Omega$ of positive measure one has*

$$\int_U \tau_U(x)\, d\mu(x) = 1.$$

Proof. Let us put τ_U^k for the kth return time, that is we put $\tau_U^1 = \tau_U$ and then define recursively

$$\tau_U^k(x) = \tau_U(\hat{T}^{k-1}x) + \tau_U^{k-1}(x)$$

where we put $\hat{T}(x) = T^{\tau_U(x)}(x)$ for the *induced transformation* on U (which exists almost surely by Poincaré). Inductively we also get

$$\tau_U^k = \tau_U + \tau_U \circ \hat{T} + \tau_U \circ \hat{T}^2 + \cdots + \tau_U \circ \hat{T}^{k-1},$$

i.e. the kth return time is the kth ergodic sum of τ_U on $(U, \hat{T})$ By the pointwise ergodic theorem we get

$$\int_U \tau_U\, d\mu = \int_\Omega \chi_U \tau_U\, d\mu = \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} (\chi_U \tau_U)(T^j x)$$

for μ-almost every $x \in \Omega$ as μ is ergodic. If we take the limit along a subsequence $n_\ell = \tau_U^\ell(x)$ and use the fact that

$$(\chi_U \tau_U)(T^j x) = \begin{cases} 0 & \text{if } \ T^j x \notin U \\ \tau_U(T^j x) & \text{if } \ T^j x \in U \end{cases}$$

then we get (with $n = \tau_U^\ell$)

$$\int_U \tau_U\, d\mu = \lim_{\ell\to\infty} \frac{1}{\tau_U^\ell(x)} \sum_{j=0}^{\tau_U^\ell - 1} (\chi_U \tau_U)(T^j x) = \lim_{\ell\to\infty} \frac{1}{\tau_U^\ell(x)} \sum_{j=0}^{\ell-1} \tau_U(\hat{T}^j x) = \lim_{\ell\to\infty} \frac{1}{\tau_U^\ell(x)} \tau_U^\ell(x) = 1.$$

It remains to show that $\chi_U \tau_U \in \mathscr{L}^1$. We use the same argument again but this time cut off the values of τ_U. For R large we put

$$\varphi_R(x) = \begin{cases} 1 & \text{if } \ \tau_U(x) \leq R \\ 0 & \text{if } \ \tau_U(x) > R \end{cases}.$$

Now, since $\varphi_R \chi_U \tau_U \in \mathscr{L}^1$, we get by the pointwise ergodic theorem

$$\int_U \varphi_R \tau_U\, d\mu = \int_\Omega \varphi_R \chi_U \tau_U\, d\mu = \lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n-1} (\varphi_R \chi_U \tau_U)(T^j x) = \lim_{n\to\infty} \frac{1}{\tau_U^\ell(x)} \sum_{j=0}^{\tau_U^\ell - 1} (\varphi_R \tau_U)(\hat{T}^\ell x) \leq 1$$

for all values of R. Let $R \to \infty$ which implies that $\chi_U \tau_U \in \mathscr{L}^1$. ∎

Second proof of Kac's theorem if T is invertible. This one uses a representation of Ω which is called a *Rokhlin tower*. Given $U \subset \Omega$ ($\mu(U) > 0$), then we put $U_k = \{x \in U : \tau_U(x) = k\}$, $k = 1, 2, \dots$, for the level sets of τ_U. Then $U = \dot{\bigcup}_{k=1}^{\infty} U_k$ is a disjoint union and the sets $T^j U_k$ for $j = 0, 1, \dots, k-1, k = 1, 2, \dots$, are pairwise disjoint. Since μ is ergodic, $\Omega = \bigcup_k \bigcup_{j=1}^{k-1} T^j U_k$ and as T is invertible $\mu(T^j U_k) = \mu(U_k)$. Hence

$$1 = \mu(\Omega) = \sum_{k=1}^{\infty} \sum_{j=0}^{k-1} \mu(T^{-1} U_k) = \sum_k k\mu(U_k) = \int_U \tau_U \, d\mu$$

This uses the representation of Ω by the following tower construction which is due to Rokhlin. Let F be a map on a space Δ_0 and assume Δ_0 is decomposed into a disjoint union $\Delta_0 = \bigcup_k \Delta_{k,0}$. Then, given a (roof) function $r : \mathbb{N} \to \mathbb{N}$, we put

$$\Delta = \bigcup_{k=1}^{\infty} \bigcup_{j=0}^{r(k)-1} \Delta_{k,j}$$

(disjoint union), where $\Delta_{k,j} = \{(x, j) : x \in U_k\}$. Then there is a map S on Δ which is defined by

$$\begin{aligned} S &: \Delta_{k,j} \to \Delta_{k,j+1} \text{ if } j \leq r(k) - 1 \\ S &: \Delta_{k,r(k)-1} \to \Delta_0 = \bigcup_k \Delta_{k,0} \end{aligned}$$

where $S(x, j) = (x, j + 1)$ for $(x, j) \in \Delta_{k,j}$ and if $j < f(k) - 1$. If $(x, j) \in \Delta_{k,r(k)-1}$ then the map is $S(x, j) = (F(x), 0)$. We call the pair (S, Δ) a *Rokhlin tower*. In the case of Kac's theorem $\Delta_0 = U$ and the roof function is $f(k) = k$.

For a subset $U \subset \Omega$, $\mu(U) > 0$, let us denote by $\hat{T} = T^{\tau_U} : U \to U$ the *induced map*. $\hat{T}$ exists by Poincaré's (or Kac's) theorem almost everywhere. We also have the *induced measure* $\hat{\mu}$ which is defined on U by $\hat{\mu}(A) = \frac{\mu(A)}{\mu(U)}$ for all measurable $A \subset U$.

Lemma 24. *The induced measure $\hat{\mu}$ is $\hat{T}$-invariant.*

Proof. Let $A \subset U$ and decompose its pullback $\hat{T}^{-1} A$ as follows:

$$\hat{T}^{-1} A = \dot{\bigcup}_{k=1}^{\infty} (T^{-k} A \cap U_k)$$

(disjoint union), where $U_k = \{x \in U : \tau_U(x) = k\}$ and

$$T^{-k} A \cap U_k = U \cap \left(T^{-k} A \setminus \bigcup_{j=1}^{k-1} T^{-j} U \right).$$

If we put $A_k = T^{-k} A \setminus \bigcup_{j=0}^{k-1} T^{-j} U$ then $A_0 = A$ and $A_k \cap U = \emptyset$ for $k \geq 1$. We have

$$T^{-k} A \cap U_k = U \cap T^{-1} A_{k-1}$$

and hence

$$\mu(\hat{T}^{-1} A) = \sum_{k=1}^{\infty} \mu(T^{-1} A_{k-1} \cap U).$$

Now note that $T^{-1}A_k = (T^{-1}A_k \cap U) \cup A_{k+1}$ (disjoint union as $A_{k+1} = T^{-1}A_k \cap U^c$) and thus

$$\mu(T^{-1}A_k \cap U) = \mu(A_k) - \mu(A_{k+1}).$$

As $\mu(A_0) = \mu(A)$ we get

$$\mu(\hat{T}^{-1}A) = \sum_{k=1}^{\infty}(\mu(A_k) - \mu(A_{k+1}))$$

which equals $\mu(A)$ if $\mu(A_{k+1}) \to 0$ (which is obvious if μ is ergodic as then $\Omega = \bigcup_k T^{-k}A$ if $\mu(A) > 0$). We get that $\mu(\hat{T}^{-1}A) \le \mu(A)$. The same argument applied to $U \setminus A$ (instead of A) yields $\mu(\hat{T}^{-1}(U \setminus A)) \le \mu(U \setminus A) = \mu(U) - \mu(A)$. Since $\hat{T}^{-1}A \cup \hat{T}^{-1}(U \setminus A) = U$ we get $\hat{\mu}(\hat{T}^{-1}A) = \hat{\mu}(A)$. ∎

Lemma 25. *Let $U \subset \Omega$, $\mu(U) > 0$, $\hat{T} = T^{\tau_U}$ the induced map and $\hat{\mu} = \mu|_U$ the restricted probability measure.*
(I) If μ is ergodic with respect to T, then $\hat{\mu}$ is ergodic with respect to $\hat{T}$.
(II) If $\hat{\mu}$ is ergodic w.r.t. $\hat{T}$ and $\Omega = \bigcup_{j=0}^{\infty} T^{-j}U$, then μ is ergodic w.r.t. T.

Proof. This is the first assignment. ∎

Lemma 26. *Assume μ is ergodic and let $U \subset \Omega$ have positive measure.*
Then for all $A \subset U^c$":

$$\mu(A) = \sum_{k=1}^{\infty} \mu\left(U \cap \left(T^{-k}A \setminus \bigcup_{j=1}^{k-1} T^{-j}U\right)\right).$$

Proof. Let $A \subset U^c$ and put

$$A_k = T^{-k}A \setminus \bigcup_{j=0}^{k-1} T^{-j}U = \left\{x \in \Omega : T^k \in A, T^j x \notin U \forall j = 0, 1, \dots, k\right\}$$

where $A_0 = A$. Then

$$U \cap \left(T^{-k}A \setminus \bigcup_{j=1}^{k-1} T^{-j}U\right) = T^{-1}A_{k-1} \setminus A_k$$

for $k = 1, 2, \dots$. Note that $A_k \subset A_{k-1} \forall k$. Hence

$$\begin{aligned} \sum_{k=1}^{n} \mu\left(U \cap \left(T^{-k}A \setminus \bigcup_{j=1}^{k-1} T^{-j}U\right)\right) &= \sum_{k=1}^{n}\big(\mu(T^{-1}A_{k-1}) - \mu(A_k)\big) \\ &= \mu(A_0) - \mu(A_n) \\ &= \mu(A) - \mu(A_n). \end{aligned}$$

Since $A_k \subset \Omega \setminus \bigcup_{j=0}^{k} T^{-j}U$ and by ergodicity $\mu\left(\bigcup_{j=0}^{\infty} T^{-j}U\right) = 1$ we conclude that

$$\mu(A_k) \le \mu\left(\bigcup_{j=0}^{k} T^{-j}U\right) \to 0$$

as $k \to \infty$. ∎

Theorem 27. *Assume μ is ergodic and $U \subset \Omega$ has positive measure. Then for all $A \subset \Omega$ one has*

$$\mu(A) = \int_U \sum_{k=0}^{\tau_U(x)-1} \chi_A \circ T^k(x)\, d\mu(x).$$

Proof. As before put $U_j = \{x : \tau_U(x) = j\}$. Then by the previous lemma

$$\begin{aligned}
\mu(A) &= \mu(A \cap U) + \mu(A \setminus U) \\
&= \mu(A \cap U) + \sum_{k=1}^{\infty} \mu\left(U \cap \left(T^{-k}A \setminus \bigcup_{j=1}^{k-1} T^{-j}U\right)\right) \\
&= \mu(A \cap U) + \sum_{k=1}^{\infty} \mu\left(T^{-k}A \cap \bigcup_{j=k}^{\infty} U_j\right) \\
&= \mu(A \cap U) + \sum_{k=1}^{\infty}\sum_{j=k}^{\infty} \mu\big(T^{-k}A \cap U_j\big) \\
&= \mu(A \cap U) + \sum_{j=1}^{\infty}\sum_{k=1}^{j} \mu\big(T^{-k}A \cap U_j\big) \\
&= \sum_{j=1}^{\infty}\sum_{k=0}^{j} \mu\big(T^{-k}A \cap U_j\big) \\
&= \sum_{j=1}^{\infty}\sum_{k=0}^{j} \int_{U_j} \chi_A \circ T^k\, d\mu \\
&= \int_U \sum_{k=0}^{\tau_U(x)-1} \chi_A \circ T^k(x)\, d\mu(x).
\end{aligned}$$

∎

This provides us with an alternate proof of Kac's theorem. Setting $A = \Omega$ we get that

$$1 = \mu(\Omega) = \int_U \sum_{k=0}^{\tau_U(x)-1} \chi_U \circ T^k(x)\, d\mu(x) = \int_U \tau_U(x)\, d\mu(x).$$

5.2. **Example.** Kac's theorem states that the return time function τ_U is integrable over U and also gives the value of the integral. Here we give an example of a system for which τ_U is not integrable over the entire space Ω and yet the measure is ergodic under the shift map.

Let $\Omega = \mathbb{N}^{\mathbb{Z}}$ where on the state space $\mathbb{N}$ we give the transition probabilities: Let $p_i \in (0,1), i = 1, 2, \ldots,$ be a sequence, then we allow for the transition $i \to i+1$ with probability p_i and for the transition $i \to 1$ with probability $q_i = 1 - p_i$. In other words, we can define a stochastic matrix M by

$$\left\{\begin{array}{rcl} M_{j,1} &=& q_j \\ M_{j,j+1} &=& p_j \\ M_{j,k} &=& 0 \text{ otherwise, i.e. if } k \neq 1 \text{ or } k \neq j+1 \end{array}\right. ,$$

where the transition probability of the transition $j \to k$ is given by the entry $M_{j,k}$. Then $M\mathbb{1} = \mathbb{1}$ as $\sum_{k=1}^{\infty} M_{j,k} = M_{j,1} + M_{j,j+1} = q_j + p_j = 1 \forall j$ and M has the left eigenvector $\vec{x} = (x_1, x_2, \dots)$ (for the dominant eigenvalue 1) which satisfies

$$\begin{array}{rcl} q_1 x_1 + q_2 x_2 + q_3 x_3 + \cdots &=& x_1 \\ x_j p_j &=& x_{j+1} \quad \text{for } j = 1, 2, \dots. \end{array}$$

One sees that the components of the left eigenvector are $x_j = x_1 P_j, j = 2, 3, \dots$, where $P_j = \prod_{i=1}^{j-1} p_i$ ($P_1 = 1$) and x_1 is chosen to make $\vec{x}$ a probability vector ($x_1^{-1} = \sum_j P_j$). The first equation above is satisfied as $\sum_j q_j x_j = \sum_j (1 - p_j) x_1 P_j = x_1 \sum_j (P_j - P_{j+1}) = x_1 P_1 = x_1$ if $P_j \to 0$ as $j \to \infty$. In this way we obtain a shift invariant probability measure μ on Ω which is ergodic as one can go from any state i to any other state j with positive probability.

Put $A_j = \{\vec{\omega} \in \Omega : \omega_0 = j\}, j = 1, 2, \dots$, let $U = A_1$ be the return set and τ_U its return/entry time function. If we put $A_{j,k} = A_j \cap \{\tau_U = k\}$ then $\vec{\omega} \in A_{j,k}$ is of the form $\omega_0 \omega_1 \cdots \omega_k = j(j+1)(j+2) \cdots (j+k-2)(j+k-1)1$ (symbol sequence of length $k+1$). One has

$$\mu(A_{j,k}) = \mu(A_j) p_j p_{j+1} \cdots p_{j+k-2} q_{j+k-1} = x_1 P_j \frac{P_{j+k-1}}{P_j} q_{j+k-1} = x_1 P_{j+k-1} q_{j+k-1}$$

as $\mu(A_j) = x_j = x_1 P_j$. The integral of τ_U over the entire space is

$$\int_\Omega \tau_U \, d\mu = \sum_{j,k} k \mu(A_{j,k}) = \sum_{j,k} k x_1 P_{j+k-1} q_{j+k-1}.$$

If we choose $p_i = \left(\frac{i}{i+1}\right)^\alpha$ for some $\alpha \in (1, 2)$ then $P_j = \prod_{i=1}^{j-1} \left(\frac{i}{i+1}\right)^\alpha = \frac{1}{j^\alpha}$ and since the P_j are summable, $x_1 = \left(\sum_j P_j\right)^{-1}$ is well defined and positive. Then

$$\begin{array}{rcl} \displaystyle\int_\Omega \tau_U \, d\mu &=& \displaystyle x_1 \sum_k k \sum_j \frac{1}{(j+k-1)^\alpha} q_{j+k-1} \\ &\geq& \displaystyle c_1 x_1 \sum_k k \sum_j \frac{1}{(j+k-1)^{\alpha+1}} \\ &\geq& \displaystyle c_2 \sum_k \frac{k}{k^\alpha} = \infty, \end{array}$$

as $\alpha < 2$, where we used that $q_{j+k-1} = 1 - \left(1 - \frac{1}{j+k-1}\right)^\alpha \geq c_1 \frac{1}{j+k-1}$ for some $c_1 > 0$. We thus see that the integral of τ_U over the entire space Ω diverges.

This can be converted to an example on a two-state shiftspace $\Sigma \subset \{0, 1\}^{\mathbb{Z}}$ by the single element mapping $\pi : \Omega \to \Sigma$ which maps $\pi(1) = 1$ and collapses all other symbols to 0, i.e. $\pi(j) = 0, j = 2, 3, \dots$. The measure μ is sent to the probability measure $\nu = \pi^*\mu$ which is invariant under the shift map.

6. Existence of invariant measures and extremality of ergodic measures

Put $\mathscr{M}(T)$ for the set of T-invariant probability measures on (Ω, T) where Ω is a compact metric space. The following result affirms that for continuous maps on compact spaces and invariant maps, invariant measures indeed exist. The mainpart of the proof is to show that the set of probability measures on Ω is weak* compact.

Theorem 28. *(Krylov-Bogolioubov) If Ω is a compact metric space and $T : \Omega \to \Omega$ is continuous, then $\mathscr{M}(T) \neq \emptyset$.*

Proof. Let $x \in \Omega$ be arbitrary and let $\mu_n = \frac{1}{n}\sum_{j=1}^{n-1} \delta_x \circ T^j$, where δ_x is the unit pointmass at x. That is, for $f \in C(\Omega)$ one has $\mu_n(f) = \frac{1}{n} f^n(x) = \frac{1}{n}\sum_{j=1}^{n-1} f(T^j x)$ (here we use continuity of T to make sure that $T_* f = f \circ T$ lies in $C(\Omega)$). Let $\mathcal{S} = \{f_1, f_2, \dots\} \subset C(\Omega)$ be dense and countable. Then $|\mu_n(f_1)| \leq |f_1|_\infty < \infty$ and consequently there exists a subsequence n_j so that the limit $\lim_{j\to\infty} \mu_{n_j}(f_1) = L(f_1)$ exists for some number $L(f_1)$. Put $\mu_{1,j} = \mu_{n_j}$ and there exists a subsequence j_ℓ so that $\mu_{1,j_\ell}(f_2)$ converges to some $L(f_2)$ as $\ell \to \infty$. Put $\mu_{2,\ell} = \mu_{1,j_\ell}$ and proceed inductively. We obtain a sequence of sequences $\mu_{k,j}$ so that $\mu_{k,j}(f_k)$ converges to a limit $L(f_k)$ as $j \to \infty$ and so that $\{\mu_{k+1,\ell} : \ell\}$ is a subsequence of $\{\mu_{k,j} : j\}$ for every k. For the diagonal sequence $\nu_k = \mu_{k,k}$ the values $\nu_k(f)$ converge to a limit $L(f)$ for every $f \in \mathcal{S}$. Thus L defines a positive linear functional on $\mathcal{S}$ which can be extended to all of $C(\Omega)$ as follows. Let $g \in C(\Omega)$ and $\varepsilon > 0$. Then there exists $f \in \mathcal{S}$ so that $|g - f|_\infty < \varepsilon$. Hence $|\nu_k(g) - \nu_k(f)| < \varepsilon$ for all k. Moreover there exists an N so that $|\nu_k(f) - \nu_\ell(f)| < \varepsilon$ for all $k, \ell \geq N$. Hence

$$|\nu_k(g) - \nu_\ell(g)| \leq |\nu_k(g) - \nu_k(f)| + |\nu_k(f) - \nu_\ell(f)| + |\nu_\ell(f) - \nu_\ell(g)| < 3\varepsilon$$

for all $k, \ell \geq N$. Thus $\{\nu_k(g) : k\}$ is a Cauchy sequence and converges to a value $L(g)$. Since L is a positive linear functional on Ω, by the Riesz representation theorem there exists a (Radon) measure μ on Ω so that $L(f) = \int_\Omega f(x)\, d\mu(x)$. Clearly the measure μ is T-invariant as it is a limit of $\{\mu_n : n\}$ and

$$\begin{aligned}
|\mu_n(f \circ T) - \mu_n(f)| &= \left| \frac{1}{n} \int_\Omega \left(\sum_{j=0}^{n-1} f \circ T^{j+1} - \sum_{j=0}^{n-1} f \circ T^j \right) \mu_n \right| \\
&= \left| \frac{1}{n} \int_\Omega (f \circ T^n - f)\mu_n \right| \\
&\leq \frac{2|f|_\infty}{n} \to 0
\end{aligned}$$

as n goes to infinity along a (sub)sequence. ∎

Example: The requirement that the map be continuous is necessary as the following example shows where we produce a non-continuous map that does not have an invariant probability measure. Let $\Omega = [0, 1]$ and $T : [0, 1] \to [0, 1]$ defined by

$$Tx = \begin{cases} \frac{1}{2}x & \text{if} \quad 0 < x \leq 1 \\ 1 & \text{if} \quad x = 0 \end{cases}$$

for $x \in [0,1]$. Then T not continuous and has no invariant probability measure since (i) the pointmass δ_0 is not T-invariant and (ii) if there were an invariant measure μ with non-zero mass on, say, the interval $(\frac{1}{2},1]$, then by T-invariance of μ one would have $\mu((\frac{1}{2},1]) = \mu((\frac{1}{4},\frac{1}{2}]) = \mu((\frac{1}{8},\frac{1}{4}]) = \cdots = \mu((2^{-(j+1)},2^{-j}])$ for $j = 0,1,2,\ldots$. Since the intervals $(2^{-(j+1)},2^{-j}]$ are pairwise disjoint, the total measure of $(0,1]$ would have be to infinite which is impossible.

Next we will identify the ergodic measures as the irreducible components of invariant measures. Since the the linear combination of two invariant measures is again an invariant measures, the set of invariant measures $\mathscr{M}(T)$ is convex. Its boundary elements are the ergodic measures.

Definition 29. *We say $\mu \in \mathscr{M}(T)$ is* not extremal *if there exist $\mu_1,\mu_2 \in \mathscr{M}(T)$ $\mu_1 \neq \mu_2$ and $\alpha \in (0,1)$ so that $\mu = \alpha\mu_1 + (1-\alpha)\mu_2$.*

Lemma 30. *μ ergodic $\iff$ $\nexists \tilde{\mu} \in \mathscr{M}(T)$ so that $\tilde{\mu} \neq \mu, \tilde{\mu} \ll \mu$.*

Proof. "$\Rightarrow$": Assume μ is ergodic and assume there exists $\tilde{\mu} \in \mathscr{M}(T)$ so that $\tilde{\mu} \ll \mu$. If $\tilde{\mu}$ is not ergodic then there exists $U \subset \Omega, T^{-1}U = U$ so that $\tilde{\mu}(U), \tilde{\mu}(U^c) > 0$. However, $\tilde{\mu}(U) > 0 \Rightarrow \mu(U) > 0$ (as $\tilde{\mu} \ll \mu$) and similarly, $\tilde{\mu}(U^c) > 0 \Rightarrow \mu(U^c) > 0$. This contradicts the assumption that μ is ergodic. Hence $\tilde{\mu}$ is ergodic. In order to show that $\tilde{\mu} = \mu$ let $U \subset \Omega$ and χ_U its characteristic function. Then by the pointwise ergodic theorem

$$\frac{1}{n}\chi_U^n \to \begin{cases} \mu(U) & \text{for } \mu\text{-a.e. } x \\ \tilde{\mu}(U) & \text{for } \tilde{\mu}\text{-a.e. } x \end{cases}.$$

Hence $\mu(U) = \tilde{\mu}(U) \forall U \subset \Omega$ and therefore $\tilde{\mu} = \mu$.

"$\Leftarrow$": Assume there is no $\tilde{\mu} \in \mathscr{M}(T)$ so that $\tilde{\mu} \ll \mu$ and $\tilde{\mu} \neq \mu$. We show that μ is ergodic by contradiction. If we assume that μ is not ergodic, then there exists $U \subset \Omega, T^{-1}U = U$ so that $\mu(U), \mu(U^c) > 0$. Put $\mu_U = \mu|_U$ for the restricted probability measure, then μ_U is T-invariant as U is T-invariant. Moreover, $\mu_U \ll \mu$ and $\mu_U \neq \mu$ which contradicts the assumption. Hence μ is ergodic. ∎

Corollary 31. *For a continuous map T on a compact metric space Ω, the set of invariant probability measures $\mathscr{M}(T)$ is convex and its extremal elements are exactly the ergodic measures.*

If there is only one invariant measure μ (i.e. $\mathscr{M}(T) = \{\mu\}$) then we call (Ω, T, μ) *uniquely ergodic* as by force μ is ergodic.

Example (Irrational rotation). We give an example of a map which is uniquely ergodic. Let $\Omega = [0,1)$ be the circle $\mathbb{T}^1$ and T be the rotation by an irrational number α, i.e. $Tx = x + \alpha \bmod 1$. Clearly, T is continuous and we can use Fourier series on $C([0,1))$. If $g_k(x) = e^{2\pi i k x}$, then

$$\frac{1}{n}\sum_{j=0}^{n-1} g_k(T^j x) = \frac{1}{n}\sum_{j=0}^{n-1} e^{2\pi i k(x+j\alpha)} = \begin{cases} 1 & \text{if } k = 0 \\ \frac{1}{n}e^{2\pi i k x}\frac{1-e^{2\pi i k\alpha n}}{1-e^{2\pi i k\alpha}} & \text{if } k \neq 0 \end{cases}$$

as $\sum_{j=0}^{n-1} t^j = \frac{1-t^n}{1-t}$ if $t \neq 1$. Letting $n \to \infty$ yields $\frac{1}{n}g_k^n(x) \to 0$ for all $x \in \mathbb{T}^1$ and for all $k \neq 0$, and $\frac{1}{n}g_0^n(x) = 1$ for all x. If μ is a T-invariant measure on $[0,1)$ then by the ergodic theorem $\frac{1}{n}g_k^n(x) \to \int_0^1 g_k\,d\mu$ as $n \to \infty$ for μ-almost every x. Hence $\int_0^1 g_k\,d\mu = 0$ for $k \neq 0$ and equal to 1 if $k = 0$. If $P(x)$ is a trigonometric polynomial $\sum_{|k|\leq m} a_k g_k(x)$ of some degree m, then

$$\frac{1}{n}P^n(x) \to a_0$$

as $n \to \infty$, where $a_0 = \int_0^1 P(x)\,d\lambda(x)$ and also equal to $\int_0^1 P(x)\,d\mu(x)$. If P_j is a sequence of trigonometric polynomials that converge to a characteristic functions $\chi_{[a,b)}$ of an interval $[a,b) \subset [0,1)$, then

$$a_0 = \int_0^1 P\,d\mu \to \int_0^1 \chi_{[a,b)}\,d\mu$$

as $j \to \infty$. Hence $a_0 = \int_0^1 \chi_{[a,b)}\,d\mu = \lambda([0,1))$ for all $0 \leq a < b \leq 1$. Thus $\mu = \lambda$ and consequently, λ is the only ergodic measure on $\mathbb{T}^1$.

Theorem 32. *(Ergodic decomposition) For every T-invariant measure on the compact metric space Ω there exists a probability measure ρ on the ergodic set $\mathscr{E} = \{\nu \in \mathscr{M}(T) : \nu \text{ ergodic}\}$ of $\mathscr{M}(T)$ so that for all $f \in C(\Omega)$*

$$\int_\Omega f\,d\mu = \int_{\mathscr{E}} \left(\int_\Omega f\,d\nu\right)\,d\rho(\nu)$$

That is $\mathscr{M}(T)$ is a *Choquet simplex* (Choquet 1959).

Example. In the shiftspace $\Sigma = \{0,1\}^{\mathbb{Z}}$ the ergodic measures form a dense set in $\mathscr{M}(\sigma)$.

7. Entropy

Definition 33. *Given (Ω,μ) with σ-algebra $\mathscr{B}$.*
(I) $\mathcal{A} \subset \mathscr{B}$ is a partition *of Ω if elements in $\mathcal{A}$ are pairwise disjoint subsets of Ω and $\bigcup_{A\in\mathcal{A}} A = \Omega$.*
(II) For two partitions $\mathcal{A},\mathcal{B}$ we call

$$\mathcal{A} \vee \mathcal{B} = \{A \cap B : A \in \mathcal{A}, B \in \mathcal{B}\}$$

the join *of $\mathcal{A}$ and $\mathcal{B}$.*
(III) We say $\mathcal{A}$ is finer *than $\mathcal{B}$ ($\mathcal{B}$ is* coarser *than $\mathcal{A}$) if for every $A \in \mathcal{A}$ there exists a $B \in \mathcal{B}$ so that $A \subset B$.*

Example. For the shiftspace $\Sigma = \{1,2,\dots,M\}^{\mathbb{Z}}$ the set $\mathcal{A} = \{U(i) : i = 1,\dots,M\}$ forms a partition ($U(i) = \{\vec{x} \in \Sigma : x_0 = i\}$ are 1-cylinder sets). Also $\mathcal{B} = \{\sigma^{-1}U(i) : i\}$ is a partition and so is $\mathcal{A} \vee \mathcal{B} = \{U(ij) : i,j = 1,\dots,M\}$.

Let us now define the function

$$\varphi(t) = \begin{cases} 0 & \text{if } t = 0 \\ -t\log t & \text{if } t > 0 \end{cases}$$

which is concave on $[0,\infty)$ which means that for $x_i \in [0,\infty)$ and weights $\alpha_i \geq 0$ so that $\sum_i \alpha_i = 1$ one has $\varphi\left(\sum_i \alpha_i x_i\right) \geq \sum_i \alpha_i \varphi(x_i)$.

Definition 34. *Let μ be a probability measure on Ω and $\mathcal{A}$ a partition. Then*

$$H(\mathcal{A}) = \sum_{A\in\mathcal{A}} \varphi(\mu(A))$$

is the entropy *of μ with respect to the partition $\mathcal{A}$.*

If we introduce the *information function* $I_{\mathcal{A}}(x) = -\log \mu(A(x)) = \sum_{A\in\mathcal{A}} |\log \mu(A)|\chi_A(x)$ where $A(x) \in \mathcal{A}$ is the partition element that contains x, then

$$H(\mathcal{A}) = \sum_{A\in\mathcal{A}} \mu(A) I_{\mathcal{A}}(A) = \int_\Omega I_{\mathcal{A}}(x)\, d\mu(x).$$

Example 1. Let $\Sigma = \{1, \ldots, M\}^{\mathbb{N}_0}$ be the full M-shift and μ the invariant measure (Bernoulli measure) generated by the probability vector $\vec{p} = (p_1, \ldots, p_m)$. For the partition $\mathcal{A} = \{U(i) : i\}$ one then has

$$H(\mathcal{A}) = \sum_i \varphi(\mu(U(i)) = \sum_i p_i |\log p_i|.$$

If all the p_i are equal to $\frac{1}{M}$, then $H(\mathcal{A}) = \log M$.

Example 2. For the affine stretching map T on the interval $[0,1)$ $(= \mathbb{T}^1)$ one has $Tx = dx \bmod 1$ for degree $d \geq 2$. For the Lebesgue measure λ and the partition $\mathcal{A} = \{[\frac{i}{d}, \frac{i+1}{d}) : i = 0, 1, \ldots, d-1\}$ one obtains

$$H(\mathcal{A}) = -\sum_i \lambda([\frac{i}{d}, \frac{i+1}{d})) \log \lambda([\frac{i}{d}, \frac{i+1}{d})) = \sum_i -\frac{1}{d} \log \frac{1}{d} = \log d$$

Definition 35. *Let μ be a probability measure on Ω and $\mathcal{A}, \mathcal{B}$ partitions. Then*

$$H(\mathcal{A}|\mathcal{B}) = \sum_B \mu(B) \sum_A \varphi\left(\frac{\mu(A\cap B)}{\mu(B)}\right) = \sum_{A,B} -\mu(A\cap B) \log \frac{\mu(A\cap B)}{\mu(B)}$$

is the conditional entropy *of $\mathcal{A}$ with respect to $\mathcal{B}$.*

Theorem 36. *For partitions $\mathcal{A}, \mathcal{B}, \mathcal{C}$ one has:*
(I) $H(\mathcal{A}\vee\mathcal{B}|\mathcal{C}) = H(\mathcal{A}|\mathcal{C}) + H(\mathcal{B}|\mathcal{A}\vee\mathcal{C})$
(II) $H(\mathcal{A}\vee\mathcal{B}) = H(\mathcal{A}) + H(\mathcal{B}|\mathcal{A})$
(III) If $\mathcal{B}$ is finer than $\mathcal{A}$ then $H(\mathcal{B}|\mathcal{C}) \geq H(\mathcal{A}|\mathcal{C})$ (and also $H(\mathcal{B}) \geq H(\mathcal{A})$)
(IV) If $\mathcal{B}$ is finer than $\mathcal{C}$ then $H(\mathcal{A}|\mathcal{B}) \leq H(\mathcal{A}|\mathcal{C})$
(V) $H(\mathcal{A}\vee\mathcal{B}|\mathcal{C}) \leq H(\mathcal{A}|\mathcal{C}) + H(\mathcal{B}|\mathcal{C})$
(VI) $H(\mathcal{A}|\mathcal{C}) \leq H(\mathcal{A}|\mathcal{B}) + H(\mathcal{B}|\mathcal{C})$

Also note that $H(\mathcal{A}|\mathcal{B}) = 0$ if $\mathcal{B}$ is finer than $\mathcal{A}$ because then

$$H(\mathcal{A}|\mathcal{B}) = -\sum_{A,B} \mu(A\cap B) \log \frac{\mu(A\cap B)}{\mu(B)} = -\sum_{A,B} \mu(A\cap B) \log \frac{\mu(B)}{\mu(B)} = 0$$

as either $A\cap B = \emptyset$, or $A\cap B = B$ as $\mathcal{A}\vee\mathcal{B} = \mathcal{B}$.

Proof. (I) One has

$$
\begin{aligned}
H(\mathcal{A}\vee\mathcal{B}|\mathcal{C}) &= -\sum_{A,B,C}\mu(A\cap B\cap C)\log\frac{\mu(A\cap B\cap C)}{\mu(C)}\\
&= -\sum_{A,B,C}\mu(A\cap B\cap C)\left(\log\frac{\mu(A\cap B\cap C)}{\mu(A\cap C)}+\log\frac{\mu(A\cap C)}{\mu(C)}\right)\\
&= H(\mathcal{B}|\mathcal{A}\vee\mathcal{C})+H(\mathcal{A}|\mathcal{C}).
\end{aligned}
$$

(II) This is a special case of (I) with the trivial partition $\mathcal{C}=\{\Omega\}$.
(III) As $\mathcal{B}$ is finer than $\mathcal{A}$ one has $\mathcal{A}\vee\mathcal{B}=\mathcal{B}$ and thus by (I)

$$
H(\mathcal{B}|\mathcal{C})=H(\mathcal{A}\vee\mathcal{B}|\mathcal{C})=H(\mathcal{A}|\mathcal{C})+H(\mathcal{B}|\mathcal{A}\vee\mathcal{C})\ge H(\mathcal{A}|\mathcal{C})
$$

as $H(\cdot|\cdot)\ge 0$.
(IV) Since $\mathcal{B}$ is finer than $\mathcal{C}$, for every $C\in\mathcal{C}$ there is a subset $\mathcal{B}_C\subset\mathcal{B}$ so that $C=\bigcup_{B\in\mathcal{B}_C}B$. Then by concavity of the function φ

$$
\begin{aligned}
H(\mathcal{A}|\mathcal{B}) &= \sum_{A\in\mathcal{A},B\in\mathcal{B}}\mu(B)\varphi\left(\frac{\mu(A\cap B)}{\mu(B)}\right)\\
&= \sum_{A\in\mathcal{A},C\in\mathcal{C}}\mu(C)\sum_{B\in\mathcal{B}_C}\frac{\mu(B)}{\mu(C)}\varphi\left(\frac{\mu(A\cap B)}{\mu(B)}\right)\\
&\le \sum_{A\in\mathcal{A},C\in\mathcal{C}}\mu(C)\varphi\left(\sum_{B\in\mathcal{B}_C}\frac{\mu(B)}{\mu(C)}\frac{\mu(A\cap B)}{\mu(B)}\right)\\
&= \sum_{A\in\mathcal{A},C\in\mathcal{C}}\mu(C)\varphi\left(\frac{\mu(A\cap C)}{\mu(C)}\right)\\
&= H(\mathcal{A}|\mathcal{C})
\end{aligned}
$$

as the weights $\frac{\mu(B)}{\mu(C)}$ add up to 1 on $\mathcal{B}_C$ for every $C\in\mathcal{C}$ and $\sum_{B\in\mathcal{B}_C}\mu(A\cap B)=\mu(A\cap C)$.
(V) By (IV) $H(\mathcal{B}|\mathcal{C})\ge H(\mathcal{B}|\mathcal{C}\vee\mathcal{A})$ and thus by (I):

$$
H(\mathcal{A}\vee\mathcal{B}|\mathcal{C})=H(\mathcal{A}|\mathcal{C})+H(\mathcal{B}|\mathcal{A}\vee\mathcal{C})\le H(\mathcal{A}|\mathcal{C})+H(\mathcal{B}|\mathcal{C}).
$$

(VI) Using (III) and then (I) yields

$$
H(\mathcal{A}|\mathcal{C})\le H(\mathcal{A}\vee\mathcal{B}|\mathcal{C})=H(\mathcal{B}|\mathcal{C})+H(\mathcal{A}|\mathcal{B}\vee\mathcal{C})\le H(\mathcal{B}|\mathcal{C})+H(\mathcal{A}|\mathcal{B}).
$$

∎

Now we come to the entropy of a map. Let T be a map on Ω and μ a T-invariant probability measures. If $\mathcal{A}$ is a partition of Ω then so is $T^{-1}\mathcal{A}$ (and higher order pull-backs). Since μ is invariant one has $H(T^{-1}\mathcal{A})=H(\mathcal{A})$. The partition $\mathcal{A}^n=\bigvee_{j=0}^{n-1}T^{-j}\mathcal{A}$ is called the *nth join* of $\mathcal{A}$ and it refines the partitions $\mathcal{A},T^{-1}\mathcal{A},\dots,T^{-(n-1)}\mathcal{A}$. The atoms A of $\mathcal{A}^n$ are of the form $A=A_{i_0}\cap A_{i_1}\cap\cdots\cap A_{i_{n-1}}$ where $A_{i_k}\in\mathcal{A}=\{A_1,A_2,\dots\}$. We call A an *n-cylinder*.

Example 1. For the full shift $\Sigma = \{1, \ldots, M\}^{\mathbb{N}_0}$ we take the partition $\mathcal{A} = \{U(i) : i\}$ of 1-cylinders. Then

$$\mathcal{A}^n = \{U(i_0 i_1 i_2 \cdots i_{n-1}) : i_k \in \{1, \ldots, M\}, k = 0, \ldots, n-1\}$$

consists of the collection of n-cylinders $U(i_0 i_1 \cdots i_{n-1}) = \{\vec{x} \in \Sigma : x_0 \cdots x_{n-1} = i_0 \cdots i_{n-1}\}$. Clearly, here $|\mathcal{A}^n| = M^n$.

Example 2. For the doubling map $T : [0, 1) \to [0, 1)$ given by $Tx = 2x \bmod 1$ we take the two element partition $\mathcal{A} = \{[0, \frac{1}{2}), [\frac{1}{2}, 1)\}$. Then

$$\mathcal{A}^n = \left\{ \left[\frac{j}{2^n}, \frac{j+1}{2^n} \right) : j = 0, 1, \ldots, 2^n - 1 \right\} = \{I_{x_1 x_2 \cdots x_n} : x_k \in \{0, 1\}, k = 1, \ldots, n\},$$

where the intervals $I_{x_1 x_2 \cdots x_n}$ consists of all points $x \in [0, 1)$ whose binary expansions begin with the digits $x_1, x_2, \ldots, x_n$. That is

$$I_{x_1 x_2 \cdots x_n} = \left[\sum_{j=1}^{n} \frac{x_j}{2^j}, \sum_{j=1}^{n} \frac{x_j}{2^j} + \frac{1}{2^n} \right).$$

7.1. **The Kolmogorov entropy.** In order to prove the existence of the limit we will need the following arithmetic lemma.

Lemma 37. *Let $\{a_n : n \in \mathbb{N}\}$ be a positive and subadditive sequence, that is $a_{n+m} \leq a_n + a_m$ for all $m, n \in \mathbb{N}$. Then the following limit exists:*

$$\lim_{n\to\infty} \frac{a_n}{n} = \inf_{n\in\mathbb{N}} \frac{a_n}{n}.$$

Proof. Let $m \geq 1$ and put $n = km + r$, where $0 \leq r < m$ (remainder). Then

$$\frac{a_n}{n} = \frac{a_{km+r}}{n} \leq \frac{a_r}{n} + \frac{ka_m}{n} \leq \frac{a_r}{n} + \frac{ka_m}{km} = \frac{1}{k} + \frac{a_m}{m}$$

since $a_{km} \leq ka_m$ by subadditivity. Now let $n \to \infty$ ($k \to \infty$) along a sequence that gives the lim sup and we get

$$\limsup_{n\to\infty} \frac{a_n}{n} \leq \frac{a_m}{m}$$

for every $m \geq 1$. Taking a liminf on the RHS gives the existence of the limit. That the limit equals the inf follows from the last inequality. ∎

Definition 38. *If the limit exists, then $h_\mu(T, \mathcal{A}) = \lim_{n\to\infty} \frac{1}{n} H(\mathcal{A}^n)$ is the* entropy *of μ with respect to T and $\mathcal{A}$.*

Theorem 39. *Let T be a map on a space Ω, μ a T-invariant probability measure and $\mathcal{A}$ a measurable partition. Then*

$$h(\mu, \mathcal{A}) = \lim_{n\to\infty} \frac{1}{n} H(\mathcal{A}^n)$$

exists.

Proof. Note that $\mathcal{A}^{n+m} = \mathcal{A}^n \vee T^{-n}\mathcal{A}^m$. Since by (II) and (III) of the theorem one has $H(\mathcal{B} \vee \mathcal{C}) \leq H(\mathcal{B}) + H(\mathcal{C})$ for any two partitions $\mathcal{B}, \mathcal{C}$, we obtain:

$$H(\mathcal{A}^{n+m}) = H(\mathcal{A}^n \vee T^{-n}\mathcal{A}^m) \leq H(\mathcal{A}^n) + H(T^{-n}\mathcal{A}^m) = H(\mathcal{A}^n) + H(\mathcal{A}^m).$$

Thus the sequence $a_n = H(\mathcal{A}^n)$, $n = 1, 2, \dots$, is subadditive and we can apply the arithmetic lemma to obtain the limit

$$h(\mu, \mathcal{A}) = \lim_{n\to\infty} \frac{1}{n} H(\mathcal{A}^n) = \inf_{n\in\mathbb{N}} \frac{1}{n} H(\mathcal{A}^n).$$

∎

Lemma 40. *$h(\mu, \mathcal{A}) \leq h(\mu, \mathcal{B}) + H(\mathcal{A}|\mathcal{B})$ for any two partitions $\mathcal{A}, \mathcal{B}$.*

Proof. Obviously $H(\mathcal{A}^n) \leq H(\mathcal{A}^n \vee \mathcal{B}^n)$ as $\mathcal{A}^n \vee \mathcal{B}^n$ is finer than $\mathcal{A}^n$. By property (II) of the theorem

$$H(\mathcal{A}^n \vee \mathcal{B}^n) = H(\mathcal{B}^n) + H(\mathcal{A}^n|\mathcal{B}^n).$$

The second term on the RHS is

$$\begin{aligned} H(\mathcal{A}^n|\mathcal{B}^n) &= H(\mathcal{A} \vee T^{-1}\mathcal{A} \vee T^{-2}\mathcal{A} \vee \cdots \vee T^{-1}\mathcal{A}|\mathcal{B}^n) \\ &\leq H(\mathcal{A}|\mathcal{B}^n) + H(T^{-1}\mathcal{A}|\mathcal{B}^n) + H(T^{-1}\mathcal{A}|\mathcal{B}^n) + \cdots + H(T^{-(n-1)}\mathcal{A}|\mathcal{B}^n) \\ &\leq H(\mathcal{A}|\mathcal{B}^n) + H(T^{-1}\mathcal{A}|T^{-1}\mathcal{B}) + H(T^{-2}\mathcal{A}|T^{-2}\mathcal{B}) + \cdots + H(T^{-(n-1)}\mathcal{A}|T^{-(n-1)}\mathcal{B}) \end{aligned}$$

because $T^{-j}\mathcal{B}$ is coarser than $\mathcal{B}^n$ and therefore $H(\mathcal{C}|\mathcal{B}^n) \leq H(\mathcal{C}|T^{-j}\mathcal{B})$ for any partition $\mathcal{C}$. Because of invariance of the measure $H(T^{-j}\mathcal{A}|T^{-j}\mathcal{B}) = H(\mathcal{A}|\mathcal{B})$ and thus

$$H(\mathcal{A}|\mathcal{B}^n) \leq nH(\mathcal{A}|\mathcal{B}).$$

Dividing by n yields

$$\frac{1}{n} H(\mathcal{A}^n) \leq \frac{1}{n} H(\mathcal{B}^n) + H(\mathcal{A}|\mathcal{B})$$

and the limit $n \to \infty$ proves the lemma. ∎

Corollary 41. *$h(\mu, \mathcal{A}) \leq h(\mu, \mathcal{B})$ if $\mathcal{B}$ is finer than $\mathcal{A}$.*

Proof. By the previous lemma $h(\mu, \mathcal{A}) \leq h(\mu, \mathcal{B}) + H(\mathcal{A}|\mathcal{B})$ and since $\mathcal{B}$ is finer than $\mathcal{A}$ one gets that $H(\mathcal{A}|\mathcal{B}) = 0$. ∎

Remark. We also have $h(\mu, \mathcal{A}^n) = h(\mu, \mathcal{A})$ for any $n \in \mathbb{N}$ because

$$(\mathcal{A}^n)^k = \bigvee_{j=0}^{k-1} T^{-j} \bigvee_{i=0}^{n-1} T^{-i}\mathcal{A} = \bigvee_{j=0}^{k+n-1} T^{-j}\mathcal{A} = \mathcal{A}^{n+k-1}$$

which implies

$$\begin{aligned} h(\mu, \mathcal{A}^n) &= \lim_{k\to\infty} \frac{1}{k} H((\mathcal{A}^n)^k) \\ &= \lim_{k\to\infty} \frac{1}{k} H(\mathcal{A}^{n+k-1}) \\ &= \lim_{k\to\infty} \frac{n+k-1}{k} \lim_{k\to\infty} \frac{1}{k+n-1} H(\mathcal{A}^{n+k-1}) \\ &= h(\mu, \mathcal{A}). \end{aligned}$$

Definition 42. *The* measure theoretic entropy *of μ is*
$$h(\mu) = \sup_{\mathcal{A}} h(\mu, \mathcal{A}),$$
where the supremum is over all finite partitions $\mathcal{A}$ of Ω.

The definition may be extended to include infinite partitions $\mathcal{A}$ under the additional assumption that $H(\mathcal{A})$ be finite.

Definition 43. *A partition $\mathcal{A}$ is a* generator *(or μ-*generator*) if $\bigcup_n \mathcal{A}^n$ generates the σ-algebra on Ω (up to μ-nullsets).*

If the map T is invertible then $\mathcal{A}$ is a μ-generator if $\{\bigvee_{j=-n}^{n} T^{-j}\mathcal{A} : n \in \mathbb{N}\}$ generates the σ-algebra.

Theorem 44. *(Kolmogorov-Sinai) If $\mathcal{A}$ is a μ-generator then*
$$h(\mu) = h(\mu, \mathcal{A}).$$

Proof. Let $\mathcal{A}$ be a μ-generator and $\mathcal{B}$ be an arbitrary finite partition. We have to show that $h(\mu, \mathcal{B}) \leq h(\mu, \mathcal{A})$. For any n we have
$$h(\mu, \mathcal{B}) \leq h(\mu, \mathcal{A}^n) + H(\mathcal{B}|\mathcal{A}^n) = h(\mu, \mathcal{A}) + H(\mathcal{B}|\mathcal{A}^n).$$
We want to show that $H(\mathcal{B}|\mathcal{A}^n)$ can be made arbitrarily small if n is large enough. Let $\mathcal{B} = \{B_1, B_2, \ldots, B_r\}$ ($r = |\mathcal{B}|$) and $\varepsilon > 0$. Since $\mathcal{A}$ is a μ-generator, we can find $A_j^{(n)} \in \sigma(\mathcal{A}^n)$, unions of n-cylinders so that $\mu(B_j \triangle A_j^{(n)}) < \varepsilon\mu(A_j^{(n)})$, $j = 1, 2, \ldots, r$. We can assume that $\mathcal{A}^{(n)} = \{A_j^{(n)} : j\}$ forms a partition of Ω. Clearly $\mathcal{A}^n$ is finer than $\mathcal{A}^{(n)}$ and therefore
$$H(\mathcal{B}|\mathcal{A}^n) \leq H(\mathcal{B}|\mathcal{A}^{(n)}) = \sum_{j,k} -\mu(A_k^{(n)})\varphi\left(\frac{\mu(A_k^{(n)} \cap B_j)}{\mu(A_k^{(n)})}\right).$$
We have two cases, (i) when $j = k$ and (ii) $j \neq k$. If $j = k$ then
$$\mu(A_j^{(n)} \triangle B_j) \leq \mu(A_j^{(n)} \cup B_j) - \mu(A_j^{(n)} \cap B_j) < \varepsilon\mu(A_j^{(n)})$$
which implies that
$$\frac{\mu(A_j^{(n)} \cap B_j)}{\mu(A_j^{(n)})} > \frac{\mu(A_j^{(n)} \cup B_j)}{\mu(A_j^{(n)})} - \varepsilon \geq 1 - \varepsilon.$$
If $j \neq k$ then
$$\mu(A_k^{(n)} \cap B_j) \leq \mu(A_k^{(n)} \cap B_k^c) \leq \mu(A_k^{(n)} \triangle B_j) < \varepsilon\mu(A_k^{(n)})$$
which implies
$$\frac{\mu(A_k^{(n)} \cap B_j)}{\mu(A_k^{(n)})} < \varepsilon.$$
Hence
$$H(\mathcal{B}|\mathcal{A}^n) \leq \sum_k \mu(A_k^{(n)}) \left(\varphi(1-\varepsilon) + \sum_{j\neq k} \varphi(\varepsilon)\right) \leq c_1 r \varphi(\varepsilon).$$

Since $\varepsilon > 0$ was arbitrary we get $h(\mu, \mathcal{B}) \leq h(\mu, \mathcal{A})$ for all partition $\mathcal{B}$. Consequently, since $\varphi(\varepsilon) \to 0$ as $\varepsilon \to 0$ and $\varepsilon > 0$ is arbitrary, we conclude that $\sup_{\mathcal{B}} h(\mu, \mathcal{B}) \leq h(\mu, \mathcal{A})$. ∎

7.2. **Examples. (I) Bernoulli shift.** Let $\Sigma = \{1, \dots, M\}^{\mathbb{N}_0}$, with the shiftmap σ and the generating partition $\mathcal{A} = \{U(i) : i\}$. Then $\bigvee_{j=0}^{n} \sigma^{-j}\mathcal{A}$ generates the σ-algebra by definition. If μ is the measure generated by the probability vector $\vec{p} = (p_1, \dots, p_M)$ ($\sum_i p_i = 1$) then

$$\begin{aligned} H(\mathcal{A}^n) &= -\sum_{x_0 x_1 \cdots x_{n-1}} p_{x_0} \cdots p_{x_{n-1}} \log p_{x_0} \cdots p_{x_{n-1}} \\ &= -\sum_{x_0 x_1 \cdots x_{n-1}} p_{x_0} \cdots p_{x_{n-1}} \log p_{x_k} \\ &= -\sum_{x_k} p_{x_k} \log p_{x_k} \\ &= -n \sum_i p_i \log p_i \end{aligned}$$

and therefore $h(\mu) = -\sum_i p_i \log p_i$.

(II) Markov measure. Let A be an $M \times M$-transition matrix and $\Sigma = \{\vec{x} \in \{1, \dots, M\}^{\mathbb{N}_0} : A_{x_i x_{i+1}} = 1 \forall i\}$ and $\sigma : \Sigma \to \Sigma$ the left shift. Again the partition $\mathcal{A} = \{U(i) : i\}$ of 1-cylinder sets is generating. Let μ be the invariant measure induced by a stochastic matrix P ($P_{ij} = 0$ if $A_{ij} = 0$) and its probability left eigenvector $\vec{p} = (p_1, \dots, p_M)$ ($\sum_i p_i = 1$). By the theorem of Kolmogorov and Sinai $h(\mu) = h(\mu, \mathcal{A})$. We have

$$\begin{aligned} H(\mathcal{A}^n) &= -\sum_{x_0 x_1 \cdots x_{n-1}} \mu(U(x_0 \cdots x_{n-1})) \log \mu(U(x_0 \cdots x_{n-1})) \\ &= -\sum_{x_0 x_1 \cdots x_{n-1}} p_{x_0} P_{x_0 x_1} \cdots P_{x_{n-2} x_{n-1}} \log p_{x_0} P_{x_0 x_1} \cdots P_{x_{n-2} x_{n-1}} \\ &= -\sum_{x_0 x_1 \cdots x_{n-1}} p_{x_0} P_{x_0 x_1} \cdots P_{x_{n-2} x_{n-1}} \left(\log p_{x_0} + \sum_{k=1}^{n-1} \log P_{x_{k-1} x_k} \right) \\ &= -\sum_i p_i \log p_i - (n-1) \sum_{i,j} p_i P_{ij} \log P_{ij}, \end{aligned}$$

and therefore $h(\mu) = \lim_n \frac{1}{n} H(\mathcal{A}^n) = -\sum_{i,j} p_i P_{ij} \log P_{ij}$.

(III) Affine stretching interval map. $\Omega = [0, 1)$, $Tx = dx \bmod 1$ ($d \geq 2$ degree). Then

$$\mathcal{A} = \left\{ \left[\frac{i}{k}, \frac{i+1}{d} \right) : i = 0, 1, \dots, d-1 \right\}$$

is generating and

$$\mathcal{A}^n = \left\{ \left[\frac{j}{d^n}, \frac{j+1}{d^n} \right) : j = 0, 1, \dots, d^n - 1 \right\} = \{I_{i_1 i_2 \cdots i_n} : i_k \in \{0, 1, \dots, d-1\}\}$$

where

$$I_{i_1 i_2 \cdots i_n} = \left[\sum_{j=1}^{n} \frac{i_j}{d^j}, \sum_{j=1}^{n} \frac{i_j}{d^j} + \frac{1}{d^n} \right)$$

consists of the numbers $x \in [0,1)$ whose base d expansion begins with the digits $i_1, i_2, \ldots, i_n$. With λ the Lebesgue measure $\lambda(I_{i_1 i_2 \cdots i_n}) = d^{-n}$ and

$$H(\mathcal{A}^n) = \sum_{i_1, i_2, \ldots, i_n} -\lambda(I_{i_1 i_2 \cdots i_n}) \log \lambda(I_{i_1 i_2 \cdots i_n})) = n \log d.$$

Therefore $h(\lambda) = \log d$.

7.3. **The theorems of Shannon-McMillan-Breiman and Ornstein-Weiss.** For $x \in \Omega$ denote by $A_n(x)$ the unique atom in $\mathcal{A}^n$ which contains x.

Theorem 45. *(Shannon-McMillan-Breiman) Let μ be a T-invariant probability measure with entropy $h(\mu)$ and $\mathcal{A}$ a finite μ-generating partition. Then*

$$\lim_{n \to \infty} \frac{1}{n} |\log \mu(A_n(x))| = h(\mu)$$

for almost every $x \in \Omega$ and in $\mathscr{L}^1$.

This theorem was first proved by Shannon [**?**] in 1948 for Markov measures. He showed that the convergence was in measure. This was in 1957 improved upon by McMillan [**?**] who proved the theorem for ergodic measures and showed that convergence was in $\mathscr{L}^1$. The final version for finite alphabets was formulated by Breiman [**?**, **?**] in 1958. Subsequent generalisations to infinite partitions are due to Carleson [**?**] in 1961 for Markov measures and convergence in measure and Chung [**?**] for ergodic measures with almost sure convergence.

In order to prove the theorem we need the Martingale convergence theorem.

Theorem 46. *(Martingale convergence theorem) Let $f \in \mathscr{L}^1$ and $\mathcal{B}_1 \subset \mathcal{B}_2 \subset \mathcal{B}_3 \subset \cdots$ be a seqence of successively finer partitions so that $\bigcup_j \mathcal{B}_j$ generates the σ-algebra $\mathcal{B}$. Then the functions*

$$f_n = f|_{\mathcal{B}_n}$$

converge almost surely and in $\mathscr{L}^1$ to the limit $f|_{\mathcal{B}}$. The convergence is in $\mathscr{L}^1$ if $\sup_n |f_n|$ is integrable.

The mainpart of the proof of the SMB theorem is to show that the limit exists.

Lemma 47. *The functions $h_n(x) = \frac{1}{n} |\log \mu(A_n(x))|$ converge almost surely and in $\mathscr{L}^1$ to a limit F^*.*

Proof. If we put $F_k(x) = \log \frac{\mu(A_k(x))}{\mu(A_{k-1}(Tx))}$ then

$$\log \mu(A_n(x)) = \sum_{k=0}^{n-1} \log \frac{\mu(A_{n-k}(T^k x))}{\mu(A_{n-k-1}(T^{k+1} x))} + \log \mu(A_0(T^n x))$$

where $\log\mu(A_0(T^nx))=0$ as $\mathcal{A}_0=\{\Omega\}$ is the trivial partition. Thus

$$h_n(x)=-\frac{1}{n}\sum_{k=0}^{n-1}F_{n-k}(T^kx)+\mathcal{O}(1/n).$$

Let us now look more closely at the functions F_k:

$$e^{F_k(x)}=\frac{\mu(A_k(x))}{\mu(A_{k-1}(Tx))}=\frac{\mu(A_1(x)\cap T^{-1}A_{k-1}(Tx))}{\mu(A_{k-1}(Tx))}=\frac{1}{\mu(A_{k-1}(Tx))}\int_{T^{-1}A_{k-1}(Tx))}\chi_{A_1(x)}\,d\mu$$

is the average of $\chi_{A_1(x)}$ over the elements in the partition $T^{-1}\mathcal{A}^{k-1}$. We now can use the Martingale convergence theorem with the function $f(x)=\chi_{A_1(x)}$ and the successively refining partitions $\mathcal{B}_k=T^{-1}\mathcal{A}^{k-1}$. Thus e^{F_k} converges to a limit e^{F_∞} almost surely.

We shall now show that $\varphi=\sup_n -F_n$ is integrable. For $A\in\mathcal{A}$ and $t>0$ note that the sets

$$S_n^A(t)=\{x\in\Omega: -F_j(x)\le t, j=2,3,\dots,n-1,\ -F_n(x)>t\}$$

are $\mathcal{B}_n$-integrable and also are disjoint, i.e. $S_n^A(t)\cap S_n^A(t)=\varnothing$ if $n\ne m$. One has

$$\mu(A\cap S_n^A(t))=\int_{S_n^A(t)}e^{F_n}\,d\mu\le e^{-t}\mu(S_n^A(t))$$

and therefore

$$\mu(A\cap\{\varphi>t\})=\sum_{n=2}^{\infty}\mu(A\cap S_n^A(t))\le e^{-t}\sum_n\mu(S_n^A(t))\le e^{-t}$$

since the sets $S_n^A(t)$ are disjoint. From this we now conclude the integrability of φ as follows:

$$\begin{aligned}\int_\Omega\varphi\,d\mu &= \int_0^\infty\mu(\{\varphi>t\})\,dt\\ &= \sum_{A\in\mathcal{A}}\int_0^\infty\mu(A\cap\{\varphi>t\})\,dt\\ &\le \sum_A\int_0^\infty\min\{e^{-t},\mu(A)\}\,dt\\ &= \sum_A\left(\int_0^{-\log\mu(A)}\mu(A)\,dt+\int_{-\log\mu(A)}^\infty e^{-t}\,dt\right)\\ &= H(\mathcal{A})+1<\infty.\end{aligned}$$

To finish the proof of the lemma we now write

$$\begin{aligned}h_n(x) &= -\frac{1}{n}\sum_{k=0}^{n-1}F_{n-k}(T^kx)+\mathcal{O}(1/n)\\ &= -\frac{1}{n}\sum_{k=0}^{n-1}(F_{n-k}-F_\infty)(T^kx)-\frac{1}{n}\sum_{k=0}^{n-1}F_\infty(T^kx)+\mathcal{O}(1/n),\end{aligned}$$

where the last sum on the RHS converges by the ergodic theorem almost surely to a constant F^* as μ is ergodic. To estimate the first sum

$$S_n = -\frac{1}{n}\sum_{k=0}^{n-1}(F_{n-k} - F_\infty)(T^k x)$$

on the RHS put $\varphi_n = \sup_{j\ge n}|F_j - F_\infty|$ and observe that by the Martingale Convergence Theorem $\varphi_n \to 0$ almost surely as $n \to \infty$. Since $\varphi_n \le 2\varphi$ conclude by the Dominated Convergence Theorem that $\int \varphi_n \, d\mu \to 0$ as $n \to \infty$. If N is a large integer then for $n > N$:

$$|S_n| \le \frac{1}{n}\sum_{k=1}^{n-N}\varphi_N \circ T^k + \frac{1}{n}\sum_{k=n-N+1}^{n-1}\varphi_1 \circ T^k \le \frac{1}{n}\sum_{k=1}^{n}\varphi_N \circ T^k + \mathcal{O}\left(\frac{N}{n}\right).$$

By the Birkhoff Ergodic Theorem, the sum on the RHS converges almost surely to $\int \varphi_N \, d\mu$ as $n \to \infty$ while the error term goes to 0. Hence

$$\limsup_n |S_n| \le \int \varphi_N \, d\mu$$

almost everywhere. Now let $N \to \infty$ which implies that the RHS goes to zero. Thus we finally get that $h_n(x) \to F^*$ almost surely as $n \to \infty$.

In order to get the convergence in $\mathcal{L}^1$ notice that

$$\int_\Omega |S_n| \, d\mu \le \frac{1}{n}\sum_{k=1}^{n-1}\int \varphi_{n-k} \, d\mu,$$

where the RHS converges to 0 as n goes to infinity. ∎

Proof of the SMB theorem. It remains to show that the constant F^* from the previous lemma is equal to the entropy. Indeed

$$\int_\Omega h_n \, d\mu = \int_\Omega -\frac{1}{n}\log\mu(A_n(x)) \, d\mu(x) = -\frac{1}{n}\sum_{A\in\mathcal{A}^n}\mu(A)\log\mu(A) = \frac{1}{n}H(\mathcal{A}^n)$$

which converges to $h(\mu)$ as $\mathcal{A}$ is a μ-generator. ∎

If the measure is sufficiently well mixing we have indeed the Central Limit Theorem:

Theorem 48. *Let μ be a β-mixing probability measure on Ω with respect to a finite, measurable and generating partition $\mathcal{A}$. Assume that β decays at least polynomially with power > 6.*

If $\sigma > 0$ then

$$\mathbb{P}\left(\frac{I_n - nh}{\sigma\sqrt{n}} \le t\right) = N(t) + \mathcal{O}(n^{-\kappa})$$

for all t and all
(i) $\kappa < \frac{1}{10} - \frac{3}{5}\frac{1}{p+2}$ if β decays polynomially with power p,
(ii) $\kappa < \frac{1}{10}$ if β decays super polynomially.

As above let $\mathcal{A}$ be a μ-generating partition and denote by $A_n(x)$ the n-cylinder that contains x. Then

$$R_n(x) = \tau_{A_n(x)}(x)$$

denotes the *return time function*. In the symbolic description, when every point x is identified by its trajectory $\vec{x} = (\dots, x_{-1}, x_0, x_1, \dots)$ then

$$R_n(x) = \min\{j \geq 1 : x_j x_{j+1} \cdots x_{j+n-1} = x_0 x_1 \cdots x_{n-1}\}$$

measures the time it takes until one sees the starting n-word again. According to Kac's theorem the value of $\tau_{A_n(x)}$ is on average $1/\mu(A_n(x))$. Since the SMB theorem says that $\mu(A_n(x)) \sim e^{-nh}$ one would expect that $R_n(x) \sim e^{nh}$. This indeed is true as the following theorem shows.

Theorem 49. *(Ornstein-Weiss* [24]*) Let μ be ergodic and $\mathcal{A}$ a finite μ-generating partition, then almost surely*

$$\lim_{n\to\infty} \frac{\log R_n(x)}{n} = h(\mu).$$

Proof. We notice that $R_{n-1}(x) \leq R_n(x)$ and also $R_{n-1} \leq R_n(x)$ for all $x \in \Omega$. Thus, if

$$R^+(x) = \limsup_{n\to\infty} \frac{1}{n} \log R_n(x), \qquad R^-(x) = \liminf_{n\to\infty} \frac{1}{n} \log R_n(x)$$

then

$$R^+ \circ T \leq R^+, \qquad R^- \circ T \leq R^-$$

from which it follows that $R^{\pm}$ are constant almost everywhere as μ is ergodic[1]. We split the proof into two parts: In part (I) we show $R^+ \leq h$ and in part (II) we show that $R^- \geq h$.

(I) Suppose $R^+ > h$ and choose b, c so that $R^+ > b > c > h$. For $A \in \mathcal{A}^n$ let

$$E_A = \{x \in A : R_n(x) \geq e^{nb}\}.$$

Then $E_A \cap T^j E_A = \emptyset$ for $j = 1, 2, \dots, S$, where $S = [e^{nb}] - 1$. This is because if $x \in E_A \cap T^j E_A \neq \emptyset$ then $R_n(x) \leq j$ thus contradicting the definition of E_A which demands that R_n be larger than S. Similarly $T^k E_A \cap T^j E_A = \emptyset$ for $j \neq k$, $j,k = 0, 1, \dots, S$, because (assming $k < j$) $T^k E_A \cap T^j E_A = T^k(E_A \cap T^{j-k} E_A) = \emptyset$. Thus the sets $E_A, TE_A, T^2 E_A, \dots, T^S E_A$ are pairwise disjoint. Moreover as $E_A \subset T^{-j}(T^j E_A)$ one has $\mu(E_A) \leq \mu(T^{-j}(T^j E_A)) = \mu(T^j E_A)$ for all $j = 1, \dots, S$, and therefore $\mu\left(\bigcup_{j=0}^S T^j E_A\right) = \sum_{j=0}^S \mu(T^j E_A) \geq (S+1)\mu(E_A)$ which implies $\mu(E_A) \leq \frac{1}{S+1} \leq e^{-nb}$. Now put

$$\mathcal{B}_n = \left\{A \in \mathcal{A}^n : \mu(A) \geq e^{-nc}\right\}$$

[1]Suppose R^+ were not constant a.s., then the set $E = \{x \in \Omega : R^+(Tx) < R^+(x)\}$ has positive measure and there exists an $a \in \mathbb{R}$ so that $E_a = \{x \in \Omega : R^+(Tx) < a < R^+(x)\}$ also has positive measure. By the Poincaré recurrence theorem almost every $x \in E_a$ returns to E_a after finite time. Since $R^+(T^j x) \leq R^+(Tx) < a \forall j \geq 1$ this contradicts the fact that $R^+(x) > a \forall x \in E_a$.

which by the drawer principle implies $|\mathcal{B}_n| \leq e^{nc}$. By the theorem of Shannon-McMillan-Breiman $\mathbb{P}(x \in \Omega : x \notin B_n \text{ i.o.}) = 0$, where $B_n = \bigcup_{A \in \mathcal{B}_n} A$. If

$$G_n = \bigcup_{A \in \mathcal{B}_n} E_A$$

then

$$\mu(G_n) \leq |\mathcal{B}_n| \max_{A \in \mathcal{A}^n} \mu(E_A) \leq e^{nc} e^{-nb} = e^{-(b-c)}$$

decays exponentially fast as $b - c > 0$. Thus $\sum_n \mu(G_n) < \infty$ and by Borel-Cantelli $\mathbb{P}(x \in \Omega : x \in G_n \text{ i.o.}) = 0$. Moreover, if we put $H_n = \bigcup_{A \in \mathcal{A}^n} E_A = \{x \in \Omega : R_n(x) \geq e^{bn}\}$, then

$$\mathbb{P}(x \in \Omega : x \in H_n \text{ i.o.}) \leq \mathbb{P}(x \in \Omega : x \in G_n \text{ i.o.}) + \mathbb{P}(x \in \Omega : x \in B_n^c \text{ i.o.}) = 0$$

as $G_n = H_n \cap B_n$ and $H_n \cap B_n^c \subset B_n^c$. Hence $R^+ \leq h$ almost surely.

(II) Now suppose $R^- < h$ and let b, c be so that $R^- < b < c < h$. For $N \in \mathbb{N}$ put

$$D_N = \left\{x \in \Omega : R_n(x) \leq e^{nb} \text{ for some } n \in [1, N]\right\}.$$

Obviously $\mu(D_N) \to 1^-$ as $N \to \infty$. Let $\varepsilon > 0$, then $\mu(D_N) > 1 - \frac{\varepsilon}{2}$ for all N large enough. Define

$$E_L = \left\{x \in \Omega : \frac{1}{L} \sum_{j=0}^{L-1} \chi_{D_N}(T^j x) > 1 - \varepsilon\right\}.$$

By the pointwise ergodic theorem $\mu(E_L) \to 1^-$ as $L \to \infty$.

Now one does a parsing argument to estimate the exponential growth rate of

$$\mathcal{A}^{(L)} = \left\{A \in \mathcal{A}^L : E_L \cap A \neq \emptyset\right\}.$$

A cylinder $A \in \mathcal{A}^{(L)}$ is given by an L-word $x_0 x_1 \cdots x_{L-1}$ (i.e. $A = U(x_0 x_1 \cdots x_{L-1})$, where $x_j \in \{1, \ldots, M\}$ and $M = |\mathcal{A}|$). The word is parsed in the following way and then we estimate the number of ways such an L-word can be composed in order to lie in $\mathcal{A}^{(L)}$.

(A) The first block: If $x \notin D_N$ then the first block is the single-element block x_0. If $x \in D_N$ then the first block is $x_0 x_1 \cdots x_{m-1}$ where $m \in [1, N]$ is so that $R_m(x) \leq e^{mb}$.

(B) Recursively: Assume the string $x_0 x_1 \cdots x_{k-1}$, $k \leq L - N$, has been parsed, then the next block is x_k if $T^k x \notin D_N$ and it is $x_k x_{k+1} \cdots x_{k+m-1}$ if $T^k x \in D_N$ where $m \leq N$ is so that $R_m(x) \leq e^{mb}$.

(C) The remaining symbols of which there are at most N many will not be parsed.

Now we estimate the number of possibilities for such L-words.

(A) There are at most εL single-element blocks. Each can be filled with any of the M symbols of $\mathcal{A}$. The number C_0 of configurations of these at most εL many single-element blocks is bounded by

$$C_0 \leq \sum_{j=0}^{[\varepsilon L]} \binom{L}{j} \leq \varepsilon L \binom{L}{[\varepsilon L]} \leq c_1 \sqrt{L} \frac{\sqrt{\varepsilon}}{\sqrt{1-\varepsilon}} \left(\frac{1}{(1-\varepsilon)^{1-\varepsilon} \varepsilon^{\varepsilon}}\right)^L$$

for some constant c_1, using Stirling's formula $j! \sim \sqrt{2\pi j} j^j e^{-j}$. The total combinatorial contribution from the single-element blocks amounts to a factor of (ε is small)

$$C_A \le c_1 M^{\varepsilon L} \left(\frac{1}{(1-\varepsilon)^{1-\varepsilon} \varepsilon^\varepsilon} \right)^L.$$

(B) Denote the length of the 'long' (i.e. not single-element) blocks by m_j, where $j = 1, 2, \dots, r$, r being the total number of non-single-element blocks. By construction $m_j \le N$. Since the return time R_{m_j} is bounded by $e^{m_j b}$, the jth word (which has length m_j) can be chosen in at most $e^{m_j b}$ many different ways because it is repeated after no more than $e^{m_j b}$ time to its 'right'. The total number of ways to fill all r words is

$$C_B \le \prod_{j=1}^{r} e^{m_j b} = e^{\sum_{j=1}^r m_j b} \le e^{bL}.$$

(C) The remaining symbols between the last 'long' word that begins on a coordinate $< L - N$ and ends on coordinate $L - 1$, of which there are no more than N many, can be filled in at most

$$C_C \le M^N$$

many different ways.

The three estimates combined give

$$|\mathcal{A}^{(L)}| \le C_A C_B C_C \le c_1 M^{N+\varepsilon L} e^{bL} \left(\frac{1}{(1-\varepsilon)^{1-\varepsilon} \varepsilon^\varepsilon} \right)^L$$

or, taking logarithms,

$$\frac{\log |\mathcal{A}^{(L)}|}{L} \le b + \left| \log(1-\varepsilon)^{1-\varepsilon} \varepsilon^\varepsilon \right| + \varepsilon \log M + \frac{1}{L} \left(c_1 + N \log M \right).$$

As $L \to \infty$ the last term vanishes and since $\lim_{\varepsilon \to 0^+} (1-\varepsilon)^{1-\varepsilon} \varepsilon^\varepsilon = 1$ we get that for all $\varepsilon > 0$ small enough $|\log(1-\varepsilon)^{1-\varepsilon}\varepsilon^\varepsilon| < \frac{b-c}{8}$ and $\varepsilon \log M < \frac{b-c}{8}$. The value of ε determines the choice of N. Hence for all L large enough

$$\frac{\log |\mathcal{A}^{(L)}|}{L} < b + \frac{b-c}{2} = \frac{b+c}{2}.$$

If

$$\mathcal{B}_L = \left\{ A \in \mathcal{A}^L : \mu(A) \le e^{-Lc} \right\},$$

then by the Shannon-McMillan-Breiman theorem $\mathbb{P}(x \in \Omega : x \notin B_L \text{ i.o.}) = 0$ where $B_L = \bigcup_{A \in \mathcal{B}_L} A$. If we put $\mathcal{G}_L = \mathcal{B}_L \cap \mathcal{A}^{(L)}$ and $G_L = \bigcup_{A \in \mathcal{G}_L} A$ then

$$\mu(G_L) \le |\mathcal{A}^{(L)}| \cdot e^{-Lc} \le e^{L\frac{b+c}{2}} e^{-Lc} = e^{-L\frac{c-b}{2}}$$

which goes to zero exponentially fast as $c > b$. Thus $\sum_L \mu(G_L) < \infty$ and by Borel-Cantelli $\mathbb{P}(x \in \Omega : x \in G_L \text{ i.o.}) = 0$. Since also $\mathbb{P}(x \in \Omega : x \in G_L \setminus B_L \text{ i.o.}) \le \mathbb{P}(x \in \Omega : x \notin B_L \text{ i.o.}) = 0$ we conclude that $R^- \ge h$.

This forces $R^+ = R^- = h$. ∎

This theorem was by Ornstein and Weiss [25] later extended to infinite partitions (alphabets).

Theorem 50. *(Ornstein 1969) Let (Σ_1, μ_1), (Σ_2, μ_2) be two Bernoulli shifts. Then $h(\mu_1) = h(\mu_2)$ if and only if the two Bernoulli shifts are isomorphic (i.e. there exists a measure preserving and invertible map $\varphi : \Sigma_1 \to \Sigma_2$).*

There are several proofs of this famous theorem all of which are very long (see e.g. [**?**, **?**]). We will skip the proof and just give an example. Let $\Sigma_1 = \{1, 2, 3, 4\}^{\mathbb{Z}}$ and μ_1 given by the probability vector $(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$. Let $\Sigma_2 = \{1, 2, 3, 4, 5\}^{\mathbb{Z}}$ and μ_2 be given by $(\frac{1}{2}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8})$. Then $h(\mu_1) = 4\frac{1}{4}|\log\frac{1}{4}| = 2\log 2$ and $h(\mu_2) = \frac{1}{2}|\log\frac{1}{2}| + 4\frac{1}{8}|\log\frac{1}{8}| = 2\log 2$ are equal. By the isomorphism theorem the two Bernoulli shifts are isomorphic. Meschalkin gave a nice description of how such an isomorphism can be constructed.

We give a short description of the Lempel-Ziv compression algorithm. Assume $\Sigma = \{0, 1\}^{\mathbb{N}}$ and let $\vec{x} \in \Sigma$, then $\vec{x} = x_0 x_1 x_2 \cdots$ is parsed into words $w_1, w_2, \ldots$ in the following way. One puts $w_1 = x_1$ for the first word and then defines recursively $w_j = w_{k_j} x_{\ell_j}$ as the word which has not been seen previously and which reduces to a word w_{k_j} which has already been seen if its last symbol x_{ℓ_j} is removed. Clearly, $\ell_j = \sum_{i=1}^{j} |w_i|$, where $|w_i|$ denotes the length of the ith word w_i. In the new description, the word w_j contains two pieces of information, namely the position k_j of its known portion w_{k_j} (which here requires $[\log_2 j] + 1$ many binary digits) and the new additional symbol x_{ℓ_j} (which requires a single binary digit). This algorithm is known to be optimal in the limit (see e.g. [**?**]) and is in its various implementations widely used in practice.

8. Pressure and topological entropy

8.1. **Pressure.** In this section we assume that Ω is a compact metric space with a metric $d(\cdot, \cdot)$ which induces the σ-algebra. We denote be $C(\Omega)$ the set of continuous functions on Ω. In section 6 we have shown that the probability measures form a compact set in the weak* topology. For a continuous map $T : \Omega \to \Omega$, the set of invariant measures $\mathscr{M}(T)$ is not empty (Theorem of Krylov and Bogolioubov).

Let $\mathcal{A}$ be a finite partition of Ω and $\mathcal{A}^n$ its nth join. If f is a continuous function on Ω then for sets $A \subset \Omega$ we write

$$f(A) = \sup_{x \in A} f(x).$$

Definition 51. *For a finite partition $\mathcal{A}$ and $f \in C(\Omega)$ we call*

$$Z_n(f, \mathcal{A}) = \sum_{A \in \mathcal{A}} e^{f^n(A)}$$

the nth partition function *of f with respect to $\mathcal{A}$.*

Lemma 52. *The limit*

$$P(f, \mathcal{A}) = \lim_{n \to \infty} \frac{1}{n} \log Z_n(f, \mathcal{A})$$

exists and is called the pressure *of f with respect to $\mathcal{A}$.*

Proof. We have to prove the existence of the limit which we do by establishing the subadditivity of the sequence $a_n = \log Z_n(f, \mathcal{A})$. Let us recall that $\mathcal{A}^{n+m} = \mathcal{A}^n \vee T^{-n}\mathcal{A}^m$ which implies that for every $C \in \mathcal{A}^{n+m}$ there are $A \in \mathcal{A}^n, B \in \mathcal{A}^m$ so that $C = A \cap T^{-n} B$.

Since $f^{n+m}(A\cap T^{-n}B)\le f^n(A)+f^m(B)$ we conclude that the Z_n are submultiplicative: as

$$\begin{aligned}
Z_{n+m}(f,\mathcal{A}) &= \sum_{C\in\mathcal{A}^{n+m}} e^{f^{n+m}(C)}\\
&= \sum_{A\in\mathcal{A}^n}\sum_{B\in\mathcal{A}^m} e^{f^{n+m}(A\cap T^{-n}B)}\\
&\le \sum_{A\in\mathcal{A}^n}\sum_{B\in\mathcal{A}^m} e^{f^n(A)+f^m(B)}\\
&= \sum_{A\in\mathcal{A}^n} e^{f^n(A)}\sum_{B\in\mathcal{A}^m} e^{f^m(B)}\\
&= Z_n(f,\mathcal{A})Z_m(f,\mathcal{A})
\end{aligned}$$

and by taking logarithms

$$a_{n+m}=\log Z_{n+m}\le\log Z_n+\log Z_m=a_n+a_m$$

which shows subadditivity. The limit is then ensured by the arithmetic lemma from section 7:

$$\lim_{n\to\infty}\frac{1}{n}\log Z_n(f,\mathcal{A})=\inf_{n\in\mathbb{N}}\frac{1}{n}\log Z_n(f,\mathcal{A}).$$

∎

Remark. As $f\ge-|f|_\infty$ we get $Z_n\ge e^{-n|f|_\infty}$ which implies $P(f)\ge-|f|_\infty$. We also have $Z_n\le e^{n|f|_\infty}|\mathcal{A}^n|\le e^{n|f|_\infty}|\mathcal{A}|^n$ which yields $P(f)\le|f|_\infty+\log|\mathcal{A}|$.

Definition 53. *$P(f)=\sup_{\{\mathcal{A}_k\}}\limsup_{k\to\infty}P(f,\mathcal{A}_k)$ is called the* pressure *of f, where the supremum is over all sequences of partitions $\mathcal{A}_k$ for which* $\text{diam}\mathcal{A}_k\to0^+$ *as $k\to\infty$.*

($\text{diam}\mathcal{A}=\max_{A\in\mathcal{A}}\text{diam}(A)$)

Lemma 54. *If $\{\mathcal{A}_k:k\}$ is a sequence of partitions whose diameters go to zero, then $P(f)=\lim_k P(f,\mathcal{A}_k)$.*

Proof. If $\{\mathcal{B}_n:n\}$ is any sequence of partitions (so that $\text{diam}\mathcal{B}_k\to0$) then we have to show that every $\varepsilon>0$

$$\limsup_n P(f,\mathcal{B}_n)\le\limsup_k P(f,\mathcal{A}_k)+\varepsilon,$$

or for every $P(f,\mathcal{B}_n)\le P(f,\mathcal{A}_k)+\varepsilon$ for all k large enough. Let $\varepsilon>0$ and since f is continuous on compact Ω there exists $\delta>0$ so that $|f(x)-f(y)|<\frac{\varepsilon}{2}$ for all $x,y\in\Omega$ for which $d(x,y)<\delta$. We let n be big enough so that $\text{diam}\mathcal{B}_n<\frac{\delta}{2}$ and also let $\delta'>0$ ($\delta'<\delta/2$) be small enough so that every $B\in\mathcal{B}_n$ contains a δ'-ball. Now choose k large enough so that $\text{diam}\mathcal{A}_k<\delta'/2$. Then for every $B\in\mathcal{B}_n$ we can find an $A_B\in\mathcal{A}_k$ so that $A_B\subset B$ and $A_B\ne A_{B'}$ if $B\ne B'$. Similarly, for every $B\in\mathcal{B}_n^m$ there exists $A_B\in\mathcal{A}_k^m$ so that $A_B\subset B$. We can now estimate the ergodic sums for $B\in\mathcal{B}_n^m$

$$|f^m(B)-f^n(x)|<m\frac{\varepsilon}{2}$$

for any point $x \in B$. Similarly we get

$$|f^m(A_B) - f^m(x)| \le m\frac{\varepsilon}{2}$$

where this time $x \in A_B \subset B$. Hence

$$|f^m(B) - f^m(A_B)| < m\varepsilon$$

for all $B \in \mathcal{B}_n^m$. Thus

$$\begin{aligned} Z_m(f, \mathcal{B}_n) &= \sum_{B \in \mathcal{B}_n} e^{f^m(B)} \\ &\le \sum_{B \in \mathcal{B}_n} e^{f^m(A_B)} e^{m\varepsilon} \\ &\le \sum_{B \in \mathcal{A}_k} e^{f^m(A)} e^{m\varepsilon} \\ &= Z_m(f, \mathcal{A}_k) e^{m\varepsilon} \end{aligned}$$

which implies that $P(f, \mathcal{B}_n) \le P(f, \mathcal{A}_k) + \varepsilon$ and since $\varepsilon > 0$ was arbitrary we obtain

$$\limsup_n P(f, \mathcal{B}_n) \le \limsup_k P(f, \mathcal{A}_k).$$

Since $\{\mathcal{B}_n : n\}$ was arbitrary we get $P(f) = \limsup_k P(f, \mathcal{A}_k)$. ∎

Definition 55. *The (finite) partition $\mathcal{A}$ is a* topological generator *if* diam$\mathcal{A}^n \to 0$ *as* $n \to \infty$.

Theorem 56. *If $\mathcal{A}$ is a topological generator then $P(f) = P(f, \mathcal{A})$.*

Proof. We will show that $P(f, \mathcal{A}^k) = P(f, \mathcal{A})$ for every k. The theorem then follows from the previous lemma. Since

$$(\mathcal{A}^k)^n = \bigvee_{j=0}^{n-1} T^{-j} \bigvee_{i=0}^{k-1} T^{-i}\mathcal{A} = \bigvee_{j=0}^{n-1}\bigvee_{i=0}^{k-1} T^{-j-i}\mathcal{A} = \bigvee_{j=0}^{k+n-2} T^{-j}\mathcal{A} = \mathcal{A}^{n+k-1}$$

we get

$$Z_n(f, \mathcal{A}^k) = \sum_{A \in \mathcal{A}^{n+k-1}} e^{f^n(A)} = \sum_{A' \in \mathcal{A}^n} \sum_{A \in \mathcal{A}^{n+k-1}, A \subset A'} e^{f^n(A)} \le \sum_{A' \in \mathcal{A}^n} e^{f^n(A')} N_{A'}$$

and thus $Z_n(f, \mathcal{A}^k) \le |\mathcal{A}|^k Z_n(f, \mathcal{A})$ where $N_{A'} = |\{A \in \mathcal{A}^{n+k-1} : A \subset A'\}| \le |\mathcal{A}|^k$. For the lower bound we obtain

$$Z_n(f, \mathcal{A}^k) \ge \sum_{A' \in \mathcal{A}^n} \sum_{A \in \mathcal{A}^{n+k-1}, A \subset A'} e^{f^n(A)} \ge \sum_{A' \in \mathcal{A}^n} e^{f^n(A')} = Z_n(f, \mathcal{A})$$

since for every $A' \in \mathcal{A}^n$ there is an $A \subset A'$, $A \in \mathcal{A}^{n+k-1}$, so that $f^n(A) = f^n(A')$ (realises the sup). Taking logarithms and dividing by n yields

$$\frac{1}{n}\log Z_n(f, \mathcal{A}) \le \frac{1}{n}\log Z_n(f, \mathcal{A}^k) \le \frac{1}{n} k \log M + \frac{1}{n}\log Z_n(f, \mathcal{A})$$

and taking limits yields

$$P(f, \mathcal{A}^k) = P(f, \mathcal{A})$$

for any k. The theorem now follows from the previous lemma since $\text{diam}\mathcal{A}^k \to 0$ as $\mathcal{A}$ is generating. ∎

Remarks. (I) $P(f + g \circ T - g) = P(f)$. This follows from the fact that

$$(f + g \circ T - g)^n = \sum_{j=0}^{n-1} (f \circ T^j + g \circ T^{j+1} - g \circ T^j) = f^n + g \circ T^n - g$$

which implies

$$Z_n(f + g \circ T - g, \mathcal{A}) \leq \sum_{A \in \mathcal{A}^n} e^{f^n(A) + g(T^n A) - g(A)} = Z_n(f, \mathcal{A}) e^{\mathcal{O}(2|g|_\infty)}.$$

Taking logarithms and dividing by n yields

$$\frac{1}{n} \log Z_n(f + g \circ T - g, \mathcal{A}) = \frac{1}{n} Z_n(f, \mathcal{A}) + \mathcal{O}\left(\frac{2|g|_\infty}{n}\right)$$

which gives the result when $n \to \infty$. The function $g \circ T - g$ is a *coboundary.*
(II) $P(f + c) = P(f) + c$ for constants c. This follows from

$$Z_n(f + c, \mathcal{A}) = \sum_{A \in \mathcal{A}^n} e^{f^n(A) + nc} = Z_n(f, \mathcal{A}) e^{nc}.$$

Taking logarithms, dividing by n and letting n go to infinity makes the statement follow.
(III) In the special case when $f = 0$

$$P(0) = \lim_{n \to \infty} \frac{1}{n} \log |\mathcal{A}^n| = h_{\text{top}}$$

is called the *topological entropy* and captures the exponential growth rate of the joins of $\mathcal{A}$.

8.2. **Variational principle.**

Theorem 57. *(Variational principle, Walters 1973)*

$$P(f) = \sup_\mu \left(h(\mu) + \int_\Omega f \, d\mu \right)$$

where the supremum is over all T-invariant probability measures μ ($h(\mu)$ is the measure theoretic entropy of μ).

In the special case when $f = 0$ then we get the topological entropy is $h_{\text{top}} = \sup_\mu h(\mu)$ where again the supremum is over all invariant probability measures.

Definition 58. *(I) If $h(\mu) + \mu(f) = P(f)$ then μ is an* equilibrium state.
(II) If $h(\mu) = h_{top}$ then μ is a measure of maximal entropy.

Bernoulli shift: Before we prove the theorem, let us do the Bernoulli shift as an example. Let $\Sigma = \{1, \ldots, M\}^{\mathbb{Z}}$, σ the left shift and $\mathcal{A} = \{U(i) : i\}$ the standard generating partition.

Let f be a function that only depends on the zeroth coordinate, i.e. $f(\vec{x}) = f(x_0) = f_{x_0}$. Since the nth partition function is

$$Z_n = \sum_{x_1\cdots x_n} e^{f^n(x_1\cdots x_n)} = \sum_{x_1\cdots x_n} e^{f_{x_1}+\cdots+f_{x_n}} = \left(\sum_i e^{f_i}\right) = Z_1^n$$

we see that the pressure is $P(f) = \log Z_1 = \log\sum_i e^{f_i}$ ($Z_1 = \sum_i e^{f_i}$). We now look for an equilibrium state μ which will be a Bernoulli measure induced by a probability vector $(p_1, \ldots, p_M)$. With a Lagrange multiplier λ the derivatives

$$\begin{aligned}\frac{\partial}{\partial p_i}\left(h(\mu) + \int f\,d\mu + \lambda\sum_j p_j\right) &= \frac{\partial}{\partial p_i}\left(-\sum_j p_j\log p_j + \sum_j f_j p_j + \lambda\sum_j p_j\right)\\ &= -\log p_i - 1 + f_i + \lambda\end{aligned}$$

have to be zero for every i, which yields $\log p_i = f_i + \lambda - 1$ or $p_i = e^{f_i}e^{\lambda-1}$. The normalisation condition is $1 = \sum_i p_i = \sum_i e^{f_i}e^{\lambda-1} = Z_1 e^{\lambda-1}$, where $Z_1 = \sum_i e^{f_i} = e^{1-\lambda}$. Hence the probabilities $p_i = \frac{e^{f_i}}{Z_1}$ define the invariant measure μ. We can verify

$$\begin{aligned}h(\mu) &= -\sum_i p_i\log p_i\\ &= -\sum_i \frac{e^{f_i}}{Z_1}\log\frac{e^{f_i}}{Z_1}\\ &= -\frac{1}{Z_1}\sum_i e^{f_i}(f_i - \log Z_1)\\ &= -\sum_i f_i p_i + \log Z_1\\ &= P(f) - \int f\,d\mu\end{aligned}$$

which means that μ is indeed an equilibrium state for f.

Let us now introduce a parameter $t > 0$ and denote by μ_{tf} the equilibrium state for the function tf. With $Z_1(tf) = \sum_i e^{tf_i}$ we obtain

$$\frac{d}{dt}\log Z_1(tf) = \frac{1}{Z_1(tf)}\frac{d}{dt}\sum_i e^{tf_i} = \frac{1}{Z_1(tf)}\sum_i f_i e^{tf_i} = \mu_{tf}(f).$$

One also has $h(\mu_{tf}) = \log Z_1(tf) - \mu_{tf}(tf)$ which, when differentiated, yields:

$$\frac{d}{dt}h(\mu_{tf}) = \mu_{tf}(f) - \frac{d}{dt}t\mu_{tf}(f) = \mu_{tf}(f) - \mu_{tf}(f) - t\frac{d}{dt}\mu_{tf}(f)$$

and thus

$$\frac{d}{dt}h(\mu_{tf}) = t\frac{d}{dt}\mu_{tf}(-f).$$

If we interpret t as the inverse temperature $\frac{1}{T}$ with T being the absolute temperature, then this last identity looks familiar to the thermodynamic relation

$$dS = \frac{1}{T}dQ$$

where S is the thermodynamic entropy and dQ is the heat energy.

For the proof of the variational principle we will need the following arithmetic lemma.

Lemma 59. *Let p_i be weights ($\sum_i p_i = 1$). Then for any $a_i \in \mathbb{R}^+$ one has*

$$\sum_i p_i(a_i - \log p_i) \leq \log \sum_i e^{a_i}$$

with equality if and only if $p_i = \frac{e^{a_i}}{Z}$, where $Z = \sum_i e^{a_i}$.

In the special case when $a_i = 0, i = 1, \ldots, M$, then $-\sum_i p_i \log p_i \leq \log M$ with equality if and only if $p_i = \frac{1}{M}$ for all i.

Proof. Recall that the function $\varphi(t) = -t\log t$ for $t > 0$ and $\varphi(0) = 0$ is concave down on $[0, \infty)$, i.e. if q_i are weights and $b_i \geq 0$ arbitrary, then $\sum_i q_i\varphi(b_i) \leq \varphi\left(\sum_i q_i b_i\right)$ and equality if and only if all the b_i are equal. Now put $q_i = \frac{e^{a_i}}{Z}$, where $Z = \sum_i e^{a_i}$, and $b_i = \frac{p_i Z}{e^{a_i}} = \frac{p_i}{q_i}$. Clearly $q_i > 0$ and $\sum_i q_i = \sum_i \frac{e^{a_i}}{Z} = \frac{Z}{Z} = 1$, i.e. the q_i are weights. Since

$$\sum_i q_i b_i = \sum_i \frac{p_i}{q_i} q_i = \sum_i p_i = 1$$

one has by concavity

$$0 = \varphi(1) = \varphi\left(\sum_i q_i b_i\right) \geq \sum_i q_i \varphi(b_i)$$

which implies

$$0 \geq \sum_i q_i \left(-\frac{p_i Z}{e^{a_i}}\right) \log \frac{p_i Z}{e^{a_i}} = -\sum_i \frac{e^{a_i}}{p_i Z}\frac{p_i Z}{e^{a_i}}\left(\log p_i + \log Z - a_i\right)$$

and finally gives

$$\log Z \geq \sum_i p_i(a_i - \log p_i).$$

Note that above in the concavity argument one gets equality if and only if the $b_i = p_i\frac{Z}{e^{a_i}}$ are all equal. Suppose $b_i = \alpha \forall i$ then $p_i = \alpha\frac{e^{a_i}}{Z}$ and $1 = \sum_i p_i = \sum_i \alpha\frac{e^{a_i}}{Z} = \alpha$ implies that $p_i = \frac{e^{a_i}}{Z}\forall i$. ∎

Lemma 60. *Let $\mu, \nu \in \mathscr{M}(T)$, $p + q = 1, p, q \geq 0$. Then $pH(\mu, \mathcal{A}^n) + qH(\nu, \mathcal{A}^n) \leq H(p\mu + q\nu, \mathcal{A}^n)$ and $ph(\mu) + qh(\nu) \leq h(p\mu + q\nu)$.*

Proof. By concavity of the function $\varphi(t) = -t\log t$ ($\varphi(0) = 0$) one has

$$\begin{aligned} H(p\mu + q\nu, \mathcal{A}^n) &= \sum_{A\in\mathcal{A}^n} \varphi(p\mu(A) + q\nu(A)) \\ &\geq \sum_{A\in\mathcal{A}^n} \left(p\varphi(\mu(A)) + q\varphi(\nu(A))\right) \\ &= pH(\mu, \mathcal{A}^n) + qH(\nu, \mathcal{A}^n). \end{aligned}$$

For the second statement divide by n and take the limit $n \to \infty$. ∎

Similarly one gets for a larger (even countably infinite) number of measures $\mu_j \in \mathscr{M}(T)$ and weights $p_j \geq 0$ ($\sum_j p_j = 1$) that $\sum_j p_j H(\mu_j, \mathcal{A}^n) \leq H(\sum_j p_j \mu_j, \mathcal{A}^n)$ and $\sum_j p_j h(\mu_j) \leq h(\sum_j p_j \mu_j)$.

Proof of the variational principle. We prove the theorem in two parts, first the lower bound on the pressure and then the upper bound on the pressure. Let $\mathcal{A}$ be a topological generator.

Part (I): We show that $P(f) \geq h(\mu) + \int_\Omega f\, d\mu$ for every invariant measure μ. Let μ be an invariant measure, then $\int f^n\, d\mu = n \int f\, d\mu$ as $\int f \circ T^j\, d\mu = \int f\, d\mu$ and therefore

$$\begin{aligned}
\frac{1}{n} H(\mathcal{A}^n) + \int f\, d\mu &= \frac{1}{n}\left(H(\mathcal{A}^n) + \int f^n\, d\mu\right)\\
&= \frac{1}{n}\sum_{A\in\mathcal{A}^n}\left(-\mu(A)\log\mu(A) + \int_A f^n\, d\mu\right)\\
&\leq \frac{1}{n}\sum_{A\in\mathcal{A}^n}\mu(A)\left(-\log\mu(A) + f^n(A)\right).
\end{aligned}$$

The arithmetic lemma with the weights $p_A = \mu(A)$ and the numbers $a_A = f^n(A)$, $A \in \mathcal{A}^n$ yields

$$\frac{1}{n} H(\mathcal{A}^n) + \int f\, d\mu \leq \frac{1}{n}\log\sum_{A\in\mathcal{A}^n} e^{f^n(A)}.$$

and letting $n \to \infty$ one gets $h(\mu) + \int f\, d\mu \leq P(f)$.

Part (II): Here we produce an invariant measure μ for which $P(f) \leq h(\mu) + \int f\, d\mu$. For every $A \in \mathcal{A}^n$ pick an arbitrary point $y_A \in A$ and put

$$\nu_n = \frac{1}{\hat{Z}_n}\sum_{A\in\mathcal{A}^n} e^{f^n(y_A)}\delta_{y_A}$$

where δ_{y_A} is the point mass at y_A and $\hat{Z}_n = \sum_{A\in\mathcal{A}^n} e^{f^n(y_A)}$ is the normalising term. Now we define

$$\mu_n = \frac{1}{n}\sum_{j=0}^{n-1}\nu_n \circ T^j.$$

By the arithmetic lemma we obtain for the measure ν_n

$$\frac{1}{n}\left(H(\nu_n, \mathcal{A}^n) + \int f^n\, d\nu_n\right) = \frac{1}{n}\sum_{A\in\mathcal{A}^n}\nu_n(A)\left(-\log\nu_n(A) + f^n(y_A)\right) = \frac{1}{n}\hat{Z}_n,$$

where we used the weights $p_A = \nu_n(A)$ and the values $a_A = e^{f^n(y_A)}$ and got equality because $p_A = e^{f^n(y_A)}/\hat{Z}_n$. To compare $\hat{Z}_n$ to the partition function Z_n let us note that for $\varepsilon > 0$ there exists $\delta > 0$ so that $|f(x) - f(y)| < \varepsilon$ if $d(x,y) < \delta$. As $\mathcal{A}$ is a generator, we can assume that $\operatorname{diam}\mathcal{A} < \delta$ because otherwise we replace $\mathcal{A}$ by a join $\mathcal{A}^k$ which has small enough diameter. Since $\sup_{A\in\mathcal{A}}\sup_{x,y\in A}|f(x) - f(y)| < \varepsilon$ we get $\hat{Z}_n = Z_n e^{\mathcal{O}(n\varepsilon)}$ and therefore $|\log\hat{Z}_n - \log Z_n| < n\varepsilon$.

In order to replace ν_n in the estimates above by μ_n we use a summation trick due to Misurewicz. Let $m > 0$ ($m \ll n$) and put $p(j) = \left[\frac{n-j}{m}\right]$ for $j = 0, 1, \ldots, m$. Then one can write

$$\mathcal{A}^n = \left(\bigvee_{k=0}^{p(j)-1} T^{-(km+j)} \mathcal{A}^m \right) \vee \bigvee_{k \in R_j} T^{-j} \mathcal{A},$$

where $R_j = \{0, 1, 2, \ldots, j-1\} \cup \{p(j)m + j + 1, p(j)m + j + 2, \ldots, n\}$ is the remainder set ($|R_j| \le 2m$). Then

$$H(\nu_n, \mathcal{A}^n) \le \sum_{k=0}^{p(j)-1} H\left(\nu_n, T^{-(km+j)}\mathcal{A}^m\right) + H\left(\bigvee_{k \in R_j} T^{-j}\mathcal{A} \right),$$

where the last term on the RHS is estimated above by $\log \left| \bigvee_{k \in R_j} T^{-j}\mathcal{A} \right| \le \log |\mathcal{A}|^{|R_j|} \le 2m \log |\mathcal{A}|$. Summation over $j = 0, 1, \ldots, m-1$ yields

$$mH(\nu_n, \mathcal{A}^n) \le \sum_{j=0}^{m-1} \sum_{k=0}^{p(j)-1} H\left(\nu_n, T^{-(km+j)}\mathcal{A}^m\right) + \mathcal{O}(m^2)$$

By convexity of the entropy function

$$\begin{aligned} \frac{m}{n} H(\nu_n, \mathcal{A}^n) &\le H\left(\frac{1}{n} \sum_{j=0}^{m-1} \sum_{k=0}^{p(j)-1} T^{km+j}\nu_n, \mathcal{A}^m \right) + \mathcal{O}\left(\frac{m^2}{n}\right) \\ &= H\left(\frac{1}{n} \sum_{i=0}^{n-1} T^i \nu_n, \mathcal{A}^m \right) + \mathcal{O}\left(\frac{m^2}{n}\right) \\ &= H\left(\mu_n, \mathcal{A}^m\right) + \mathcal{O}\left(\frac{m^2}{n}\right). \end{aligned}$$

Since $\frac{1}{n}(H(\nu_n, \mathcal{A}^n) + \int f^n \, d\nu_n) = \log \hat{Z}_n = \log Z_n + \mathcal{O}(\varepsilon)$ one has

$$\frac{1}{n} \log Z_n \le \frac{1}{n} \left(H(\nu_n, \mathcal{A}^n) + \int f^n \, d\nu_n \right) + \varepsilon \le \frac{1}{m} H(\mu_n, \mathcal{A}^m) + \int f \, d\mu_n + \varepsilon + \mathcal{O}\left(\frac{m^2}{n}\right)$$

as $\frac{1}{n} \int f^n \, d\nu_n = \int f \, d\mu_n$. Let $\mu_{n_j} \to \mu$ in the weak* topology along a convergent subsequence n_j and one obtains

$$P(f, \mathcal{A}) \le \frac{1}{m} H(\mu, \mathcal{A}^m) + \int f \, d\mu + \varepsilon.$$

Letting $m \to \infty$ yields

$$P(f, \mathcal{A}) \le h(\mu) + \int f \, d\mu + \varepsilon$$

as $\mathcal{A}$ is a generator. The result follows as $\varepsilon > 0$ was arbitrary. The limiting measure is an equilibrium state for f as is any other accumulation point of the sequence $\{\mu_n : n\}$. ∎

Remark: Oftentimes the pressure is introduced using separated and spanning sets. We say a set $E_{\varepsilon,n} \subset \Omega$ is (ε, n)-*separated* if for every $x, y \in E_{\varepsilon,n}$, $x \ne y$ one has $d(T^j x, T^j y) \ge \varepsilon$

for $j = 0, 1, \ldots, n-1$. Similarly, a set $F_{\varepsilon,n} \subset \Omega$ is (ε, n)-*spanning* if for every $x \in \Omega$ there exists a $y \in F_{\varepsilon,n}$ so that $d(T^j x, T^j y) < \varepsilon$ for $j = 0, 1, \ldots, n-1$. Equivalently one can say that $F_{\varepsilon,n}$ is an (ε, n)-spanning set if $\Omega = \bigcup_{x \in F_{\varepsilon,n}} B_{\varepsilon,n}(x)$, where $B_{\varepsilon,n}(x) = \{y \in \Omega : d_n(x, y) < \varepsilon\}$ is an (ε, n)-*Bowen ball* given by the metric $d_n(x, y) = \max_{0 \le j < n} d(T^j x, T^j y)$.

One then defines

$$Z_{\varepsilon,n}(f) = \sup_{E_{\varepsilon,n}} \sum_{x \in E_{\varepsilon,n}} e^{f^n(x)}$$

where the supremum is over all (ε, n)-separated sets $E_{\varepsilon,n}$. The pressure is then

$$P(f) = \lim_{\varepsilon \to 0} \lim_{n \to \infty} \frac{1}{n} \log Z_{\varepsilon,n}(f)$$

and the variational principle applies (see [**?**]). In a similar way one can use minimal spanning sets. If the map T is expansive, then the limit $\varepsilon \to 0$ can is achieved if ε is an expansive constant. T is *expansive* if there exists an $\varepsilon > 0$ so that $d(T^j x, t^j y) < \varepsilon \forall j$ implies that $x = y$ (and such ε is then an *expansive constant.*

8.3. **The Parry measure.** We look at the equilibrium states on a subshift of finite type for locally constant functions. The subshift $\Sigma = \{\vec{x} \in \{1, \ldots, M\}^{\mathbb{Z}} : A_{x_i x_{i+1}} = 1 \forall i\}$ are the doubly infinite sequences over an alphabet with M elements and $M \times M$-transition matrix A. We assume that A is irreducible and aperiodic, that is $A^n > 0$ for all large enough n. The usual left shift map is σ and $\mathcal{A} = \{U(i) : i\}$ is a topologically generating partition.

Let $f : \Sigma \to \mathbb{R}$ be locally constant. We can assume that f depends only on the first two coordinates: $f(\vec{x}) = f(x_0, x_1) = f_{x_0 x_1}$. ($f$ is locally constant if it depends on only finitely many coordinates. A recoding can reduce this to only two coordinates.) Put B for the $M \times M$-matrix with the entries $B_{ij} = A_{ij} e^{f_{ij}}$. Clearly B is non-negative, irreducible and aperiodic and thus has by the theorem of Perron-Frobenius a simple largest eigenvalue $\lambda \in (0, \infty)$ and all the other eigenvalues have modulus strictly less than λ. There are also strictly positive left and right eigenvectors $\vec{v}, \vec{w}$ to the dominating eigenvalue: $\vec{v}B = \lambda \vec{v}, B\vec{w} = \lambda \vec{w}$. We can assume that $\sum_i v_i w_i = 1$. Now define a probability vector $\vec{p}$ by putting $p_i = v_i w_i$ and define the $M \times M$-matrix P by $P_{ij} = \frac{1}{\lambda} B_{ij} \frac{w_j}{w_i}$. Then P is a stochastic matrix with left eigenvector $\vec{p}$ as

$$(P\mathbb{1})_i = \sum_j P_{ij} = \frac{1}{\lambda w_i} \sum_j B_{ij} w_j = \frac{1}{\lambda w_i} \lambda w_i = 1$$

for all i, and also

$$(\vec{p}P)_j = \sum_i p_i P_{ij} = \sum_i v_i w_i \frac{1}{\lambda} B_{ij} \frac{w_j}{w_i} = \frac{w_j}{\lambda} \sum_i v_i B_{ij} = \frac{w_j}{\lambda} \lambda v_j = p_j$$

for all j. Thus $(P,\vec{p})$ defines a σ-invariant probability measure μ on Σ which on cylinder sets is given by

$$\begin{aligned}\mu(U(x_0\cdots x_{n-1})) &= p_{x_0}P_{x_0x_1}P_{x_1x_2}\cdots P_{x_{n-2}x_{n-1}}\\ &= \frac{1}{\lambda^{n-1}}v_{x_0}w_{x_0}B_{x_0x_1}\frac{w_{x_1}}{w_{x_0}}B_{x_1x_2}\frac{w_{x_2}}{w_{x_1}}\cdots B_{x_{n-2}x_{n-1}}\frac{w_{x_{n-1}}}{w_{x_{n-2}}}\\ &= \frac{1}{\lambda^{n-1}}v_{x_0}B_{x_0x_1}B_{x_1x_2}\cdots B_{x_{n-2}x_{n-1}}w_{x_{n-1}}\\ &= \frac{1}{\lambda^{n-1}}v_{x_0}e^{f_{x_0x_1}+f_{x_1x_2}+\cdots+f_{x_{n-2}x_{n-1}}}w_{x_{n-1}}\\ &= \frac{1}{\lambda^{n-1}}v_{x_0}e^{f^{n-1}(x_0x_1\cdots x_{n-1})}w_{x_{n-1}}.\end{aligned}$$

By section 7 its entropy is $h(\mu)=-\sum_{ij}p_iP_{ij}\log P_{ij}$ for which we can also write ($P_{ij}=\frac{1}{\lambda}A_{ij}e^{f_{ij}}\frac{w_j}{w_i}$)

$$\begin{aligned}h(\mu) &= -\sum_{ij}p_iP_{ij}\log\frac{1}{\lambda}A_{ij}e^{f_{ij}}\frac{w_j}{w_i}\\ &= \log\lambda-\sum_{ij}p_iP_{ij}f_{ij}-\sum_{ij}p_iP_{ij}\log A_{ij}\frac{w_j}{w_i}\\ &= \log\lambda-\int_\Sigma f\,d\mu-\sum_{ij}p_iP_{ij}\log w_j+\sum_{ij}p_iP_{ij}\log w_i\\ &= \log\lambda-\int_\Sigma f\,d\mu\end{aligned}$$

since $\sum_i p_iP_{ij}=p_j$ and $\sum_j P_{ij}=1$. Or $\log\lambda=h(\mu)+\int f\,d\mu$, where $\log\lambda=\lim_{n\to\infty}\log\|B^n\|$ is the spectral radius of B. Notice that

$$(B^n)_{ij}=\sum_{x_0x_1\cdots x_n}B_{x_0x_1}B_{x_1x_2}\cdots B_{x_{n-1}x_n}$$

where the sum is over all $(n+1)$-words $x_0x_1\cdots x_n$ which begin with $x_0=i$ and end on $x_n=j$. Thus

$$\begin{aligned}\|B^n\|=\sum_{ij}(B^n)_{ij} &= \sum_{x_0x_1\cdots x_n}B_{x_0x_1}B_{x_1x_2}\cdots B_{x_{n-1}x_n}\\ &= \sum_{x_0x_1\cdots x_n}e^{f_{x_0x_1}+f_{x_1x_2}+\cdots+f_{x_{n-1}x_n}}\\ &= \sum_{x_0x_1\cdots x_n}e^{f^n(x_0x_1\cdots x_n)}\\ &= Z_n(f,\mathcal{A}),\end{aligned}$$

where the sum is over all $(n+1)$-words in Σ. We conclude that $\log\lambda=P(f)$ or the leading eigenvalue of B is $\lambda=e^{P(f)}$. The measure μ is called the *Parry measure*.

The special case $f=0$ gives the measure of maximal entropy. Here $B=A$ and P is the stochastic matrix given by $P_{ij}=A_{ij}\frac{w_j}{w_i}$ where $\vec{w}$ is the right eigenvector of A to its leading eigenvalue λ which satisfies

$$\log\lambda = \lim_{n\to\infty}\frac{1}{n}\log\|A^n\|.$$

The entries $(A^n)_{ij}$ count the number of $(n+1)$-words in Σ that begin with the symbol i and end with the symbol j. Thus the norm $\|A^n\|$ is the total number of $(n+1)$-words the space Σ allows ($\|A^n\| = |\mathcal{A}^{n+1}|$). If $\vec{v}$ denotes the left eigenvector for A to the leading eigenvalue λ then, with suitable normalisation, $p_i = v_i w_i, i = 1, \dots, M$, defines a strictly positive probability vector $\vec{p}$. The measure of maximal entropy μ is then defined by $(P, \vec{p})$ on cylinder sets by

$$\mu(U(x_0 x_1 \cdots x_{n-1})) = \frac{1}{\lambda^{n-1}} v_{x_0} w_{x_{n-1}}.$$

Moreover $\log\lambda = h_{\text{top}}$.

Theorem 61. *If the entropy function $\mu \mapsto h(\mu)$ is upper semi continuous, then there exists at least one equilibrium state for $f \in C(\Omega)$.*

Proof. For $\delta > 0$ put

$$S_\delta = \{\mu \in \mathscr{M}(T) : h(\mu) + \mu(f) \geq P(f) - \delta\}.$$

It follows from the variational principle, Theorem 57 $S_\delta \neq \varnothing$. We moreover note that S_δ is convex, since for $\mu, \nu \in \mathscr{M}(T)$ and $p + q = 1, p, q \geq 0$ one has by convexity of the entropy function (Lemma 60)

$$h(p\mu + q\nu) + (p\mu + q\nu)(f) \geq ph(\mu) + qh(\nu) + p\mu(f) + q\nu(f) \geq P(f) - \delta.$$

Hence $p\mu + q\nu \in S_\delta$. Also, S_δ is compact since if $\mu_n \in S_\delta$ is a sequence which converges to μ, then by the upper semi continuity:

$$h(\mu) + \mu(f) \geq \limsup_{n\to\infty}\left(h(\mu_n) + \mu_n(f)\right) \geq P(f) - \delta,$$

and hence $\mu \in S_\delta$. Since the S_δ form a nested sequence we conclude that

$$\{\mu \in \mathscr{M}(T) : h(\mu) + \mu(f) = P(f)\} = \bigcap_{\delta > 0} S_\delta$$

is non-empty. ∎

Assuming upper semi continuity of the entropy function one can prove the dual variational principle

$$h(\mu) = \inf_{f \in C(\Omega)}(P(f) - \mu(f))$$

for every T-invariant probability measure μ.

Lemma 62. *Assume the entropy function h is upper semi continuous. Then the pressure function $P(f)$ is continuous. That is, let $f, g_n \in C(\Omega)$ so that $|g_n|_\infty \to 0$ as $n \to \infty$. Then*
(i) $P(f + g_n) \to P(f)$ as $n \to \infty$.
(ii) If μ_n are equilibrium states for $f + g_n$ and $\mu_n \to \mu$ (weakly) for some μ, then μ is an equilibrium state for f.

Proof. To get continuity of the pressure function we use the variational principle:

$$\begin{aligned} P(f+g_n) &= \sup_{\nu\in\mathscr{M}(T)}(h(\nu)+\nu(f+g_n)) \\ &\geq \sup_{\nu}(h(\nu)+\nu(f))-|g_n|_\infty \\ &= P(f)-|g_n|_\infty. \end{aligned}$$

Similarly one shows the upper bound and obtains

$$|P(f+g_n)-P(f)|\leq |g_n|_\infty\to 0$$

as $n\to\infty$. In order to show that $\mu=\lim_{n\to\infty}\mu_n$ is an equilibrium state for f we need to verify the following lower bound where we use the upper semicontinuity of the entropy function:

$$\begin{aligned} h(\mu)+\mu(f) &\geq \limsup_{n\to\infty}(h(\mu_n)+\mu_n(f)) \\ &= \limsup_{n}(h(\mu_n)+\mu_n(f+g_n)-\mu_n(g_n)) \\ &\geq \limsup_{n}\left((h(\mu_n)+\mu_n(f+g_n))-|g_n|_\infty\right) \\ &\geq \limsup_{n}\left(P(f+g_n)-|g_n|_\infty\right) \\ &\geq \limsup_{n}(P(f)-2|g_n|_\infty)=P(f). \end{aligned}$$

Hence μ is an equilibrium state for f. ∎

Corollary 63. *Assume the entropy function h is USC. Let $f,g_n\in C(\Omega)$ so that $|g_n|_\infty\to 0$ and μ_n the equilibrium states for $f+g_n$.*

If μ is a unique equilibrium state for f, then $\mu=\lim_{n\to\infty}\mu_n$.

Proof. By the previous lemma, a limit point ν of $\{\mu_n : n\}$ is an equilibrium state for f. Hence $\nu=\mu$. ∎

Theorem 64. *Assume the entropy function h is USC and μ is a unique equilibrium state for $f\in C(\Omega)$. For any $g\in C(\Omega)$ one the derivative exits:*

$$\lim_{t\to 0}\frac{P(f+tg)-P(f)}{t}=\mu(g).$$

Proof. Assume $t>0$ and let μ_t be an equilibrium state for $f+tg$. Then

$$\begin{aligned} \mu_t(f+tg)+h(\mu_t) &= P(f+tg) \\ \mu_t(f)+h(\mu) &\leq P(f) \end{aligned}$$

which yields

$$t\mu_t(g)\geq P(f+tg)-P(f).$$

Similarly one obtains from

$$\begin{aligned} \mu(f+tg)+h(\mu) &\leq P(f+tg) \\ \mu(f)+h(\mu) &= P(f) \end{aligned}$$

the bound

$$t\mu(g) \leq P(f+tg) - P(f).$$

Hence

$$\mu(g) \leq \frac{P(f+tg)-P(f)}{t} \leq \mu_t(g)$$

for positive t. Similar estimates from above and below can be obtained for $t<0$. If we let $t \to 0$ then by Lemma 62 and its corollary we obtain $\mu = \lim_{t\to 0}\mu_t$ and therefore

$$\mu(g) \leq \lim_{t\to 0} \frac{P(f+tg)-P(f)}{t} \leq \mu_(g).$$

This proves the derivative and its limit. ▮.

9. The transfer operator method for subshifts of finite type

9.1. **Transfer operator.** Let A be an $M \times M$-transition matrix and put

$$\Sigma^+ = \left\{\vec{x} \in \{1,\dots,M\}^{\mathbb{N}} : A_{x_i x_{i+1}} = 1 \forall i \geq 1\right\}$$

for the one-sided shift space. The left shift map $\sigma : \Sigma^+ \to \Sigma^+$ is defined as before ($(\sigma x)_i = x_{i+1} \forall i$) but now is only locally invertible. It is an at most M to 1 map. In fact $\sigma^{-1}x = \{\eta x : \eta \in \{1,\dots,M\}\}$ where it is understood that the point ηx lies in Σ^+, i.e. $A_{\eta x_1} = 1$. Similarly $\sigma^{-n}x = \{\eta x\}$, where $\eta = \eta_1\eta_2\cdots\eta_n$ ranges over all permissible n-words in Σ^+ that satisfy $A_{\eta_n x_1} = 1$. The topology is generated by cylinder sets $U(\eta)$ where η ranges over all finite words in Σ^+. Let $f \in C(\Sigma^+)$ be a (real valued) function on Σ^+, then

$$\mathrm{var}_n f = \sup_{x_i=y_i, i=1,\dots,n} |f(x)-f(y)| = \max_{A\in\mathcal{A}^n} \sup_{x,y\in A} |f(x)-f(y)|$$

is the *n-variation* of f. For $\vartheta \in (0,1)$ we put

$$|f|_\vartheta = \sup_n \vartheta^{-n}\mathrm{var}_n f$$

and define a norm

$$\|f\|_\vartheta = |f|_\infty + |f|_\vartheta.$$

The space

$$C_\vartheta(\Sigma^+) = \left\{f \in C(\Sigma^+) : \|f\|_\vartheta < \infty\right\}$$

is the space of ϑ-Hölder continuous functions on Σ^+ which forms in fact a Banach space (Exercise). Let $f \in C_\vartheta(\Sigma^+)$ and define the *transfer operator* $\mathcal{L}_f$ on $C_\vartheta(\Sigma^+)$ by

$$(\mathcal{L}_f\varphi)(x) = \sum_{y\in\sigma^{-1}x} e^{f(y)}\varphi(y), \quad \varphi \in C_\vartheta(\Sigma^+).$$

We also get

$$(\mathcal{L}_f^2\varphi)(x) = \sum_{y\in\sigma^{-1}x} e^{f(y)}(\mathcal{L}_f\varphi)(y) = \sum_{y\in\sigma^{-1}x} e^{f(y)} \sum_{z\in\sigma^{-1}y} e^{f(z)}\varphi(z) = \sum_{z\in\sigma^{-2}x} e^{f(\sigma z)+f(z)}\varphi(z),$$

as $\sigma z = y$, and inductively $(\mathcal{L}^n\varphi)(x) = \sum_{y\in\sigma^{-n}x} e^{f^n(y)}\varphi(y)$.

Lemma 65. *$\mathcal{L}$ maps $C_\vartheta(\Sigma^+)$ into $C_\vartheta(\Sigma^+)$.*

Proof. Let $\varphi \in C_\vartheta(\Sigma^+)$, then

$$|\mathcal{L}\varphi|_\infty \le |\varphi|_\infty \sum_{y\in\sigma^{-1}x} e^{f(y)} \le M|\varphi|_\infty e^{|f|_\infty} < \infty$$

as $\mathcal{L}\mathbb{1}(x) \le c_1 \forall x \in \Sigma^+$ for a constant $c_1 < \infty$. To get a bound on the variation let $x, y \in \Sigma^+$ so that $x_i = y_i \forall i \le n$ for some n. Then for some constants c_2, c_3, c_4 independent of φ:

$$\begin{aligned}
|\mathcal{L}\varphi(x) - \mathcal{L}\varphi(y)| &= \left|\sum_\eta \left(\varphi(\eta x)e^{f(\eta x)} - \varphi(\eta y)e^{f(\eta y)}\right)\right| \\
&\le \sum_\eta e^{f(\eta x)}|\varphi(\eta x) - \varphi(\eta x)| + |\varphi(\eta y)| \sum_\eta \left|e^{f(\eta x)} - e^{f(\eta y)}\right| \\
&\le \sum_\eta e^{f(\eta x)}\mathrm{var}_{n+1}\varphi + |\varphi|_\infty \sum_\eta e^{f(\eta y)} \left|1 - e^{f(\eta x)-f(\eta y)}\right| \\
&\le \mathcal{L}\mathbb{1}(x)\mathrm{var}_{n+1}\varphi + c_2|\varphi|_\infty \mathcal{L}\mathbb{1}(y)\mathrm{var}_{n+1}f \\
&\le c_3|\varphi|_\vartheta \vartheta^{n+1} + c_4|\varphi|_\infty \vartheta^n,
\end{aligned}$$

where we used that $f(\eta x) - f(\eta y) = \mathcal{O}(\mathrm{var}_{n+1}f) = \mathcal{O}(\vartheta^{n+1})$. Thus $\mathrm{var}_n\mathcal{L}\varphi \le c_5\|\varphi\|_\vartheta\vartheta^n$ (for some c_5) and consequently $\|\mathcal{L}\varphi\|_\vartheta < \infty$. ∎

9.2. Ruelle's Perron-Frobenius theorem.

Theorem 66. *(Ruelle's Perron-Frobenius theorem) Let A be irreducible and aperiodic and $f \in C_\vartheta(\Sigma^+)$. Then:*
(I) There exists a simple real positive eigenvalue λ and a positive $h \in C_\vartheta(\Sigma^+)$ so that $\mathcal{L}h = \lambda h$.
(II) There exists a $\nu \in C^*_\vartheta(\Sigma^+)$ so that $\mathcal{L}^*\nu = \lambda\nu$.
(III) $\lambda^{-n}\mathcal{L}^n\chi \to \nu(\chi)h$ as $n \to \infty$ for all $\chi \in C_\vartheta(\Sigma^+)$.
(IV) (Quasicompactness) $\{$spectrum of $\mathcal{L}\} \setminus \{\lambda\}$ is contained in a disk strictly smaller than λ.

Proof. (I) Let us define the map G on the σ-invariant probability measures $\mathcal{M}(\sigma)$ on Σ^+ by

$$\mathcal{G}(\mu) = \frac{\mathcal{L}^*\mu}{\mathcal{L}^*\mu(1)}.$$

Since $\mathcal{M}(\sigma)$ is a convex and compact set (in the weak topology), by Schauder-Tychonoff[2] there exists a fixed point $\nu \in \mathcal{M}(\sigma)$, that is $\mathcal{G}(\nu) = \nu$. Consequently $\mathcal{L}^*\nu = \lambda\nu$, where $\lambda = \mathcal{L}^*\nu(1) \in (0, \infty)$

Now we prove the existence of h. Put

$$B_n = \exp\left(|f|_\vartheta \frac{\vartheta^n}{1-\vartheta}\right) = \exp\left(|f|_\vartheta \sum_{k=n}^\infty \vartheta^k\right) = B_{n+1}e^{|f|_\vartheta\vartheta^n}$$

[2] It says that a continuous map from a convex compact Banach space into itself has a fixed point (a generalisation of Brower's fixed point theorem).

and

$$\Lambda = \left\{ g \in C_\vartheta(\Sigma^+) : g \geq 0, \nu(g) = 1, g(x) \leq B_n g(y) \text{ if } x_i = y_i \forall i \leq n \right\}.$$

Note that Λ is closed and convex. Define the operator $\mathcal{M}$ on Λ by $\mathcal{M}\chi = \frac{\mathcal{L}\chi}{\lambda}$.

Lemma 67. $\mathcal{M} : \Lambda \to \Lambda$

Proof. We check the three conditions in the definition of $\mathcal{M}$. First note that $\mathcal{M}\chi \geq 0$ if $\chi \geq 0$ as $\mathcal{L}$ is a positive operator. Second we note that $\nu(\mathcal{M}\chi) = \frac{1}{\lambda}\nu(\mathcal{L}\chi) = \frac{1}{\lambda}\mathcal{L}^*\nu(\chi) = \nu(\chi) = 1$. Third we check on the regularity condition. Let $x, y \in \Sigma^+$ be so that $x_i = y_i$ for $i \leq n$ for some n, then

$$\begin{aligned}
\mathcal{M}\chi(x) &= \frac{1}{\lambda}\sum_\alpha e^{f(\alpha x)}\chi(\alpha x) \\
&\leq \frac{1}{\lambda}\sum_\alpha e^{f(\alpha x)} B_{n+1}\chi(\alpha y) \\
&\leq \frac{B_{n+1}}{\lambda}\sum_\alpha \chi(\alpha y) e^{f(\alpha y)} e^{f(\alpha x) - f(\alpha y)} \\
&\leq \frac{B_{n+1}}{\lambda}\sum_\alpha \chi(\alpha y) e^{f(\alpha y)} e^{|f|_\vartheta \vartheta^{n+1}} \\
&\leq \frac{B_n}{\lambda}\mathcal{L}\chi(y) \\
&= B_n \mathcal{M}\chi(y)
\end{aligned}$$

as $|f(\alpha x) - f(\alpha y)| \leq \mathrm{var}_{n+1} f \leq |f|_\vartheta \vartheta^{n+1}$. ∎

Lemma 68. *There exists a constant K so that $|\chi|_\infty \leq K$ for all $\chi \in \Lambda$.*

Proof. Let $\chi \in \Lambda$ and y and arbitrary point in Σ^*. Since A is irreducible and aperiodic, there exists and integer N so that A^N is strictly positive. So, if $x \in \Sigma$ is chosen then there exists a point $z \in \sigma^{-n}x$ so that $z_0 = y_0$ and therfore $g(y) \leq B_0 g(z)$. Moreover

$$\mathcal{L}^N\chi(x) = \sum_{w \in \sigma^{-n}x} e^{f^N(w)} g(w) \geq e^{f^N(z)} g(z) \geq e^{-N|f|_\infty} B_0^{-1} g(y)$$

and since $\frac{1}{\lambda^N}$
$nu(\mathcal{L}^N\chi) = 1$ we can choose $x \in \Sigma^+$ so that $\frac{1}{\lambda^N}\mathcal{L}^n\chi(x) \leq 1$ and therefore conclude that $g(y) \leq K$ where $K \leq e^{N|f|_\infty}\lambda^N B_0$. Since $y \in \Sigma^+$ is arbitrary we get $|\chi|_\infty \leq K$. ∎

Let us now observe that Λ is a family of equicontinuous functions since by the third property we have

$$|\chi(x) - \chi(y)| \leq (B_n - 1)|\chi|_\infty \leq (B_n - 1)K \leq K e^{\frac{1}{1-\vartheta}|f|_\vartheta} \vartheta^n$$

for every $\chi \in \Lambda$, every $x, y \in \Sigma$ so that $x_i = y_i \forall i \leq n$ and every n.

We thus conclude that by the theorem of Arzela-Ascoli, Λ is compact in the $|\cdot|_\infty$-norm. Since moreover Λ is convex, we can apply the theorem of Schauder-Tychonoff and conclude that $\mathcal{M}$ has a fixed point $h \in \Lambda$: $\mathcal{M}h = h$. This implies that $\mathcal{L}h = \lambda h$, where

$\lambda = |\mathcal{L}h|_\infty$. It remains to show that (a) h is strictly positive, i.e. has no zeroes, and (b) h is unique up to scalar multiples, i.e. λ is simple.

(a) To get positivity of h let us note that $h \geq 0$ as $h \in \Lambda$ and suppose $h(x) = 0$ for some $x \in \Sigma^+$. Then for all $n = 1, 2, \dots,$

$$0 = h(x) = \frac{1}{\lambda^n}\mathcal{L}^n h(x) = \frac{1}{\lambda^n}\sum_{|\alpha|=n} e^{f^n(\alpha x)} h(\alpha x).$$

Since $e^{f^n(\alpha x)} > 0$ for all $\alpha x \in \Sigma^+$ we must have $h(\alpha x) = 0$ for all α, or h vanishes on $\sigma^{-n}x$ for any $n \in \mathbb{N}$. Hence h is zero on $\bigcup_n \sigma^{-n}x$ which is dense in Σ^+. Hence h is identically zero. This is a contradition.

(b) To show that λ is a simple eigenvalue suppose there exists another eigenfunction $h' \in \Lambda$ so that $\mathcal{L}h' = \lambda h'$. If $t = \inf_x \frac{h'(x)}{h(x)}$ then $h' - th \geq 0$ and $h'(x) - th(x) = 0$ for some $x \in \Sigma^+$. As $h' - th \geq 0 \Rightarrow h' - th \in \Lambda$ (by convexity) and $\mathcal{L}(h' - th) = \lambda(h' - th)$ we see that $h' - th$ is and eigenfunction with a zero and thus by part (a) above must be identically zero. Thus h' is a scalar multiple of h. This finishes the proof of part (I) of the Ruelle Perron-Frobenius theorem.

(II) By standard Banach space theory there exists a $\nu \in C^*_\vartheta(\Sigma^+)$ so that $\mathcal{L}^*\nu = \lambda\nu$. We can normalise so that $\nu(h) = 1$. Then $\chi \to \nu(\chi)h$ is the projection onto the eigenspace spanned by h. This concludes part (II) of the RFP.

(III) To prove convergence $\lambda^{-n}\mathcal{L}^n\chi \to \nu(\chi)h$ we normalise the transfer operator. Put $\bar{f} = f - \log\lambda - \log h \circ \sigma + h$ which lies in C_ϑ since $h \in C_\vartheta$ is strictly positive. Then $\bar{f}^n = f^n - n\log\lambda + \log h \circ \sigma^n + \log h$. Put $\bar{\mathcal{L}} = \mathcal{L}_{\bar{f}}$ for the normalised transfer operator which acts on functions as $\bar{\mathcal{L}}\varphi = \frac{1}{\lambda h}\mathcal{L}(h\varphi)$ and for higher powers $\bar{\mathcal{L}}^n\varphi = \frac{1}{\lambda h}\mathcal{L}^n(h\varphi)$. The leading eigenvalue of $\bar{\mathcal{L}}$ is 1 which has the associated eigenfunction $\mathbb{1}$ as $\bar{\mathcal{L}}\mathbb{1} = \frac{1}{\lambda h}\mathcal{L}(h\mathbb{1}) = \frac{1}{\lambda h}\lambda h = \mathbb{1}$ which means that $\sum_\alpha e^{\bar{f}(\alpha x)} = 1$ for all $x \in \Sigma^+$.

Lemma 69. *(Lasota-Yorke or Doeblin-Fortet inequality) There exists a constant C so that*

$$\left|\bar{\mathcal{L}}^n\varphi\right|_\vartheta \leq \vartheta^n|\varphi|_\vartheta + C|\varphi|_\infty$$

for all $n \in \mathbb{N}$ and $\varphi \in C_\vartheta$.

Proof. To estimate $\text{var}_k\bar{\mathcal{L}}^n\varphi$ let $x, y \in \Sigma^+$ be so that $x_1 \cdots x_k = y_1 \cdots y_k$. Then

$$\begin{aligned}
\left|\bar{\mathcal{L}}^n\varphi(x) - \bar{\mathcal{L}}^n\varphi(y)\right| &= \left|\sum_{|\alpha|=n}\left(e^{\bar{f}^n(\alpha x)}\varphi(\alpha x) - e^{\bar{f}^n(\alpha y)}\varphi(\alpha y)\right)\right| \\
&\leq \sum_{|\alpha|=n} e^{\bar{f}^n(\alpha x)}|\varphi(\alpha x) - \varphi(\alpha y)| + \sum_{|\alpha|=n}|\varphi(\alpha y)|e^{\bar{f}^n(\alpha y)}\left|1 - e^{\bar{f}^n(\alpha y) - \bar{f}^n(\alpha y)}\right| \\
&\leq (\text{var}_{n+k}\varphi)\sum_{|\alpha|=n} e^{\bar{f}^n(\alpha x)} + |\varphi|_\infty\sum_{|\alpha|=n} e^{\bar{f}^n(\alpha y)}c_1\vartheta^n|\bar{f}|_\vartheta \\
&\leq \vartheta^{n+k}|\varphi|_\vartheta + c_1|\varphi|_\infty\vartheta^n|\bar{f}|_\vartheta
\end{aligned}$$

for a constant c_1 where we used that

$$|\bar{f}^n(\alpha y) - \bar{f}^n(\alpha y)| \leq \sum_{j=0}^{n-1} |\bar{f}(\sigma^j \alpha y) - \bar{f}(\sigma^j \alpha y)| \leq \sum_{j=0}^{n-1} \mathrm{var}_{n+k-j} \bar{f} \leq \sum_{j=0}^{n-1} \vartheta^{n+k-j} |\bar{f}|_\vartheta \leq \frac{\vartheta^n}{1-\vartheta} |\bar{f}|_\vartheta$$

(as $(\sigma^j \alpha x)_i = (\sigma^j \alpha y)_i$ for $i = 1, \ldots, n+k-j$) and the normalisation $\sum_{|\alpha|=n} e^{\bar{f}^n(\alpha x)} = 1$ for all x. Thus

$$\mathrm{var}_k \bar{\mathcal{L}}^n \varphi \leq \vartheta^k (\vartheta^n |\varphi|_\vartheta + C|\varphi|_\infty)$$

which implies $|\bar{\mathcal{L}}^n \varphi|_\vartheta \leq \vartheta^n |\varphi|_\vartheta + C|\varphi|_\infty$. ∎

The lemma in particular implies that for every $\varphi \in C_\vartheta$ the set $\{\bar{\mathcal{L}}^n \varphi : n \in \mathbb{N}\}$ is equicontinuous and has by the theorem of Arzela-Ascoli an accumulation point $\ell_\varphi \in C(\Sigma^+)$ (though not necessarily in $C_\vartheta(\Sigma^+)$.)

We will now show that ℓ_φ is a constant. Indeed, since

$$\inf \varphi \leq \inf \bar{\mathcal{L}} \varphi \leq \inf \bar{\mathcal{L}}^2 \varphi \leq \inf \bar{\mathcal{L}}^3 \varphi \leq \cdots$$

we conclude $\inf \ell_\varphi = \inf \bar{\mathcal{L}}^n \ell_\varphi$ for all n. If $x \in \Sigma^+$ is such that $\ell_\varphi(x) = \inf \ell_\varphi$ then by convexity $\ell_\varphi(x) = \sum_{y \in \sigma^{-n}x} e^{\bar{f}^n(y)} \ell_\varphi(y)$ implies $\ell_\varphi(y) = \inf \ell_\varphi$ for all $y \in \sigma^{-n}x$ as $\sum_{y \in \sigma^{-n}x} e^{\bar{f}^n(y)} = 1$. Since this applies to every n we must have $\ell_\varphi = \inf \ell_\varphi$ on the set $\bigcup_n \sigma^{-n}x$ which is dense in Σ^+. So ℓ_φ is constant.

In this way we obtain a positive linear functional $\varphi \mapsto \ell_\varphi$ on C_ϑ which by Riesz's representation theorem implies the existence of a measure μ so that $\ell_\varphi = \int_{\Sigma^+} \varphi \, d\mu$ for all $\varphi \in C_\vartheta$. We assume μ is normalised to have mass 1. Since $\ell_\varphi = \bar{\mathcal{L}} \ell_\varphi = \ell_{\bar{\mathcal{L}}\varphi}$ one has

$$\mu(\varphi) = \ell_\varphi = \bar{\mathcal{L}} \ell_\varphi = \mu(\bar{\mathcal{L}} \varphi) = (\bar{\mathcal{L}}^* \mu)(\varphi) \ \ \forall \varphi \in C_\vartheta$$

i.e. $\bar{\mathcal{L}}^* \mu = \mu$.

It remains to verify that $\mu = h\nu$ where ν is the eigenfunctional to the eigenvalue λ (from part (II)). Indeed, since for every $\varphi \in C_\vartheta$,

$$\left(\frac{1}{h}\mu\right)(\varphi) = \mu\left(\frac{\varphi}{h}\right) = \mu\left(\bar{\mathcal{L}} \frac{\varphi}{h}\right) = \mu\left(\frac{1}{\lambda h} \mathcal{L} h \frac{\varphi}{h}\right) = \mu\left(\frac{1}{\lambda h} \mathcal{L} \varphi\right) = \frac{1}{\lambda}\left(\bar{\mathcal{L}}^* \frac{1}{h}\mu\right)(\varphi)$$

the functional $\frac{1}{h}\mu \in C_\vartheta^*$ is an eigenfunctional to the simple eigenvalue λ of the non-normalised transfer operator $\mathcal{L}$. Since $\mu(\Sigma^+) = 1$ and $\nu(h) = 1$ we conclude that $\nu = \frac{1}{h}\mu$.

Hence $\bar{\mathcal{L}}^n \varphi \to \ell_\varphi = (h\nu)(\varphi)$ in the infinity norm as $n \to \infty$ for all $\varphi \in C_\vartheta$. If we write $\psi = h\varphi$ (as h is positive), then

$$\frac{1}{\lambda^n} \mathcal{L}^n \psi = h \frac{1}{\lambda^n h} \mathcal{L}^n (h\varphi) = h \bar{\mathcal{L}}^n \varphi \to h\mu(\varphi) = \nu(\psi) h.$$

This concludes part (III) of the RPF theorem.

(IV) We use the normalised transfer operator $\bar{\mathcal{L}}$ and put

$$R^\perp = \{\varphi \in C_\vartheta : \mu(\varphi) = 0\}$$

for the orthogonal complement to the eigenspace $\mathbb{R}\mathbb{1}$ in C_ϑ. We will show that the spectral radius of $\bar{\mathcal{L}}$ restricted to $R^\perp$ is strictly less than 1, that is $\bar{\rho} = \limsup_{n \to \infty} \|\bar{\mathcal{L}}^n \varphi\|_\vartheta^{\frac{1}{n}} < 1$

for all $\varphi \in R^\perp$. Clearly, for $\varphi \in R^\perp$ one has by part (III) $\bar{\mathcal{L}}^n\varphi \to \mu(\varphi) = 0$ as $n \to \infty$. By Lasota-Yorke-Doeblin-Fortet

$$\left|\bar{\mathcal{L}}^{2n}\varphi\right|_\vartheta \le \vartheta^n \left|\bar{\mathcal{L}}^n\varphi\right|_\vartheta + C\left|\bar{\mathcal{L}}^n\varphi\right|_\infty.$$

Without loss of generality we can assume that $\|\varphi\|_\vartheta = 1$. By compactness of the unit-sphere $\{\varphi \in C_\vartheta : \|\varphi\|_\vartheta = 1\}$ in the supremum norm, one has

$$\sup_{\varphi\in R^\perp, \|\varphi\|_\vartheta = 1} \left|\bar{\mathcal{L}}^n\varphi\right|_\infty \to 0$$

as $n \to \infty$. Thus for $\varepsilon > 0$ small $\sup_{\varphi\in R^\perp, \|\varphi\|_\vartheta=1} \left|\bar{\mathcal{L}}^n\varphi\right|_\infty < \varepsilon$ for large enough n. Then

$$\begin{array}{rcl} \left|\bar{\mathcal{L}}^{2n}\varphi\right|_\vartheta & \le & \vartheta^n\left(\vartheta^n|\varphi|_\vartheta + C|\varphi|_\infty\right) + C\left|\bar{\mathcal{L}}^n\varphi\right|_\infty \\ & \le & \vartheta^n c_1\|\varphi\|_\vartheta + C\varepsilon \\ & \le & \sqrt{\varepsilon} \end{array}$$

($c_1 \le \max(1, C)$) for n large enough and ε small enough. Therefore

$$\|\bar{\mathcal{L}}^{2n}\varphi\|_\vartheta \le |\bar{\mathcal{L}}^{2n}\varphi|_\infty + |\bar{\mathcal{L}}^{2n}\varphi|_\vartheta \le 2\sqrt{\varepsilon}$$

and consequently $\|\bar{\mathcal{L}}^{2nk}\varphi\|_\vartheta \le (2\sqrt{\varepsilon})^k\|\varphi\|_\varphi$ for all $\varphi \in R^\perp$ and k. Finally we obtain $\bar{\rho} \le \lim_{k\to\infty} \| \left.\bar{\mathcal{L}}^{2nk}\right|_{R^\perp} \|^{\frac{1}{2nk}} \le (2\sqrt{\varepsilon})^{\frac{1}{2n}} < 1$ (ε small). ∎

As an example we describe an ingredient which is used to determine the page rank in Google's search algorithm. The page rank is partly based on the number of times a page is linked to. Assume there are M websites ($M \gg 1$) and add the 'super site' 0 to make it the alphabet $\mathcal{A} = \{0, 1, 2, \dots, M\}$. The site 0 makes the system irreducible. One defines the $(M+1)\times(M+1)$-transition matrix A by

$$A_{ij} = \begin{cases} 0 & \text{if } i = j = 0 \\ 1 & \text{if either } i = 0 \text{ or } j = 0 \text{ but not } (0,0) \\ 1 & \text{if } i \text{ links to } j \\ 0 & \text{otherwise} \end{cases}$$

and the compatible stochastic matrix B as follows

$$\begin{array}{rcl} B_{00} & = & 0 \\ B_{0j} & = & \dfrac{1}{M} \quad \text{for } j \ne 0 \\ B_{i0} & = & \begin{cases} 1 & \text{if } \sum_j A_{ij} = 1 \text{ (i.e. no links from } i\text{th site for } i \ge 1) \\ 1-\beta & \text{if } \sum_j A_{ij} \ge 2 \end{cases} \\ B_{ij} & = & \begin{cases} 0 & \text{if } A_{ij} = 0 \\ \frac{\beta}{\sum_k A_{ik}} & \text{if } A_{ij} = 1 \end{cases} \end{array}$$

for a parameter $\beta \in (0,1)$. Note that this corresponds to choosing a potential function f which depends only on two coordinates ($f(\vec{x}) = f_{x_0x_1}$):

$$\begin{aligned} f_{00} &= 0 \\ f_{0j} &= -\log M \quad \text{for } j \neq 0 \\ f_{i0} &= \begin{cases} 0 & \text{if } \sum_j A_{ij} = 1 \\ \log(1-\beta) & \text{if } \sum_j A_{ij} \geq 2 \end{cases} \\ f_{ij} &= \begin{cases} 0 & \text{if } A_{ij} = 0 \\ \log\beta - \log\sum_k A_{ik} & \text{if } A_{ij} = 1 \end{cases} . \end{aligned}$$

By the Perron-Frobenius theorem there exists a left eigenvector $\vec{p} = (p_0, p_1, \dots, p_M)$ so that $\vec{p}B = \vec{p}$, $B\mathbb{1} = \mathbb{1}$. The value p_j is the page rank of the jth site ($j \geq 1$). To compute $\vec{p}$ one uses the fact that $p_j = (B^\infty)_{ij}$ for any i, where $B^\infty = \lim_n B^n$ (see section 4.5). In practice it is enough to compute some iterates of B in order to get a reasonable approximation of $\vec{p}$.

Let us observe that the measure μ from RPF theorem is σ-invariant because

$$(\sigma^*\mu)(\varphi) = \nu(h(\varphi\circ\sigma)) = \frac{1}{\lambda}(\mathcal{L}^*\nu)(h(\varphi\circ\sigma)) = \frac{1}{\lambda}(\mathcal{L}(h(\varphi\circ\sigma))) = \frac{1}{\lambda}\nu(\varphi\mathcal{L}h) = \nu(\varphi h) = \mu(h)$$

for all $\varphi \in C_\vartheta$.

9.3. **Spectrum of the transfer operator.** Next we determine the spectrum of $\mathcal{L}$ of which we currently only know the dominant eigenvalue λ.

Theorem 70. *The spectrum of $\mathcal{L}$ has a simple dominating eigenvalue $\lambda \in (0,\infty)$, discrete eigenvalues of finite multiplicities in the annulus $\{z : \vartheta\lambda < |z| < \lambda\}$ and an essential spectrum that fills the disk $\{z : |z| \leq \vartheta\lambda\}$.*

Equivalently the spectrum of the normalised transfer operator $\bar{\mathcal{L}}$ has a simple dominating eigenvalue 1 discrete eigenvalues of finite multiplicities in the annulus $\{z : \vartheta < |z| < 1\}$ and an essential spectrum that fills the disk $\{z : |z| \leq \vartheta\}$. To find the essential spectrum we proceed in two stages.

Lemma 71. *(Keller) The essential spectrum of $\mathcal{L}$ has radius $\leq \vartheta\lambda$.*

Proof. We use an essential spectrum formula due to Nussbaum according to which the essential spectrum has radius ρ which is given by

$$\rho = \inf_{\{K_n\}} \limsup_{n\to\infty} \|\mathcal{L}^n - K_n\|_\vartheta^{\frac{1}{n}}$$

where the infimum is over all sequences $\{K_n : n\}$ of compact operators in C_ϑ. In order to define a sequence of compact operators K_n which will give us an upper bound, we introduce the projection S_n given by

$$S_n\varphi = \sum_{|\alpha|=n} \chi_\alpha \frac{1}{\mu(\chi_\alpha)} \int_{U(\alpha)} \varphi(x)\, d\mu(x)$$

for $\varphi \in C_\vartheta$, where χ_α is the characteristic function of the cylinder set $U(\alpha)$. Clearly $S_n\varphi$ is a locally constant function on the n-cylinders $U(\alpha)$, where α ranges over all permissible n-words. The range of S_n is a finite dimensional subspace of C_ϑ of dimension $|\mathcal{A}^n| = \|A^{n-1}\|$. We define $K_n = \mathcal{L}^n S_n$ which is a compact operator on C_ϑ for every n.

If $\varphi \in C_\vartheta$ and $\psi = \varphi - S_n\varphi = (\mathbb{1} - S_n)\varphi$ then

$$|\psi|_\infty \le \mathrm{var}_n\varphi \le |\varphi|_\vartheta \vartheta^n$$

and

$$\mathrm{var}_k\psi \le \begin{cases} 2|\psi|_\infty & \text{if } k \le n \\ \mathrm{var}_k\vartheta & \text{if } k > n \end{cases}.$$

Then $|\bar{\mathcal{L}}\psi|_\infty \le |\psi|_\infty \le |\varphi|_\vartheta\vartheta^n$ and therefore $|\mathcal{L}\psi|_\infty \le c_1|\varphi|_\vartheta\vartheta^n\lambda^n$. To estimate $\mathrm{var}_k\mathcal{L}^n\psi$ let $x, y \in \Sigma^+$ so that $x_1\cdots x_k = y_1\cdots y_k$. Then

$$\begin{aligned}
|\mathcal{L}^n\psi(x) - \mathcal{L}^n\psi(y)| &\le \sum_{|\alpha|=n} \left|e^{f^n(\alpha x)}\psi(\alpha x) - e^{f^n(\alpha y)}\psi(\alpha y)\right| \\
&\le \sum_{|\alpha|=n} e^{f^n(\alpha x)}|\psi(\alpha x) - \psi(\alpha y)| + \sum_{|\alpha|=n} |\psi(\alpha y)e^{f^n(\alpha y)}\left|1 - e^{f^n(\alpha x)-f^n(\alpha y)}\right| \\
&\le (\mathrm{var}_{n+k}\psi)\mathcal{L}^n\mathbb{1}(x) + |\psi|_\infty\mathcal{L}^n\mathbb{1}(y)c_1\mathrm{var}_k f \\
&\le c_2(\vartheta\lambda)^n\vartheta^k|\varphi|_\vartheta + c_3(\vartheta\lambda)^n|\varphi|_\vartheta\vartheta^k \\
&\le c_4(\vartheta\lambda)^n\vartheta^k|\varphi|_\vartheta
\end{aligned}$$

where we have used that $\mathcal{L}^n\mathbb{1}(x) \le c_2\lambda^n$ for a constant c_2. Consequently $|\mathcal{L}^n\psi|_\vartheta \le c_4(\vartheta\lambda)^n|\varphi|_\vartheta$ and $\|\mathcal{L}^n\psi\|_\vartheta \le c_5(\vartheta\lambda)^n\|\varphi\|_\vartheta$. Since $\mathcal{L}^n\psi = \mathcal{L}^n(\mathbb{1} - S_n)\varphi = (\mathcal{L}^n - K_n)\varphi$ we get $\|\mathcal{L}^n - K_n\|_\vartheta \le c_6(\vartheta\lambda)^n$ and therefore $\rho \le \vartheta\lambda$. ∎

Lemma 72. *(Parry) The essential spectrum of $\mathcal{L}$ contains the disk $\{z : |z| < \vartheta\lambda\}$, that is $\rho \ge \vartheta\lambda$.*

Proof. We show that for the normalised transfer operator $\bar{\mathcal{L}}$ every value γ with $|\gamma| < \vartheta$ is an eigenvalue of $\bar{\mathcal{L}}$.

(I) We find a $\varphi \in C_\vartheta$ so that $\bar{\mathcal{L}}\varphi = 0$. For this we choose $\varphi|_U(1) \in C_\vartheta$ arbitrarily and we assume that there are points $x \in \Sigma^+$ so that $A_{1x_1} = A_{2x_1} = 1$ (otherwise replace the symbols $1, 2$ by some other two symbols of the alphabet). Put

$$\begin{cases} \varphi(x) = 0 & \text{if } x_1 \ne 1, 2 \\ \varphi(2x) = -\varphi(1x)e^{\bar{f}(1x)-\bar{f}(2x)} & \text{otherwise} \end{cases}$$

Then

$$(\bar{\mathcal{L}}\varphi)(x) = \sum_{\alpha=1}^{M} e^{\bar{f}(\alpha x)}\varphi(\alpha x) = e^{\bar{f}(1x)}\varphi(1x) + e^{\bar{f}(2x)}\varphi(2x) = \varphi(1x)\left(e^{\bar{f}(1x)} + e^{\bar{f}(2x)}\left(-e^{\bar{f}(1x)-\bar{f}(2x)}\right)\right) = 0$$

and we define

$$\chi = \sum_{k=0}^{\infty} \gamma^k(\varphi \circ \sigma^k)$$

for $|\gamma| < \vartheta$.

(II) We now show that χ lies in C_ϑ. The infinity norm estimates as follows

$$|\chi|_\infty \le \sum_k |\gamma|^k |\varphi\circ\sigma^k|_\infty \le \frac{|\varphi|_\infty}{1-|\gamma|} < \infty$$

and the Hölder constant like this

$$\begin{aligned}
\mathrm{var}_\ell\chi &\le \sum_k |\gamma|^k \mathrm{var}_\ell \varphi\circ\sigma^k \\
&\le \sum_{k=0}^{\ell} |\gamma|^k \mathrm{var}_{\ell-k}\varphi + \sum_{k=\ell+1}^{\infty} |\gamma|^k |\varphi|_\infty \\
&\le \sum_{k=0}^{\ell} |\gamma|^k \vartheta^{\ell-k} |\varphi|_\vartheta + |\varphi|_\vartheta \frac{|\gamma|^{\ell+1}}{1-|\gamma|} \\
&\le \vartheta^\ell |\varphi|_\vartheta \sum_{k=0}^{\ell} \left(\frac{|\gamma|}{\vartheta}\right)^k + \frac{|\varphi|_\vartheta}{1-|\gamma|}|\gamma|^{\ell+1} \\
&\le c_1 \|\varphi\|_\vartheta \vartheta^\ell
\end{aligned}$$

as $|\gamma| < \vartheta$. Therefore $|\chi|_\vartheta \le c_1\|\varphi\|_\vartheta$ which implies that $\chi \in C_\vartheta$.

(III) Now we show that χ is an eigenfunction for $\bar{\mathcal{L}}$ to the eigenvalue γ. Indeed

$$\bar{\mathcal{L}}\chi = \sum_{k=0}^{\infty} \gamma^k \bar{\mathcal{L}}(\varphi\circ\sigma^k) = \bar{\mathcal{L}}\varphi + \sum_{k=1}^{\infty} \gamma^k \varphi\circ\sigma^{k-1}\bar{\mathcal{L}}\mathbb{1}$$

as $\bar{\mathcal{L}}(\varphi\circ\sigma^k\mathbb{1}) = \varphi\circ\sigma^{k-1}\bar{\mathcal{L}}\mathbb{1}$ for $k \ge 1$. Since $\bar{\mathcal{L}}\varphi = 0$ by construction and $\bar{\mathcal{L}}\mathbb{1} = \mathbb{1}$ we obtain

$$\bar{\mathcal{L}}\chi = \sum_{k=1}^{\infty} \gamma^k \varphi\circ\sigma^{k-1} = \gamma\chi$$

∎

The last two lemmas prove the theorem.

9.4. **Gibbs states.**

Definition 73. *An invariant measure $\mu \in \mathscr{M}(\sigma)$ is a* Gibbs state *for $f \in C(\Sigma^+)$ if there exists a $P \in \mathbb{R}$ and a constant $C > 0$ so that*

$$\frac{1}{C} \le \frac{\mu(U(x_1x_2\cdots x_n))}{e^{f^n(x)-nP}} \le C$$

for all $x \in \Sigma$ and for all n.

Lemma 74. *If μ, μ' are Gibbs for some $f \in C(\Sigma^+)$ then $P = P'$ and μ, μ' are equivalent and equal if one of them is ergodic.*

Proof. μ Gibbs implies that $\mu(U(x_1\cdots x_n))e^{f^n(x)-nP} \in \left[\frac{1}{C}, C\right]$ and μ' Gibbs implies $\mu'(U(x_1\cdots x_n))e^{f^n(x)-nP'} \in \left[\frac{1}{C'}, C'\right]$.

(I) We first show $P = P'$. Since

$$\sum_{x_1\cdots x_n} e^{f^n(x)-nP} \le \sum_{x_1\cdots x_n} C\mu(U(x_1\cdots x_n)) = C$$

and

$$\sum_{x_1\cdots x_n} e^{f^n(x)-nP'} \ge \sum_{x_1\cdots x_n} \frac{1}{C'}\mu'(U(x_1\cdots x_n)) = \frac{1}{C'}$$

one has $P' \le \lim_n \frac{1}{n}\log Z_n \le P$. In a similar fashion one obtains $P \le \lim_n \frac{1}{n}\log Z_n \le P'$. This implies that $P = P' = P(f)$ the pressure of f.

(II) Now we show that μ and μ' are equivalent. Indeed, for all $x \in \Sigma^+$ and for all n one has

$$\frac{1}{CC'}\mu'(U(x_1\cdots x_n)) \le \mu(U(x_1\cdots x_n)) \le CC'\mu'(U(x_1\cdots x_n))$$

which implies $\mu' \ll \mu \ll \mu'$. In particular $\mu' = g\mu$ where the Radon-Nikodym derivative g is positive on full measure set and σ-invariant. Thus, if either μ or μ' is ergodic, g is a constant and consequently $\mu = \mu'$. ∎

Lemma 75. *Let $f \in C_\vartheta$ and $\mu = h\nu \in \mathscr{M}(\sigma)$, $\mathcal{L}h = \lambda h$, $\mathcal{L}^*\nu = \lambda\nu$ where λ is the dominante eigenvalue of $\mathcal{L}$ given by the RPF theorem. Then μ is a Gibbs state for f and $\lambda = e^P$.*

Proof. Let $x \in \Sigma^+$, then

$$\mu(U(x_1\cdots x_n)) = \mu(\chi_{x_1\cdots x_n}) = \nu(h\chi_{x_1\cdots x_n}) = \frac{1}{\lambda^n}\nu(\mathcal{L}^n h\chi_{x_1\cdots x_n}),$$

where (with $\alpha = x_1\cdots x_n$ and χ_α being the characteristic function of the cylinder set $U(\alpha)$)

$$\mathcal{L}^n(h\chi_\alpha)(y) = \sum_{|\beta|=n} e^{f(\beta y)}\chi_\alpha(\beta y)h(\alpha y)e^{f^n(\alpha y)} \le |h|_\infty e^{f^n(x)+\sum_{j=0}^{n-1}\mathrm{var}_j f} \le c_1 e^{f^n(x)}$$

for a constant c_1, where we used that $\sum_{j=0}^{n-1}\mathrm{var}_j f \le |f|_\vartheta \frac{1}{1-\vartheta}$. Similarly one gets the lower bound $\mathcal{L}^n(h\chi_\alpha)(y) \ge c_2 e^{f^n(x)}$ for a constant $c_2 > 0$. Thus

$$\frac{1}{C}e^{f^n(x)}\lambda^{-n} \le \mu(U(x_1\cdots x_n)) \le Ce^{f^n(x)}\lambda^{-n}$$

for some $C > 0$. Consequently $\lambda = e^P$. ∎

Lemma 76. *If μ is an equilibrium state for $f \in C_\vartheta(\Sigma^+)$, then it is a Gibbs state for f.*

Proof. We use the fact from the proof of the variational principle that μ is a weak* limit of the sequence μ_k as $k \to \infty$ where

$$\mu_k = \frac{1}{k}\sum_{j=0}^{k-1}\sigma^{*j}\nu_k$$

and

$$\nu_k = \frac{1}{Z_k}\sum_{|\alpha|=k} e^{f^k(\alpha)}\delta_\alpha$$

where with $f^k(\alpha)$ we mean it evaluated by an arbitrarily chosen (and fixed) point in $U(\alpha)$ and similarly δ_α is the point mass at that point. Now let $x \in \Sigma^+$ and put $\beta = x_1 \cdots x_n$. Then

$$\mu(U(\beta)) \approx \mu_k(U(\beta)) = \frac{1}{kZ_k} \sum_{j=0}^{k-1} \sum_{|\alpha|=k} e^{f^k(\alpha)} \sigma^{*j} \delta_\alpha(\chi_\beta) = \frac{1}{kZ_k} \sum_{j=0}^{k-1} \sum_{|\alpha|=k} e^{f^k(\alpha)} \delta_\alpha(\chi_{\sigma^{-j}U(\beta)}).$$

Now we write $\alpha = \omega\beta\gamma$ where $|\omega| = j$, $|\beta| = n$ and $|\gamma| = k - j - n$. Then

$$\mu_k(U(\beta)) = \frac{1}{kZ_k} \sum_{j=0}^{k-1} \sum_{|\omega|=j, |\gamma|=k-j-n} e^{f^k(\omega\beta\gamma)}$$

and with the decomposition

$$f^k(\omega\beta\gamma) = f^j(\omega) + f^n(\beta) + f^{k-j-n}(\gamma) + \mathcal{O}\left(\frac{3|f|_\vartheta}{1-\vartheta}\right)$$

one obtains

$$\begin{aligned} \mu_k(U(\beta)) &= \mathcal{O}(1) e^{f^n(\beta)} \frac{1}{kZ_k} \sum_{j=0}^{k-1} \sum_{|\omega|=j} \sum_{|\gamma|=k-j-n} e^{f^j(\omega)} e^{f^{k-j-n}(\gamma)} \\ &= \mathcal{O}(1) e^{f^n(\beta)} \frac{1}{kZ_k} \sum_{j=0}^{k-1} Z_{j-N} Z_{k-j-n-N} e^{\mathcal{O}(2N|f|_\infty)} \end{aligned}$$

where N is a (fixed) number so that $A^N > 0$. We now also use the fact that

$$Z_k = Z_{j-N} Z_n Z_{k-j-n-N} e^{\mathcal{O}(2N|f|_\infty)}$$

and obtain

$$\mu_k(U(\beta)) = \mathcal{O}(1) \frac{1}{Z_n} e^{f^n(\beta)} \frac{1}{k} \sum_{j=0}^{k-1} e^{\mathcal{O}(4N|f|_\infty)} = \mathcal{O}(1) \frac{1}{Z_n} e^{f^n(x)} = \mathcal{O}(1) e^{-nP} e^{f^n(x)}.$$

Now take a subsequence $k_j \to \infty$ for which μ_{k_j} converges to μ. ∎

9.5. **Correlation function.** For $f \in C_\vartheta(\Sigma^+)$ we have as above the transfer operator $\mathcal{L}_f : C_\vartheta \to C_\vartheta$. For simplicity's sake let us assume that the discrete eigenvalues $\lambda_j \in \{z : \vartheta\lambda < |z| < \lambda\}$, $j = 1, 2, \ldots$, are simple. Let $|\lambda_1| \geq |\lambda_2| \geq |\lambda_3| \geq \ldots$, $\lambda_0 = \lambda$. Moreover let $h_j \in C_\vartheta$ be the eigenfuctions of λ_j and $\nu_j \in C_\vartheta^*$ the eigenfunctionals where we assume the normalisation $\nu_j(h_j) = 1 \forall j$. Then P_j is the projection onto the eigenspace spanned by h_j, i.e. $P_j\varphi = \nu_j(\varphi)h_j$. Then

$$\mathcal{L} = \sum_j \lambda_j P_j + R$$

where the remainder term $R : C_\vartheta \to C_\vartheta$ has spectral radius $\vartheta\lambda$, i.e. $\|R^n\|_\vartheta \leq c_1(\vartheta'\lambda)^n$ for any $\vartheta' > \vartheta$ and some constant c_1 (depending on ϑ'). Similarly $\mathcal{L}^n = \sum_j \lambda_j^n h_j + R^n$.

Let $G \in C_\vartheta$, $H \in \mathscr{L}^1$ be two functions on Ω and define the *correlation function*

$$\rho(n) = \int_{\Sigma^+} G \cdot (H \circ \sigma^n)\, d\mu,$$

where $\mu = h\nu$ is the equilibrium state for f. We expect that $\rho(n) \to \mu(G)\mu(H)$ as $n \to \infty$. Indeed we get for the powerspectrum

$$\begin{aligned}
\hat{\rho}(\omega) &= \sum_{n=0}^{\infty} e^{-i\omega n}\rho(n) \\
&= \sum_{n=0}^{\infty} e^{-i\omega n}\nu(hG(H\circ\sigma^n)) \\
&= \sum_{n=0}^{\infty} e^{-i\omega n}\frac{1}{\lambda^n}(\mathcal{L}^{*n}\nu)(hG(H\circ\sigma^n)) \\
&= \sum_{n=0}^{\infty} e^{-i\omega n}\frac{1}{\lambda^n}\nu(\mathcal{L}^n hG(H\circ\sigma^n)) \\
&= \sum_{n=0}^{\infty} e^{-i\omega n}\frac{1}{\lambda^n}\nu(H\mathcal{L}^n hG) \\
&= \sum_{n=0}^{\infty} e^{-i\omega n}\frac{1}{\lambda^n}\left(\sum_j \nu(H\lambda_j^n h_j\nu_j(hG)) + \nu(HR^n hG)\right) \\
&= \sum_{n=0}^{\infty} e^{-i\omega n}\left(\frac{\lambda_j}{\lambda}\right)^n \sum_j \nu(Hh_j)\nu_j(hG) + \tilde{\rho}(\omega),
\end{aligned}$$

where

$$\tilde{\rho}(\omega) = \sum_{n=0}^{\infty} e^{-i\omega n}\frac{1}{\lambda^n}\nu(HR^n hG)$$

is majorised by the series $\sum_{n=0}^{\infty} e^{-i\Im\omega n}\frac{1}{\lambda^n}\|HR^n hG\|_\vartheta$ which converges for $|\Im\omega| < |\log\vartheta'|$ as $\|R^n hG\|_\vartheta \le c_2(\vartheta'\lambda)^n$ for some constant c_2. Since $\vartheta' > \vartheta$ is arbitrary we conclude that $\tilde{\rho}(\omega)$ is analytic for $|\Im\omega| < |\log\vartheta|$. Therefore

$$\hat{\rho}(\omega) = \sum_j \frac{\nu(Hh_j)\nu_j(hG)}{1 - \frac{\lambda_j}{\lambda}e^{-i\omega}} + \tilde{\rho}(\omega)$$

is meromorphic in the strip $|\Im\omega| < |\log\vartheta|$ with poles whenever $\frac{\lambda_j}{\lambda}e^{-i\omega} = 1$ that when ω equals the values $\omega_j = \log\frac{\lambda_j}{\lambda}$. The residue at ω_j is $\nu(Hh_j)\nu_j(hG)$ $(h = h_0, \nu = \nu_0)$ and in particular the principal pole at $\omega_0 = 0$ has residue $\nu(Hh)\nu(hG) = \mu(H)\mu(G)$ which vanishes if $\mu(H)\mu(G) = 0$ or if the integrand in the definition of the correlation function $\rho(n)$ is replaced by $G(H\circ\sigma^n) - \mu(H)\mu(G)$. Then $\hat{\rho}$ is analytic in the strip $|\Im\omega| < \log\frac{|\lambda_1|}{\lambda}$ and by Paley-Wiener

$$\left|\int_{\Sigma^+} G\cdot(H\circ\sigma^n)\,d\mu - \int_{\Sigma^+} G\,d\mu\int_{\Sigma^+} H\,d\mu\right| \le c_3\gamma^n$$

decays exponentially fast for any $\gamma > \left|\frac{\lambda_1}{\lambda}\right|$. Thus the rate of decay is given by the 'spectral gap' between the leading eigenvalue λ and the rest of the spectrum.

9.6. **Dynamical Zeta function for subshifts.** Let us put $P_n = \{x \in \Sigma^+ : \sigma^n x = x\}$ for the set of periodic points with period n. For $f \in C_\vartheta$ we put $\zeta_n = \sum_{x\in P_n} e^{f^n(x)}$ and define the *dynamical zeta function* by

$$\zeta(z) = \exp - \sum_n \frac{1}{n}\zeta_n z^n$$

where z is a complex variable.

In the special case when $f = 0$ we get the Artin-Mazur zeta function which is

$$\zeta(z) = \exp - \sum_n \frac{z^n}{n}\text{trace}A^n = \exp - \sum_n \frac{z^n}{n}\sum_\lambda \lambda^n = \exp - \log\det(\text{id}-zA) = \frac{1}{\prod_\lambda(1-z\lambda)}$$

as $|P_n| = \text{trace}A^n$, where λ are the eigenvalues of A. This is a meromorphic function in the entire plane and is analytic for $|z| < e^{-h_{\text{top}}}$ and has a simple pole at $e^{-h_{\text{top}}}$ (Bowen-Lanford).

Theorem 77. [16] *Let $f \in C_\vartheta$, $A^n > 0$ for all n large enough. Then*
(I) $\zeta(z)$ is analytic for in the disk $|z| < e^{P(f)}$
(II) $\zeta(z)$ has a meromorphic extension to $|z| < \vartheta^{-1}e^{P(f)}$ with a simple pole at $e^{P(f)}$.

Proof. (I) In order to get analyticity note that

$$\zeta_n = \sum_{x\in P_n} e^{f^n(x)} \le \sum_{|\alpha|=n} e^{f^n(\alpha)} = Z_n.$$

Thus

$$\limsup_n \left(\frac{\zeta_n}{n}\right)^{\frac{1}{n}} \le \limsup_n \left(\frac{Z_n}{n}\right)^{\frac{1}{n}} = \limsup_n \left(\frac{Z_n}{n}\right)^{\frac{1}{n}} = P(f)$$

which by Hadamard's formula implies that $\zeta(n)$ is analytic for $|z| < \frac{1}{e^P}$. We can also note that since $A^N > 0$ for a some $N \in \mathbb{N}$ one has

$$\zeta_n \ge \sum_{|\alpha|=n-N} e^{f^n(\alpha) - \frac{|f|_\vartheta}{1-\vartheta} - N|f|_\infty} \ge \text{const.} Z_{n-N}.$$

(II) To get meromorphicity we use the spectral decomposition of $\mathcal{L}$ as

$$\mathcal{L}^n\varphi = \sum_{|\lambda|>\vartheta e^P} \lambda^n h_\lambda \nu_\lambda(\varphi) + R^n\varphi$$

where $\|R^n\varphi\|_\vartheta \le c_1\|\varphi\|_\vartheta(\vartheta' e^P)^n$ for any $\vartheta' > \vartheta$ and where h_λ, ν_λ are the eigenfunctions and eigenfunctionals to the eigenvalues λ (for simplicity's sake we assume the eigenvalues are simple). Denote by

$$\omega_\alpha(\varphi) = \frac{1}{\mu(U(\alpha))}\int_{U(\alpha)} \varphi(x)\, d\mu(x)$$

the average value of the function φ over the cylinder $U(\alpha)$ where α is an n-word. We have

$$\omega_\alpha(\mathcal{L}^n\chi_\alpha) = \frac{1}{\mu(U(\alpha))}\int_{U(\alpha)} \sum_{|\beta|=n} e^{f^n\beta x)}\chi_\alpha(\beta x)\, d\mu(x) = \frac{1}{\mu(U(\alpha))}\int_{U(\alpha)} e^{f^n\alpha x)}\, d\mu(x)$$

as $\chi_\alpha(\beta x)=1$ if and only if $\beta=\alpha$. In particular (as $\alpha^\infty\in P_n$ is a periodic point of period n if α can be concatenated with itself)

$$\left|\omega_\alpha(\mathcal{L}^n\chi_\alpha)-e^{f(\alpha^\infty)}\right|\le e^{f(\alpha^\infty)}\left(e^{\frac{|f|_\vartheta\vartheta^n}{1-\vartheta}}-1\right)$$

as

$$|f^n(\alpha x)-f^n(\alpha^\infty)|\le\sum_{j=0}^{n-1}\mathrm{var}_{2n-j}f\le\frac{|f|_\vartheta\vartheta^n}{1-\vartheta}$$

for $x\in U(\alpha)$. Thus

$$\left|\sum_{|\alpha|=n}\omega_\alpha(\mathcal{L}^n\chi_\alpha)-e^{f(\alpha^\infty)}\right|\le c_2\sum_{|\alpha|=n}e^{f(\alpha^\infty)}=c_2\zeta_n$$

and therefore

$$\left|\sum_{|\alpha|=n}\omega_\alpha(\mathcal{L}^n\chi_\alpha)-\zeta_n\right|\le c_3(\vartheta' e^P)^n.$$

The next step is to compare $\sum_\alpha\omega_\alpha(\mathcal{L}^n\chi_\alpha)$ to $\sum_\lambda\lambda^n$. Here we use the decomposition of $\mathcal{L}$:

$$\omega_\alpha(\mathcal{L}^n\chi_\alpha)=\omega_\alpha\left(\sum_\lambda\lambda^n h_\lambda\nu_\lambda(\chi_\alpha)+R^n\chi_\alpha\right)=\sum_\lambda\lambda^n\omega_\alpha(h_\lambda)\nu_\lambda(\chi_\alpha)+\omega_\alpha(R^n\chi_\alpha).$$

In order to replace the term $\omega_\alpha(h_\lambda)$ by the function h_λ inside the functional ν_λ let us look at the error function

$$S_{\alpha,\lambda}(\alpha y)=\omega_\alpha(h_\lambda)\mathbb{1}(y)-h_\lambda(\alpha y)$$

where $\alpha y\in\Sigma^+$ is the variable. Clearly $S_{\alpha,\lambda}\in C_\vartheta$, although its norm will be too big for us in that space. We estimate

$$|S_{\alpha,\lambda}|_\infty\le\mathrm{var}_n h_\lambda\le|h_\lambda|_\vartheta\vartheta^n$$

and also

$$\mathrm{var}_k S_{\alpha,\lambda}\le\mathrm{var}_k h_\lambda\le|h_\lambda|_\vartheta\vartheta^k$$

for $k>n$ as $S_{\alpha,\lambda}$ defined only on the cylinder $U(\alpha)$. The function $S_\lambda=\sum_\alpha\chi_\alpha S_{\alpha,\lambda}$ is defined on the entire space Σ^+ and satisfies

$$|S_\lambda|_\infty\le\max_\alpha|S_{\alpha,\lambda}|_\infty\le c_4\vartheta^n$$

and

$$\mathrm{var}_k S_\lambda\le\begin{cases}2|S_\lambda|_\infty\le c_5\vartheta^n & \text{if } k\le n\\ \max_\alpha\mathrm{var}_k S_{\alpha,\lambda}\le c_5\vartheta^k & \text{if } k>n\end{cases}.$$

For ever λ we estimate now in a better function space, namely we take $\vartheta_\lambda<\frac{|\lambda|}{e^P}$ and note that λ is a discrete eigenvalue of $\mathcal{L}:C_{\vartheta_\lambda}\to C_{\vartheta_\lambda}$ (of the same multiplicity) and has the eigenfunction $h_\lambda\in C_{\vartheta_\lambda}$ and the eigenfunctional $\nu_\lambda\in C^*_{\vartheta_\lambda}$. Then

$$|S_\lambda|_{\vartheta_\lambda}=\sup_k\vartheta_\lambda^{-k}\mathrm{var}_k S_\lambda\le c_5\left(\frac{\vartheta}{\vartheta_\lambda}\right)^n$$

and therefore

$$\|S_\lambda\|_{\vartheta_\lambda} \leq c_6 \left(\frac{\vartheta' e^P}{|\lambda|}\right)^n.$$

Since $\omega_\alpha(h_\lambda)\nu_\lambda(\chi_\alpha) = \nu_\lambda(\chi_\alpha \omega_\alpha(h_\lambda)) = \nu_\lambda(\chi_\alpha(h_\lambda + S_{\alpha,\lambda}))$ we get

$$\sum_\alpha \omega_\alpha(h_\lambda)\nu_\lambda(\chi_\alpha) = \nu_\lambda\left(\sum_\alpha \chi_\alpha\left(h_\lambda + S_{\alpha,\lambda}\right)\right) = \nu_\lambda(h_\lambda) + \nu_\lambda(S_\lambda) = 1 + \nu_\lambda(S_\lambda)$$

where

$$|\nu_\lambda(S_\lambda)| \leq \|\nu_\lambda\|_{\vartheta_\lambda}\|S_\lambda\|_{\vartheta_\lambda} \leq c_6\|\nu_\lambda\|_\vartheta \left(\frac{\vartheta' e^P}{|\lambda|}\right)^n.$$

Summarising what we have done till now:

$$\begin{aligned}
\zeta_n &= \sum_\alpha \omega_\alpha(\mathcal{L}^n \chi_\alpha) + \mathcal{O}((\vartheta' e^P)^n) \\
&= \sum_\lambda \lambda^n \sum_\alpha \omega_\alpha(h_\lambda)\nu_\lambda(\chi_\alpha) + Q_n + \mathcal{O}((\vartheta' e^P)^n) \\
&= \sum_\lambda \lambda^n \left(1 + \mathcal{O}\left(\left(\frac{\vartheta e^P}{|\lambda|}\right)^n\right)\right) + Q_n + \mathcal{O}((\vartheta' e^P)^n) \\
&= \sum_\lambda \lambda^n + Q_n + \mathcal{O}((\vartheta' e^P)^n)
\end{aligned}$$

where $Q_n = \sum_\alpha \omega_\alpha(R^n \chi_\alpha)$ has still to be estimated.

To estimate the remainder term Q_n which comes from the essential spectrum let $\mathcal{P}$ be the projection operator for which $R^n = \mathcal{L}^n\mathcal{P} = \mathcal{P}\mathcal{L}^n$. For every α pick a point x_α so that $\alpha x_\alpha \in \Sigma^+$ depends only on the last symbol of α. Then

$$\omega_\alpha(R^n\chi_\alpha) = \omega_\alpha(\mathcal{P}\mathcal{L}^n\chi_\alpha) = \omega_\alpha(\mathcal{P}e^{f^n(\alpha\cdot)}) = e^{f^n(\alpha x_\alpha)}\omega_\alpha(\mathcal{P}\Phi_\alpha),$$

where $\Phi_\alpha(\alpha x) = e^{f^n(\alpha x) - f^n(\alpha x_\alpha)} \in C_\vartheta$ is defined on $U(\alpha)$. Clearly $|\Phi_\alpha|_\infty \leq e^{\frac{|f|_\vartheta}{1-\vartheta}} = \mathcal{O}(1)$. As $\mathcal{P}$ is a bounded operator on C_ϑ one has

$$|\omega_\alpha(\mathcal{P}\Phi_\alpha) - \mathcal{P}\Phi_\alpha(\alpha x_\alpha)| \leq \|\mathcal{P}\Phi_\alpha\|_\vartheta \vartheta^n$$

and thus

$$\begin{aligned}
\sum_\alpha \omega_\alpha(R^n\chi_\alpha) &= \sum_\alpha \omega_\alpha(\mathcal{P}\mathcal{L}^n\chi_\alpha) \\
&= \sum_\alpha e^{f^n(\alpha x_\alpha)}\left(\mathcal{P}\Phi_\alpha(\alpha x_\alpha) + \mathcal{O}(\vartheta^n)\right) \\
&= \sum_\alpha e^{f^n(\alpha x_\alpha)}\mathcal{P}\Phi_\alpha(\alpha x_\alpha) + \mathcal{O}((\vartheta e^P)^n).
\end{aligned}$$

In order to estimate $\sum_\alpha e^{f^n(\alpha x_\alpha)}\mathcal{P}\Phi_\alpha(\alpha x_\alpha)$ let us write $\Phi_\alpha = \sum_{i=1}^n \Phi_{i,\alpha}$, where we put

$$\Phi_{1,\alpha}(\alpha x) = e^{f\sigma^{n-1}(\alpha x) - f\sigma^{n-1}(\alpha x_\alpha)}$$

and then successively

$$\Phi_{i,\alpha}(\alpha x) = e^{\sum_{j=1}^{i}(f\sigma^{n-j}(\alpha x) - f\sigma^{n-j}(\alpha x_\alpha))} - e^{\sum_{j=1}^{i-1}(f\sigma^{n-j}(\alpha x) - f\sigma^{n-j}(\alpha x_\alpha))}$$

and notice that if we write $\alpha = \beta\gamma$ with $|\beta| = n - i$ the first $n - i$ symbols and $|\gamma| = i$ the last i symbols of α then $\Phi_{i,\alpha} = \Phi_{i,\gamma}$ is independent of the first $n - i$ symbols β. Since $\left|e^{f\sigma^{n-i}(\alpha x) - f\sigma^{n-i}(\alpha x_\alpha)} - 1\right| \le e^{\mathrm{var}_{k+i} f} - 1$ one has

$$\mathrm{var}_k \Phi_{i,\alpha}(\alpha x) \le e^{\sum_{j=1}^{i-1}(f\sigma^{n-j}(\alpha x) - f\sigma^{n-j}(\alpha x_\alpha))}\left(e^{\mathrm{var}_{k+i} f} - 1\right)$$

which implies $|\Phi_{i,\gamma}|_\vartheta \le c_7 \vartheta^i$ and also $\|\Phi_{i,\gamma}\|_\vartheta \le c_8 \vartheta^i$ and allows us to proceed as follows ($x_\alpha = x_\gamma$)

$$\begin{aligned} \sum_\alpha e^{f^n(\alpha x_\alpha)} \mathcal{P}\Phi_\alpha(\alpha x_\alpha) &= \sum_i \sum_{|\alpha|=n} e^{f^n(\alpha x_\alpha)} \mathcal{P}\Phi_{i,\alpha}(\alpha x_\alpha) \\ &= \sum_i \sum_{|\gamma|=i} e^{f^i(\gamma x_\alpha)} \sum_{|\beta|=n-i} e^{f^{n-i}(\beta\gamma x_\gamma)} \mathcal{P}\Phi_{i,\gamma}(\beta\gamma x_\gamma) \\ &= \sum_i \sum_{|\gamma|=i} e^{f^i(\gamma x_\alpha)} \sum_{|\beta|=n-i} \mathcal{L}^{n-i} \mathcal{P}\Phi_{i,\gamma}(\gamma x_\alpha). \end{aligned}$$

Since $|\mathcal{L}^{n-i}\mathcal{P}\Phi_{i,\gamma}| \le \|\mathcal{L}^{n-i}\mathcal{P}\|_\vartheta \|\Phi_{i,\gamma}\|_\vartheta \le c_9(\vartheta' e^P)^{n-i}\vartheta^i$ one obtains

$$\left|\sum_\alpha e^{f^n(\alpha x_\alpha)} \mathcal{P}\Phi_\alpha(\alpha x_\alpha)\right| \le c_{10} \sum_i \sum_\gamma e^{f^i(\gamma)} (\vartheta' e^P)^{n-i} \vartheta^i \le n c_{11} (\vartheta' e^P)^n$$

for all $\vartheta' > \vartheta$ which implies $|Q_n| \le c_{12}(\vartheta' e^P)^n$.

Combining all the estimates yields

$$\left|\zeta_n - \sum_\lambda \lambda^n\right| \le \mathrm{const.}(\vartheta' e^P)^n$$

for all $\vartheta' > \vartheta$ and therefore

$$\zeta(z) = \exp - \sum_n \frac{z^n}{n}\zeta_n = \exp - \sum_n \frac{z^n}{n}\left(\sum_\lambda \lambda^n + \mathcal{O}(\vartheta' e^P)^n\right) = \frac{1}{\prod_\lambda (1 - z\lambda)}\psi(z)$$

where $\psi(z)$ is analytic for $|z|\vartheta' e^P < 1 \forall \vartheta' > \vartheta$, i.e. for $|z| < \frac{1}{\vartheta e^P}$. ∎

10. Coboundaries

In this section we show how a two-sided function can be reduced to a one-sided function by adding a coboundary. Then we also show that two equilibrium states are equal if and only if the potentials differ by a constant and a coboundary.

Let us note that if μ is a σ-invariant probability measure on Σ^+ then it has an extension to the two-sided shift-space Σ by putting for cylinder sets ($k < \ell$)

$$\mu(U(x_{-k} \cdots x_\ell)) = \mu(\sigma^{-k} U(x_{-k} \cdots x_\ell))$$

using that σ is invertible on Σ. (As before we assume $A^n > 0$ for all large n.)

Theorem 78. *Let $g \in C_\vartheta(\Sigma)$. Then there exists a $u \in C_{\sqrt{\vartheta}}(\Sigma)$ so that $f = g + u - u \circ u$ depends only on positive coordinates, that is $f(x) = f(y)$ if $x_i = y_i$ for $i < 0$.*

Proof. For each symbol $a \in \{1, \dots, M\}$ we pick a left-infinite sequence $y^a = \cdots y_{-2}y_{-1}$ in Σ so that $A_{y_{-1}a} = 1$ i.e. $y^a a$ is an allowed sequence in Σ. Then we define $\pi : \Sigma \to \Sigma$ by $\pi(x) = y^{x_0}x_0x_1x_2\cdots$ for $x \in \Sigma$. Define

$$u = \sum_{j=0}^{\infty}(g \circ \sigma^j \pi - g \circ \sigma^j).$$

To show convergence of the sum note that $(\sigma^j \pi x)_i = (\sigma^j x)_i$ for all $|i| \leq j$. Hence

$$|u(x)| \leq \sum_{j=0}^{\infty} \mathrm{var}_j g \leq |g|_\vartheta \sum_{j=0}^{\infty} \vartheta^j < \infty$$

which in particular proves that $|u|_\infty < \infty$. To show that u belongs to $C_{\sqrt{\vartheta}}(\Sigma)$ we have to estimate $\mathrm{var}_k u$ for any k. To that end let $x, y \in \Sigma$ so that $x_i = y_i$ for all $|i| \leq k$. Then

$$\begin{aligned}
|u(x) - u(y)| &= \left|\sum_{j=0}^{\infty}(g\sigma^j\pi x - g\sigma^j x) - \sum_{j=0}^{\infty}(g\sigma^j\pi y - g\sigma^j y)\right| \\
&\leq \sum_{j=0}^{[k/2]} |g\sigma^j y - g\sigma^j x| + \sum_{j=0}^{[k/2]} |g\sigma^j\pi x - g\sigma^j\pi x| \\
&\qquad + \sum_{[k/2]+1}^{\infty} |g\sigma^j\pi x - g\sigma^j x| + \sum_{[k/2]+1}^{\infty} |g\sigma^j\pi y - g\sigma^j y| \\
&\leq \sum_{j=0}^{[k/2]} 2\mathrm{var}_{k-j} g + \sum_{[k/2]+1}^{\infty} 2\mathrm{var}_j g \\
&\leq 2|g|_\vartheta \sum_{j=0}^{[k/2]} \vartheta^{k-j} + 2|g|_\vartheta \sum_{[k/2]+1}^{\infty} \vartheta^j \\
&\leq c_1|g|_\vartheta \vartheta^{\frac{k}{2}}
\end{aligned}$$

for some c_1. Hence $\|u\|_{\sqrt{\vartheta}} < \infty$ and thus $u \in C_{\sqrt{\vartheta}}(\Sigma)$.

It remains to show that $f = g + u - u \circ u$ is independent of negative coordinates. One has for every M

$$\begin{aligned}
f(x) &= g(x) + \sum_{j=0}^{\infty}(g\sigma^j\pi x - g\sigma^j x) - \sum_{j=0}^{\infty}(g\sigma^j\pi\sigma x - g\sigma^{j+1}x) \\
&= g(x) - \sum_{j=0}^{M}(g\sigma^j x - g\sigma^{j+1}x) + \sum_{j=0}^{M}(g\sigma^j\pi x - g\sigma^j\pi\sigma x) \\
&\qquad + \sum_{j=M+1}^{\infty}(g\sigma^j\pi x - g\sigma^j x) - \sum_{j=M+1}^{\infty}(g\sigma^j\pi\sigma x - g\sigma^{j+1}x) \\
&= g(\pi x) + g\sigma^{M+1}x - g\sigma^{M+1}\pi x + \sum_{j=0}^{M}(g\sigma^{j+1}\pi x - g\sigma^j\pi\sigma x) \\
&\qquad + \sum_{j=M+1}^{\infty}(g\sigma^j\pi x - g\sigma^j x) - \sum_{j=M+1}^{\infty}(g\sigma^j\pi\sigma x - g\sigma^{j+1}x)
\end{aligned}$$

and for $M \to \infty$ we obtain

$$f = g \circ \pi + \sum_{j=0}^{\infty}(g \circ \sigma^{j+1} \circ \pi - g \circ \sigma^j \circ \pi \circ \sigma)$$

which is independent of negative coordinates since all terms involve the projection π. ∎

Recall that μ_f is a Gibbs state for $f \in C_\vartheta(\Sigma)$ is $\sigma^*\mu_f = \mu_f$ and there exists $C > 0, P_f \in \mathbb{R}$ so that

$$\frac{1}{C} \leq \frac{\mu_f(U(x_0x_1\cdots x_{n-1}))}{e^{f^n(x)-nP_f}} \leq C$$

for all $x \in U(x_0x_1\cdots x_{n-1})$ and for all n.

Theorem 79. *Let $f, g \in C_\vartheta(\Sigma)$ and μ_f, μ_g be respective Gibbs states. Then the following are equivalent:*
(I) $\mu_f = \mu_g$,
(II) there exists a constant K so that $f^n(x) = g^n(x) + nK$ for all periodic points x of period n,
(III) there exists a K and $u \in C_\vartheta(\Sigma)$ so that $f = g + u - u \circ u + K$,
(IV) there exists K, S so that $|f^n(x) - g^n(x) + nK| < S$ for all $x \in \Sigma$ and all $n \in \mathbb{N}$.

Proof. "(III)⇒(IV)" is obvious.
"(IV)⇒(I)": Assume

$$\frac{\mu_f(U(x_0\cdots x_{n-1}))}{e^{f^n(x)-nP_f}} \in \left[\frac{1}{C}, C\right], \qquad \frac{\mu_g(U(x_0\cdots x_{n-1}))}{e^{g^n(x)-nP_g}} \in \left[\frac{1}{C}, C\right]$$

and $g^n = nK + f^n + \mathcal{O}(S)$. Then

$$\frac{\mu_g(U(x_0\cdots x_{n-1}))}{e^{f^n(x)-nP_g+nK}} \in \left[\frac{1}{C}e^{-S}, Ce^{S}\right],$$

for all $x \in U(x_0 \cdots x_{n-1})$ and all n. Since $\mu_f(U(x_0 \cdots x_{n-1})) \geq C^{-1} e^{f^n(x) - nP_f}$ and $P_g - K = P_f$ (cf. section 9.4) we get

$$\mu_g(U(x_0 \cdots x_{n-1})) \geq C^2 e^S \mu_f(U(x_0 \cdots x_{n-1}))$$

and similarly $\mu_f(U(x_0 \cdots x_{n-1})) \geq C^2 e^S \mu_g(U(x_0 \cdots x_{n-1}))$. Hence μ_f, μ_g are equivalent. We now show that they are ergodic which then implies $\mu_f = \mu_g$. We use the characterisation that μ_f is ergodic if and only if for all $V, W \subset \Sigma$ of non-zero measures there exists a j so that $\mu_f(V \cap \sigma^{-j} W) > 0$ (see section 3). It is enough to assume that V, W are cylinder sets. Hence let $V = U(\alpha), W = U(\beta)$ for some words α, β in Σ. Let N be so that $A^N > 0$ i.e. any two symbols can be connected by a string of length $N - 1$. Take $j > N + |\alpha|$. Then there exists a word γ of length $|\gamma| = j - |\alpha| > N$ so that $\alpha\gamma\beta$ is an allowed word. Since

$$U(\alpha\gamma\beta) \subset U(\alpha) \cap \sigma^{-j} U(\beta)$$

we get that

$$\mu_f(U(\alpha) \cap \sigma^{-j} U(\beta)) \geq \mu_f(U(\alpha\gamma\beta)) \geq \frac{1}{C} e^{f^{j+|\beta|}(x) - (j+|\beta|)P_f} > 0$$

for any $x \in U(\alpha\gamma\beta)$. Since α, β were arbitrary, μ_f, μ_g are ergodic and thus equal.
"(I)⇒(II)": Since by assumption $\mu_f = \mu_g$ one has

$$\frac{1}{C^2} \leq \frac{e^{f^n(x) - nP_f}}{e^{g^n(x) - nP_g}} \leq C^2$$

for all $x \in \Sigma$ and n. Hence $f^n(x) - g^n(x) + n(P_g - P_f) = \mathcal{P}(1)$ and if x is periodic with period n, i.e. $\sigma^n x = x$, then also $\sigma^{kn} x = x$ for all integers k and therefore $f^{kn}(x) - g^{kn}(x) - knK = \mathcal{O}(1)$ where we put $K = P_g - P - f$. Since $f^{kn}(x) = kf^n(x), g^{kn}(x) = kg^n(x)$ we get

$$f^n(x) - g^n(x) - nK = \mathcal{O}\left(\frac{1}{k}\right).$$

Letting $k \to \infty$ yields $f^n(x) - g^n(x) - nK = 0$ for all periodic x of period n.
"(II)⇒(III)": This is the hard part of the theorem. Assume that $f^n(x) = g^n(x) + nK$ for some constant K and all periodic points x of period n and for all periods n. Put $\varphi = f - g + K$. Clearly $\varphi \in C_\vartheta(\Sigma)$ and $\varphi^n(x) = 0$ for all periodic x of period n. We have to find $u \in C_\vartheta(\Sigma)$ so that $\varphi = u \circ \sigma - u$ is a coboundary.

Let $x \in \Sigma$ be a transitive point, i.e. its orbit $\Gamma = \{\sigma^j x : j \in \mathbb{N}_0\}$ is dense in Σ. We have $\bar{\Gamma} = \Sigma$. Define $u : \Gamma \to \mathbb{R}$ as follows: if $y \in \Gamma$ and $j \geq 0$ is so that $\sigma^j x = y$, then we put

$$u(y) = u(\sigma^j x) = \varphi^j(x) = \sum_{k=0}^{j-1} \varphi(\sigma^k x)$$

We now claim u is continuous and has an extension to the full space Σ which lies in $C_\vartheta(\Sigma)$. To that end let $y, z \in \Gamma$ so that $y_i = z_i$ for $|i| \leq k$ for some $k \geq 1$ and estimate $u(z) - u(y)$. Let $m, n \geq 0$ be so that $y = \sigma^m x, z = \sigma^n x$ and assume (without loss of generality) that $n > m$. Put $w = \sigma^{-m}((x_m x_{m+1} \cdots x_{n-2} x_{n-1})^\infty)$ for the periodic point in Σ of period $n - m$ which agrees with x on the coordinates from m to $n - 1$. This point exists as $k \geq 1$. Evidently $w \notin \Gamma$ and in fact satisfies $w_i = x_i \forall i = m-k, \ldots, n+k$. Since by assumption $\varphi^{n-m}(w) = 0$ we can thus estimate

$$\begin{aligned}
|u(z) - u(y)| &= |\varphi^n(x) - \varphi^m(x)| \\
&= \left| \sum_{j=m}^{n-1} \varphi(\sigma^j x) \right| \\
&= \left| \sum_{j=m}^{n-1} \varphi(\sigma^j x) - \varphi^{n-m}(\sigma^m w) \right| \\
&\leq \sum_{j=m}^{n-1} \left| \varphi(\sigma^j x) - \varphi(\sigma^j w) \right| \\
&\leq \sum_{j=m}^{n-1} \mathrm{var}_{\min(j-(m-k), n+k-j)} \varphi \\
&\leq 2 \sum_{\ell=k}^{\infty} \mathrm{var}_\ell \varphi \\
&\leq 2|\varphi|_\vartheta \vartheta^k \frac{1}{1-\vartheta}
\end{aligned}$$

Hence $\mathrm{var}_k u|_\Gamma \leq c_1 \vartheta^k$ and therefore u has a continuous extension to $\bar{\Gamma} = \Sigma$ such that $|u|_\vartheta < \infty$ and therefore $u \in C_\vartheta(\Sigma)$.

To show that $\varphi = u \circ \sigma - u$ observe that for $y = \sigma^m x \in \Gamma$ one has $\sigma y = \sigma^{m+1} x$ and therefore

$$u(\sigma y) - u(y) = \varphi^{m+1}(x) - \varphi^m(x) = \sum_{k=0}^{m} \varphi(\sigma^k x) - \sum_{k=0}^{m-1} \varphi(\sigma^k x) = \varphi(\sigma^m x) = \varphi(y).$$

By continuity $\varphi = u \circ \sigma - u$ on Σ. ∎

11. References

References

[1] M Abadi: Exponential Approximation for Hitting Times in Mixing Stochastic Processes; Mathematical Physics Electronic Journal 7 (2001)

[2] S Bernstein: Sur l'extension du théorème limite du calcul des probabilités aux sommes du quantités dépendantes; Math. Ann. 97 (1926), 1–59

[3] P Billingsley: *Convergence in Probability Measures*; Wiley 1968

[4] X Bressaud: Subshifts on an infinite alphabet; Ergod. Th. & Dynam. Sys. 19 (1999), 1175-1200

[5] M Brin and A Katok: On local entropy; *Geometric Dynamics* 30–38, Springer LNM #1007, 1983

[6] A Broise: *Transformations dilatantes de l'intervalle et théorèmes limites*; Asterisque #238, 1996

[7] H Bruin and S Vaienti: Return times for unimodal maps; submitted to Forum Math.

[8] N Chernov: Limit theorems and Markov approximations for chaotic dynamical systems; Prob. Th. Rel. Fields **101** (1995), 321–362

[9] P Collet and A Galves: Repetition times for Gibbsian sources; Nonlinearity 12 (1999), 1225–1237

[10] P Collet, A Galves and B Schmitt: Fluctuations of repetition times for Gibbsian sources; Nonlinearity 12 (1999), 1225–1237

[11] P Doukhan: *Mixing: Properties and examples*; Lecture Notes in Statistics 85, Springer 1995
[12] P Ferrero, N Haydn and S Vaienti: Entropy fluctuations for parabolic maps; Nonlinearity 16 (2003), 1203–1218
[13] M Field, I Melbourne and A Török: Decay of Correlations, Central Limit Theorems and Approximations by Brownian Motion for Compact Lie Group Extensions; Ergod. Th. & Dynam. Syst. 23 (2003), 87–110
[14] A Galves and B Schmitt: Inequalities for hitting times in mixing dynamical systems; Random and Computational Dynamics 5 (1997), 337–348
[15] M Gordin: The central limit theorem for stationary processes; Soviet Math. Doklady 10 (1969), 1174–1176
[16] N Haydn: Meromorphic extension of the zeta function for Axiom A flows; Ergod. Th. & Dynam. Sys. **10** (1990), 347–360
[17] N Haydn and S Vaienti: Fluctuations of the metric entropy for mixing measures; Stochastics and Dynamics 4 (2004), 595–627
[18] N Haydn and S Vaienti: The distribution of the measure of cylinders for non-Gibbsian measures; *Complex Dynamics and Related Topics* 147–162, New Studies in Advanced Mathematics #5, 2003
[19] I A Ibragimov: Some limit theorems for stationary processes; Theory Prob. Appl. 7 (1962), 349–382
[20] I Kontoyiannis: Asymptotic Recurrence and Waiting Times for Stationary Processes; J. Theor. Prob. 11 (1998), 795–811
[21] C Liverani: Central Limit theorem for deterministic systems; Intern. Congress on Dyn. Syst., Montevideo 1995 (Proc. Research Notes in Math. Series), Pitman (1996), 56–75
[22] S Nagaev: Some limit theorems for stationary Markov chains; Theory Prob. Appl. 2 (1957), 378–406
[23] A Nobel and A Wyner: A Recurrence Theorem for Dependent Processes with Applications to Data Compression; IEEE Vol. 38 No. 5 (1992), 1561–1564
[24] Ornstein and Weiss; Entropy and Data Compression Schemes; IEEE Transactions on Information Theory 39 (1993), 78–83
[25] Ornstein and Weiss; Entropy and Recurrence Rates for Stationary Random Fields; IEEE Transactions on Information Theory 48 (6) (2002), 1694–97
[26] F Paccaut: *Propriétés Statistiques de Systèmes Dynamiques Non Markovian*; Thèse (Doctorat) Dijon 2000
[27] F Pène: Rates of Convergence in the CLT for Two-Dimensional Dispersive Billiards; Commun. Math. Phys. 225 (2002), 91–119
[28] V V Petrov: On a relation between an estimate of the remainder in the central limit theorem and the law of the iterated logarithm; Teor. Veroyatn. Primen. 11 (3), 514–518. English translation: Theor. Probab. Appl. 11(3), 454–458.
[29] W Philipp and W Stout: *Almost sure invariance principles for partial sums of weakly dependent random variables*; AMS Memoirs vol. 2 issue 2 No. 161, 1975
[30] Y Rinott and V Rotar: Normal approximations by Stein's method; Decis. Econ. Finance 23 (2000), 15–29
[31] A Röllins: A note on the exchangeability condition in Stein's method; Statist. Probab. Lett. 78 (2008), 1800–1806
[32] V I Romanovskii: *Discrete Markov Chains*; Wolters-Noordhoff Publishing Groningen 1970
[33] M Rosenblatt: A central limit theorem and a strong mixing condition; Proc. Nat. Acad. Sci. USA 42 (1956), 43–47
[34] M Rosenblatt: *Markov Processes. Structure and Asymptotic Behavior*; Springer, Grundlagen #184, 1971
[35] O Sarig: Thermodynamic formalism for countable Markov shifts; Ergod. Th. & Dynam. Syst. 19 (1999), 1565–1593
[36] C Stein: *Approximate Computation of Expectations*; IMS Lecture Notes #7, 1986
[37] L-S Young: Statistical properties of dynamical systems with some hyperbolicity: Ann. of Math. 147 (1998), 585–650

[38] A Yushkevich: On limit theorems connected with the concept of the entropy of Markov chains (in Russian); Uspehi Math. Nauk 8 (1953) 177–180