Notes on Ergodic Theory.

H. Bruin

May 8, 2023

Abstract

These are notes in the making for the course Ergodic Theory 1 & 2, Summer & Winter Semester 2017, University of Vienna

Contents

1 Notation

Throughout, (X, d) will be a metric space, possibly compact, and $T : X \to X$ will be a (piecewise) continuous map. The combination (X, T) defines dynamical systems by means of iteration. The **orbit** of a point $x \in X$ is the set

$$\mathrm{orb}(x) = \{x, T(x), T \circ T(x), \dots, \underbrace{T \circ \cdots \circ T}_{n \text{ times}}(x) =: T^n(x), \cdots\} = \{T^n(x) : n \geqslant 0\},$$

and if T is invertible, then $\mathrm{orb}(x) = \{T^n(x) : n \in \mathbb{Z}\}$ where the negative iterates are defined as $T^{-n} = (T^{inv})^n$. In other words, we consider $n \in \mathbb{N}$ (or $n \in \mathbb{Z}$) as discrete time, and $T^n(x)$ is the position the point x takes at time n.

Definition 1. *We call x a* **fixed point** *if $T(x) = x$;* **periodic** *if there is $n \geqslant 1$ such that $T^n(x) = x$;* **recurrent** *if $x \in \overline{\mathrm{orb}(x)}$.*

In general *chaotic* dynamical systems most orbits are more complicated than periodic (or quasi-periodic as the irrational rotation R_α discussed below). The behaviour of such orbits is hard to predict. Ergodic Theory is meant to help in predicting the behaviour of **typical** orbits, where typical means: almost all points x for some (invariant) measure μ.

To define measures properly, we need a σ-algebra $\mathcal{B}$ of "measurable" subsets. σ-algebra means that the collection $\mathcal{B}$ is closed under taking complements, countable unions and countable intersections, and also that $\varnothing, X \in \mathcal{B}$. Then a measure μ is a function $\mu : \mathcal{B} \to \mathbb{R}^+$ that is countably subadditive: $\mu(\cup_i A_i) \leqslant \sum_i \mu(A)_i$ (with equality if the sets A_i are pairwise disjoint). To makeour lives a bit easier, in these notes, we let $\mathcal{B}$ be the σ-algebra of Borel sets; this is the small σ-algebra that contains all open sets.

Example: For a subset $A \subset X$, define

$$\nu(A) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} 1_A \circ T^i x,$$

for the **indicator function** 1_A, assuming for the moment that this limit exists. We call this the visit frequency of x to the set A. We can compute

$$\begin{aligned}
\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} 1_A \circ T^i x &= \lim_{n\to\infty} \frac{1}{n} \left(\sum_{i=0}^{n-1} 1_A \circ T^{i+1} x + 1_A x - 1_A(T^n x) \right) \\
&= \lim_{n\to\infty} \frac{1}{n} \left(\sum_{i=0}^{n-1} 1_{T^{-1}A} \circ T^i x + 1_A x - 1_A(T^n x) \right) \\
&= \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} 1_{T^{-1}A} \circ T^i x = \nu(T^{-1}(A))
\end{aligned}$$

That is, visit frequency measures, when well-defined, are **invariant** under the map. This allows us to use invariant measure to make statistical predictions of what orbit do "on average".

Let $\mathcal{B}_0$ be the collection of subsets $A \in \mathcal{B}$ such that $\mu(A) = 0$, that is: $\mathcal{B}_0$ are the **null-sets** of μ. We say that an event happens almost surely (a.s.) or μ-almost everywhere (μ-a.e.) if it is true for all $x \in X \setminus A$ for some $A \in \mathcal{B}_0$.

A measure μ on $(X, T, \mathcal{B})$ is called

- non-singular if $A \in \mathcal{B}_0$ implies $T^{-1}(A) \in \mathcal{B}_0$.
- non-atomic if $\mu(\{x\}) = 0$ for every $x \in X$
- T-invariant if $\mu(T^{-1}(A)) = \mu(A)$ for all $A \in \mathcal{B}$.
- finite if $\mu(X) < \infty$. In this case we can always rescale μ so that $\mu(X) = 1$, *i.e.,* μ is a probability measure.
- σ-finite if there is a countable collection X_i such that $X = \cup_i X_i$ and $\mu(X_i) \leqslant 1$ for all i. In principle, finite measures are also σ-finite, but we would like to reserve the term σ-finite only for infinite measures (*i.e.,* $\mu(X) = \infty$).

Lemma 1. *Let $T : X \to X$ be a continuous map on a compact space X. Then μ is T-invariant if and only if*

$$\int_X f d\mu = \int_X f \circ T \ d\mu$$

for every $f \in C(X)$. (Here $C(X)$ is the space of all **continuous** *functions on X, equipped with the norm $\| \ \|_\infty$.)*

Proof. Assume that μ is T-invariant and $A \in \mathcal{B}$ is some measurable set. Then

$$\int_X 1_A \circ T \ d\mu = \mu(T^{-1}A) = \mu(A) = \int_X 1_A \ d\mu.$$

A similar expression works for linear combinations of indicator sets. Now if f is continuous, and $\varepsilon > 0$ is arbitrary, then due to uniform continuity, there is a partition of X into measurable sets A_j and factors $a_j \in \mathbb{R}$ such that for $g = \sum_j a_j 1_{A_j}$, we have $\|f - g\|_\infty < \varepsilon$. Now

$$\begin{aligned} \left| \int f \circ T \ d\mu - \int f \ d\mu \right| &= \left| \int (f-g) \circ T \ d\mu - \int f - g \ d\mu \right. \\ &\quad \left. + \sum_j a_j \int 1_{A_j} \circ T \ d\mu - \sum_j a_j \int 1_{A_j} \ d\mu \right| \\ &\leq 2\varepsilon \end{aligned}$$

Since ε is arbitrary, $\int f \circ T \, d\mu = \int f \, d\mu$.

Conversely, for every closed set A and $\varepsilon > 0$, we can find a function $f \in C(X)$ such that $f \equiv 1$ on A and $\int |f - 1_A| \, d\mu < \varepsilon$ as well as $\int |f - 1_A| \circ T \, d\mu < \varepsilon$. Then

$$\begin{aligned} |\mu(T^{-1}A) - \mu(A)| &= |\int 1_A \circ T \, d\mu - \int 1_A \, d\mu| \\ &= |\int (f - 1_A) \circ T \, d\mu - \int (f - 1_A) \, d\mu| \leq 2\varepsilon. \end{aligned}$$

Since ε is arbitrary, we get $\mu(T^{-1}A) = \mu(A)$. Because the closed sets generate the Borel σ-algebra, this property carries over to all $A \in \mathcal{B}$. □

2 What are the invariant measures of the cat map?

The following example is called Arnol'd's cat map.

Example: Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be defined by

$$T \begin{pmatrix} x \\ y \end{pmatrix} = M \begin{pmatrix} x \\ y \end{pmatrix} \qquad \text{for matrix} \qquad M = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}.$$

Note that T is a bijection of $\mathbb{R}^2$, with 0 as single fixed point. Therefore the Dirac measure δ_0 is T-invariant. However, also Lebesgue measure m is invariant because (using coordinate transformation $x = T^{-1}(y)$)

$$m(T^{-1}A) = \int_{T^{-1}A} dm(x) = \int_A \det(M^{-1}) dm(y) = \int_A \frac{1}{\det(M)} dm(y) = m(A)$$

because $\det(M) = 1$. This is a general fact: If $T : \mathbb{R}^n \to \mathbb{R}^n$ is a bijection with Jacobian $J = |\det(DT)| = 1$, then Lebesgue measure is preserved. However, Lebesgue measure is not a probability measure (it is σ-finite). In the above case of the integer matrix with determinant 1, T preserves (and is a bijection) on $\mathbb{Z}^2$. Therefore we can factor out over $\mathbb{Z}^2$ and obtain a map on the two-torus $\mathbb{T}^2 = \mathbb{R}^2/\mathbb{Z}^2$:

$$\begin{aligned} T \ : \ & \mathbb{T}^2 \to \mathbb{T}^2 \\ & \begin{pmatrix} x \\ y \end{pmatrix} \mapsto M \begin{pmatrix} x \\ y \end{pmatrix} \pmod 1 \end{aligned}$$

This map is called Arnol'd's cat-map, and it preserves Lebesgue measure, which on $\mathbb{T}^2$ is a probability measure.

A special case of the above is:

Proposition 1. *If $T : U \subset \mathbb{R}^n \to U$ is an isometry (or piecewise isometric bijection), then T preserves Lebesgue measure.*

Let $\mathcal{M}(X, T)$ denote the set of T-invariant Borel[1] probability measures. In general, there are always invariant measures.

Theorem 1 (Krylov-Bogol'ubov)**.** *If $T : X \to X$ is a continuous map on a nonempty compact metric space X, then $\mathcal{M}(X, T) \neq \varnothing$.*

Proof. Let ν be any probability measure and define Césaro means:

$$\nu_n(A) = \frac{1}{n} \sum_{j=0}^{n-1} \nu(T^{-j}A),$$

these are all probability measures. The collection of probability measures on a compact metric space is known to be compact in the weak* topology, *i.e.,* there is limit probability measure μ and a subsequence $(n_i)_{i \in \mathbb{N}}$ such that for every continuous function $\psi : X \to \mathbb{R}$:

$$\int_X \psi \, d\nu_{n_i} \to \int \psi \, d\mu \text{ as } i \to \infty. \tag{1}$$

On a metric space, we can, for any $\varepsilon > 0$ and closed set A, find a continuous function $\psi_A : X \to [0, 1]$ such that $\psi_A(x) = 1$ if $x \in A$ and

$$\mu(A) \leqslant \int_X \psi_A d\mu \leqslant \mu(A) + \varepsilon \text{ and } \mu(T^{-1}A) \leqslant \int_X \psi_A \circ T \, d\mu \leqslant \mu(T^{-1}A) + \varepsilon.$$

Here we use outer regularity of the measure μ: $\mu(A) = \inf\{\mu(G) : G \supset A \text{ is open}\}$. We take $G \supset A$ so small that $\mu(G) - \mu(A) < \varepsilon$ and make sure that $\psi_A = 0$ for all $x \notin G$. Note that it is important that A is closed, because if there exists $a \in \partial A \setminus A$, then the above property fails for $\mu = \delta_a$.

By Lemma 1 and the definition of μ

$$\begin{aligned}
|\mu(T^{-1}(A)) - \mu(A)| &\leqslant \left| \int \psi_A \circ T \, d\mu - \int \psi_A \, d\mu \right| + \varepsilon \\
&= \lim_{i \to \infty} \left| \int \psi_A \circ T \, d\nu_{n_i} - \int \psi_A \, d\nu_{n_i} \right| + \varepsilon \\
&= \lim_{i \to \infty} \frac{1}{n_i} \left| \sum_{j=0}^{n_i - 1} \left(\int \psi_A \circ T^{j+1} \, d\nu - \int \psi_A \circ T^j \, d\nu \right) \right| + \varepsilon \\
&\leqslant \lim_{i \to \infty} \frac{1}{n_i} \left| \int \psi_A \circ T^{n_i} \, d\nu - \int \psi_A \, d\nu \right| + \varepsilon \\
&\leqslant \lim_{i \to \infty} \frac{2}{n_i} \|\psi_A\|_\infty + \varepsilon = \varepsilon.
\end{aligned}$$

[1]that is, sets in the σ-algebra of sets generated by the open subsets of X.

Since $\varepsilon > 0$ is arbitrary, $\mu(T^{-1}(A)) = \mu(A)$. The closed sets generate the Borel sets, so $\mu(T^{-1}(A)) = \mu(A)$ for arbitrary Borel sets too. $\square$

3 Ergodicity and unique ergodicity

Definition 2. *A measure is called* **ergodic** *if* $T^{-1}(A) = A \pmod{\mu}$ *for some* $A \in \mathcal{B}$ *implies that* $\mu(A) = 0$ *or* $\mu(A^c) = 0$.

Here $\mod \mu$ means "up to a set of μ-measure zero. Here specifically, it means thatthe symmetric difference has measure $\mu(A \triangle T^{-1}A) = 0$.

Proposition 2. *The following are equivalent:*

(i) μ *is ergodic;*

(ii) If $\psi \in L^1(\mu)$ *is* T*-invariant,* i.e., $\psi \circ T = \psi$ μ*-a.e., then* ψ *is constant* μ*-a.e.*

Proof. (i) $\Rightarrow$ (ii): Let $\psi : X \to \mathbb{R}$ be T-invariant μ-a.e., but not constant. Thus there exists $a \in \mathbb{R}$ such that $A := \psi^{-1}((-\infty, a])$ and $A^c = \psi^{-1}((a, \infty))$ both have positive measure. By T-invariance, $T^{-1}A = A \pmod{\mu}$, and we have a contradiction to ergodicity.
(ii) $\Rightarrow$ (i): Let A be a set of positive measure such that $T^{-1}A = A$. Let $\psi = 1_A$ be its indicator function; it is T-invariant because A is T-invariant. By (ii), ψ is constant μ-a.e., but as $\psi(x) = 0$ for $x \in A^c$, it follows that $\mu(A^c) = 0$. $\square$

The rotation $R_\alpha : \mathbb{S}^1 \to \mathbb{S}^1$ is defined as $R_\alpha(x) = x + \alpha \pmod{1}$.

Theorem 2 (Poincaré)**.** *If* $\alpha \in \mathbb{Q}$*, then every orbit is periodic.*

If $\alpha \notin \mathbb{Q}$*, then every orbit is dense in* $\mathbb{S}^1$*. In fact, for every interval* J *and every* $x \in \mathbb{S}^1$*, the visit frequency*

$$v(J) := \lim_{n\to\infty} \frac{1}{n} \#\{0 \leqslant i < n : R_\alpha^i(x) \in J\} = |J|.$$

Proof. If $\alpha = \frac{p}{q}$, then clearly

$$R_\alpha^q(x) = x + q\alpha \pmod{1} = x + q\frac{p}{q} \pmod{1} = x + p \pmod{1} = x.$$

Conversely, if $R_\alpha^q(x) = x$, then $x = x + q\alpha \pmod{1}$, so $q\alpha = p$ for some integer p, and $\alpha = \frac{p}{q} \in \mathbb{Q}$.

Therefore, if $\alpha \notin \mathbb{Q}$, then x cannot be periodic, so its orbit is infinite. Let $\varepsilon > 0$. Since $\mathbb{S}^1$ is compact, there must be $m < n$ such that $0 < \delta := d(R_\alpha^m(x), R_\alpha^n(x)) < \varepsilon$. Since R_α is an isometry, $|R_\alpha^{k(n-m)}(x) - R_\alpha^{(k+1)(n-m)}(x)| = \delta$ for every $k \in \mathbb{Z}$, and $\{R_\alpha^{k(n-m)}(x) : k \in \mathbb{Z}\}$ is a collection of points such that every two neighbours are exactly δ apart. Since $\varepsilon > \delta$ is arbitrary, this shows that $\text{orb}(x)$ is dense, but we want to prove more.

Let $J_\delta^0 = [R_\alpha^m(x), R_\alpha^n(x))$ and $J_\delta^k = R_\alpha^{k(n-m)}(J_\delta)$. Then for $K = \lfloor 1/\delta \rfloor$, $\{J_\delta^k\}_{k=0}^K$ is a cover $\mathbb{S}^1$ of adjacent intervals, each of length δ, and $R_\alpha^{j(n-m)}$ is an isometry from J_δ^i to J_δ^{i+j}. Therefore the visit frequencies

$$\underline{v}_k = \liminf_n \frac{1}{n} \#\{0 \leqslant i < n : R_\alpha^i(x) \in J_\delta^k\}$$

are all the same for $0 \leqslant k \leqslant K$, and together they add up to at most $1 + \frac{1}{K}$. This shows for example that

$$\frac{1}{K+1} \leqslant \underline{v}_k \leqslant \overline{v}_k := \limsup_n \frac{1}{n} \#\{0 \leqslant i < n : R_\alpha^i(x) \in J_\delta^k\} \leqslant \frac{1}{K},$$

and these inequalities are **independent** of the point x. Now an arbitrary interval J can be covered by $\lfloor |J|/\delta \rfloor + 2$ such adjacent J_δ^k, so

$$v(J) \leqslant \left(\frac{|J|}{\delta} + 2 \right) \frac{1}{K} \leqslant (|J|(K+1) + 2) \frac{1}{K} \leqslant |J| + \frac{3}{K}.$$

A similar computation gives $v(J) \geqslant |J| - \frac{3}{K}$. Now taking $\varepsilon \to 0$ (hence $\delta \to 0$ and $K \to \infty$), we find that the limit $v(J)$ indeed exists, and is equal to $|J|$. $\square$

Definition 3. *A transformation (X, T) is called* **uniquely ergodic** *if there is exactly one invariant probability measure.*

The above proof shows that Lebesgue measure is the only invariant measure if $\alpha \notin \mathbb{Q}$, so $(\mathbb{S}^1, R_\alpha)$ is uniquely ergodic. However, there is a missing step in the logic, in that we didn't show yet that Lebesgue measure is ergodic. This will be shown in Example 4 and also Theorem 10.

Questions: Does R_α preserve a σ-finite measure? Does R_α preserve a **non-atomic** σ-finite measure?

Theorem 3. *[Oxtoby's Theorem] Let X be a compact space and $T : X \to X$ continuous. A transformation (X, T) is uniquely ergodic if and only if, for every continuous function, the Birkhoff averages $\frac{1}{n}\sum_{i=0}^{n-1} f \circ T^i(x)$ converge* **uniformly** *to a constant function.*

Remark 1. *Every continuous map on a compact space has an invariant measure, as we showed in Theorem 1. Theorem 10 later on shows that if there is only one invariant measure, it has to be ergodic as well.*

Proof. If μ and ν were two different ergodic measures, then we can find a continuous function $f : X \to \mathbb{R}$ such that $\int f d\mu \neq \int f d\nu$. Using the Ergodic Theorem for both measures (with their own typical points x and y), we see that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f \circ T^k(x) = \int f d\mu \neq \int f d\nu = \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f \circ T^k(y),$$

so there is not even convergence to a constant function.

Conversely, we know by the Ergodic Theorem that $\lim_n \frac{1}{n} \sum_{k=0}^{n-1} f \circ T^k(x) = \int f d\mu$ is constant μ-a.e. But if the convergence is not uniform, then there is a sequence $(y_i) \subset X$ and $(n_i) \subset \mathbb{N}$, such that $B := \lim_i \frac{1}{n_i} \sum_{k=0}^{n_i-1} f \circ T^k(y_i) \neq \int_X f\, d\mu$. Define probability measures $\nu_i := \frac{1}{n_i} \sum_{k=0}^{n_i-1} \delta_{T^k(x_i)}$. This sequence (ν_i) has a weak accumulation points ν which is shown to be T-invariant measures in the same way as in the proof of Theorem 1. But $\nu \neq \mu$ because $\int f\, d\nu = B \neq \int f\, d\mu$. Hence (X, T) cannot be uniquely ergodic. □

4 The Ergodic Theorem

Theorem 2 is an instance of a very general fact observed in ergodic theory:

Space Average = Time Average (for typical points).

This is expressed in the

Theorem 4 (Birkhoff Ergodic Theorem). *Let μ be a probability measure and $\psi \in L^1(\mu)$. Then the* **ergodic average**

$$\psi^*(x) := \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \psi \circ T^i(x)$$

exists μ-a.e., and ψ^* is T-invariant, i.e., *$\psi^* \circ T = \psi^*$ μ-a.e. If in addition μ is ergodic then*

$$\psi^* = \int_X \psi\, d\mu \qquad \mu\text{-a.e.} \tag{2}$$

Remark 2. *A point $x \in X$ satisfying* (2) *is called* **typical** *for μ. To be precise, the set of μ-typical points also depends on ψ, but for different functions $\psi, \tilde{\psi}$, the (μ, ψ)-typical points and $(\mu, \tilde{\psi})$-typical points differ only on a null-set.*

Corollary 1. *Lebesgue measure is the only R_α-invariant probability measure.*

Proof. Suppose R_α had two invariant measures, μ and ν. Then there must be an interval J such that $\mu(J) \neq \nu(J)$. Let $\psi = 1_J$ be the indicator function; it will belongs to $L^1(\mu)$

and $L^1(\nu)$. Apply Birkhoff's Ergodic Theorem for some μ-typical point x and ν-typical point y. Since their visit frequencies to J are the same, we have

$$\begin{aligned}\mu(J) &= \int \psi \, d\mu = \lim_n \frac{1}{n} \#\{0 \leqslant i < n : R_\alpha(x) \in J\} \\ &= \lim_n \frac{1}{n} \#\{0 \leqslant i < n : R_\alpha(y) \in J\} = \int \psi \, d\nu = \nu(J),\end{aligned}$$

a contradiction to μ and ν being different. □

If $0.x_1x_2x_3\dots$ is the decimal expansion of some $x \in [0,1]$, you would expect that "typically", all digits $0, 1, \dots, 9$ appear with the same frequency. In fact, so should all blocks of N digits, that is: for every $a_1 \dots a_N \in \{0, 1, \dots, 9\}^N$, the frequency

$$\lim_{n\to\infty} \frac{1}{n} \#\{0 \leq i < n : x_i + 1 \dots x_{i+N} = a_1 \dots a_N\} = 10^{-N}.$$

If a number $x \in [0,1]$ has this property, it is called a **normal number**; they are in way the most random numbers with the least special structure that one can hope for. Borel proved in 1909:

Theorem 5. *Lebesgue-a.e. $x \in [0,1]$ is normal. In fact, this results holds for every base $b \in \{2, 3, 4, \dots\}$.*

Borel's theorem predates Birkhoff's theorem (1931), but this theorem gives a very short way of proving Borel's result.

Proof. Fix the base $b \in \{2, 3, 4, \dots\}$ and let $T : [0,1] \to [0,1]$ be defined as $T(x) = bx \pmod 1$. It has b branches with domains denoted $[a]$, $a \in \{0, \dots b-1\}$. Lebesgue measure μ is invariant and ergodic (we prove that somewhere else). Take $a_1 \dots a_N \in \{0, \dots, b-1\}^N$ arbitrary, and define the cylinder set $[a_1 \dots a_N] = \{x \in [0,1] : T^{k-1}(x) \in [a_k]\}$. That means that $x \in [a_1 \dots a_N]$ if its b-nary expansion starts with $0.a_1 \dots a_N$.

The indicator function $1_{[a_1 \dots a_N]} \in L^1(\mu)$, so by Theorem 4,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} 1_{[a_1 \dots a_N]} \circ T^k(x) = \int_0^1 1_{[a_1 \dots a_N]} \, dx = b^{-N}$$

for μ-a.e. x. This concludes the proof. □

Now that we know that almost every $x \in [0,1]$ is normal, it is tempting to find one such number explicitly. That is not easy! The standard example is

$$x = 0.12345678910111213141516171819202122\dots$$

This x is known as Champernowne's number, but it doesn't look random at all! Similar normal numbers can be obtained by concatenating the primes $0.23571113\dots$ (Copeland & Erdös, 1946) or the squares $0.149162536\dots$ (Besicovich, 1953).

Now we start with the proof of Theorem 4. The **Koopman operator** $U_T : L^1(\mu) \to L^1(\mu)$ is defined as $U_T f = f \circ T$. Clearly U_T is linear and **positive**, i.e., $f \geq 0$ implies $U_T f \geq 0$. For the next result, we write the **ergodic sums** as

$$S_n = S_n f = \sum_{k=0}^{n-1} f \circ T^k \quad \text{ and } quad S_0 \equiv 0.$$

Theorem 6 (Maximal Ergodic Theorem)**.** *Let $(X, T, \mathcal{B}, \mu)$ be a probability measure preserving dynamical system. Take $M_N = \max\{S_n : 0 \leq n \leq N\}$. Then*

$$\int_{A_N} f \, d\mu \geq 0 \qquad \textit{for } A_N = \{x \in X : M_N(x) > 0\}.$$

Proof. Clearly $M_N \geq S_n$ for all $0 \leq n \leq N$ and by positivity of the Koopman operator, also $U_T M_N \geq U_T S_n$. Add f on both sides: $U_T M_N + f \geq U_T S_n + f = S_{n+1}$. For $x \in x \in A_N$, this means

$$\begin{aligned} U_T M_N(x) + f(x) &\geq \max_{1 \leq n \leq N} S_n(x) \\ &= \max_{0 \leq n \leq N} S_n(x) \qquad \text{since } S_0 = 0 \text{ and } M_N(x) > 0 \\ &= M_N(x). \end{aligned}$$

Therefore $f \geq M_N - U_T M_N$ on A_N, and

$$\begin{aligned} \int_{A_N} f \, d\mu &\geq \int_{A_N} M_N \, d\mu - \int_{A_N} U_T M_N \, d\mu \\ &= \int_X M_N \, d\mu - \int_{A_N} U_T M_N \, d\mu \qquad \text{because } M_N = 0 \text{ outside } A_N \\ &= \int_X M_N \, d\mu - \int_X U_T M_N \, d\mu \qquad \text{because } U_T M_N \geq M_N \geq 0 \\ &= 0 \qquad \text{by } T\text{-invariance of } \mu. \end{aligned}$$

This completes the proof. □

Remark 3. *In fact, the only property of U_T we really need is that U_T is positive and $\|U_T\| = 1$. This follows by T-invariance of μ, because*

$$\|U_T f\|_1 = \int_X |U_T f| \, d\mu = \int_X |f| \circ T \, d\mu = \int_X |f| \, d\mu = \|f\|_1.$$

Lemma 2. *Let $(X, T, \mathcal{B}, \mu)$ be a probability measure preserving dynamical system, and $E \subset X$ a T-invariant subset. Let $B_\alpha := \{x \in X : \sup_n \frac{1}{n} S_n g(x) > \alpha\}$. Then $\int_{B_\alpha \cap E} g \, d\mu \geq \alpha \mu(B_\alpha \cap E)$.*

Proof. If $\mu(E) = 0$ then there is nothing to prove. So assume that $\mu(E) > 0$ and let $\mu_E = \frac{1}{\mu(E)} \mu|_E$ be a new invariant (because E is a T-invariant set) probability measure. Take $f = g - \alpha$, so $B_\alpha = \cup_N A_N$ with the notation of Theorem 6. Note also that $A_N \subset A_{N+1}$ for all N. Therefore for each $\varepsilon > 0$ there exists $N \in \mathbb{N}$ such that $\int_{B_\alpha} f \, d\mu_E \geq \int_{A_N} f d\mu_E \geq -\varepsilon$. Since ε is arbitrary, $\int_{B_\alpha} f \, d\mu_E \geq 0$. Adding α again we have $\int_{B_\alpha} g \, d\mu_E = \int_{B_\alpha} f + \alpha \, d\mu_E \geq \alpha \mu_E(B_\alpha \cap E)$. Finally, multiply everything by $\mu(E)$ to get the required result. $\square$

Proof of Theorem 4. Define

$$\overline{\psi} = \limsup_{n \to \infty} \frac{1}{n} S_n \psi \quad \text{and} \quad \underline{\psi} = \liminf_{n \to \infty} \frac{1}{n} S_n \psi.$$

Since $|\frac{n+1}{n} \frac{1}{n+1} S_{n+1} \psi - \frac{1}{n} S_n \psi \circ T| = \frac{1}{n} |f(x)| \to 0$ as $n \to \infty$, we have $\overline{\psi} \circ T = \overline{\psi}$ and similarly $\underline{\psi} \circ T = \underline{\psi}$. We want to show that $\overline{\psi} = \underline{\psi}$ μ-a.e.

Let

$$E_{\alpha,\beta} = \{x \in X : \psi_*(x) < \beta, \alpha < \psi^*(x)\}$$

Then $E_{\alpha,\beta}$ is T-invariant, and

$$\{x \in X : \psi_*(x) < \psi^*(x)\} = \bigcup_{\alpha, \beta \in \mathbb{Q}, \beta < \alpha} E_{\alpha,\beta}.$$

This is a countable union, and therefore it suffices to show that $\mu(E_{\alpha,\beta}) = 0$ for every pair of rationals $\beta < \alpha$. Write $B_\alpha := \{x \in X : \sup_n \frac{1}{n} S_n \psi(x) > \alpha\}$ as in Lemma 2. Since $E_{\alpha,\beta} = E_{\alpha,\beta} \cap B_\alpha$, this corollary gives

$$\int_{E_{\alpha,\beta}} \psi \, d\mu = \int_{E_{\alpha,\beta} \cap B_\alpha} \psi \, d\mu \geq \alpha \mu(E_{\alpha,\beta} \cap B_\alpha) = \alpha \mu(E_{\alpha,\beta}).$$

We repeat this argument replacing ψ, α, β by $-\psi, -\alpha, -\beta$. Note that $(-\psi)^* = -\psi_*$ and $(-\psi)_* = -\psi^*$. This gives

$$\int_{E_{\alpha,\beta}} \psi \, d\mu \leq \beta \mu(E_{\alpha,\beta}).$$

Since $\beta < \alpha$, this can only be true if $\mu(E_{\alpha,\beta}) = 0$. Therefore $\overline{\psi} = \underline{\psi} = \psi^*$, i.e., the $\limsup / \liminf$ is actually a $\lim$ μ-a.e.

The next step is to show that $\overline{\psi} \in L^1(\mu)$. **Fatou's Lemma** (from Measure Theory) states that

If $(g_n)_{n\in\mathbb{N}}$ are non-negative $L^1(\mu)$-functions and $g(x) = \liminf_n g_n(x)$, then $g \in L^1(\mu)$ and $\int_X g\, d\mu \leq \liminf_n \int_X g_n\, d\mu$.

Here we apply this to $g_n = |\frac{1}{n}S_n\psi|$, which belong to $L^1(\mu)$ because (by T-invariance) $\int_X |\frac{1}{n}S_n\psi|\, d\mu \leq \frac{1}{n}\sum_{k=0}^{n-1}\int_X |\psi|\circ T^k\, d\mu = \int |\psi|\, d\mu < \infty$. Hence $\int_X |\psi^*|\, d\mu \leq \liminf_n \int_X |\psi|\, d\mu < \infty$.

Next, we need to show that $\int \psi^*\, d\mu = \int \psi\, d\mu$ (so without absolute value signs). Take

$$D_{k,n} = \{x \in X : \frac{k}{n} \leq \psi^*(x) < \frac{k+1}{n}\}.$$

Then $D_{k,n}$ is T-invariant, and $\cup_{k\in\mathbb{Z}} D_{k,n} = X$. For $\varepsilon > 0$ small, $D_{k,n} \cap B_{\frac{k}{n}-\varepsilon} = D_{k,n}$. Therefore Lemma 2 gives

$$\int_{D_{k,n}} \psi\, d\mu = \int_{D_{k,n}\cap B_{\frac{k}{n}-\varepsilon}} \psi\, d\mu \geq (\frac{k}{n} - \varepsilon)\mu(D_{k,n} \cap B_{\frac{k}{n}-\varepsilon}) = (\frac{k}{n} - \varepsilon)\mu(D_{k,n}).$$

Since ε is arbitrary, we have $\frac{k}{n}\mu(D_{k,n}) \leq \int_{D_{k,n}} \psi\, d\mu$. Therefore

$$\int_{D_{k,n}} \psi^*\, d\mu \leq \frac{k+1}{n}\mu(D_{k,n}) \leq \frac{1}{n}\mu(D_{k,n}) + \int_{D_{k,n}} \psi\, d\mu.$$

Summing over all $k \in \mathbb{Z}$, we find $\int_X \psi^*\, d\mu \leq \frac{1}{n} + \int_X \psi\, d\mu$, and since $n \in \mathbb{N}$ is arbitrary, $\int_X \psi^*\, d\mu \leq \int_X \psi\, d\mu$, Applying the same argument to $-\psi$, we find $\int_X \psi_*\, d\mu \geq \int_X \psi\, d\mu$. Hence $\int_X \psi_* = \int_X \psi\, d\mu$.

Finally, if μ is ergodic, the T-invariant function ψ^* has to be constant μ-a.e., so $\psi^* = \int \psi\, d\mu$. $\square$

Theorem 7 (The L^p Ergodic Theorem)**.** *Let $(X, T, \mathcal{B}, \mu)$ be a probability measure preserving dynamical system. If μ is ergodic, and $\psi \in L^p(\mu)$ for some $1 \leq p < \infty$ then there exists $\psi^* \in L^p(\mu)$ with $\psi^* \circ T = \psi^*$ μ-a.e. such that*

$$\|\frac{1}{n}S_n\psi - \psi^*\|_p \to 0 \text{ as } n \to \infty.$$

This is a generalisation of Von Neumann's L^2 version of the Ergodic Theorem, which predates[2] Birkhoff's Theorem, but nowadays, it is usually proved as a corollary of the pointwise ergodic theorem, and that is also what we do in the proof below.

[2]named after George Birkhoff (1884–1944). There was a controversy on priority of the Ergodic Theorem: John von Neumann was earlier in proving his L^1-version, but Birkhoff delayed its publication until after the appearance of his own paper.

Proof. First assume that ψ is bounded (and hence in $L^p(\mu)$. By Theorem 4, there is ψ^* such that $\frac{1}{n}S_n\psi(x) \to \psi^*(x)$ μ-a.e., and ψ^* is bounded (and hence in $L^p(\mu)$ too). In particular, $|\frac{1}{n}S_n\psi(x) - \psi^*(x)|^p \to 0 \qquad \mu$-a.e. By the Bounded Convergence Theorem, we can swap the limit and the integral:

$$\begin{aligned}\lim_{n\to\infty} \|\frac{1}{n}S_n\psi - \psi^*\|_p &= \lim_{n\to\infty} \left(\int_X |\frac{1}{n}S_n\psi(x) - \psi^*(x)|^p \, d\mu\right)^{1/p} \\ &= \left(\int_X \lim_{n\to\infty} |\frac{1}{n}S_n\psi(x) - \psi^*(x)|^p \, d\mu\right)^{1/p} = 0.\end{aligned}$$

In particular, $\frac{1}{n}S_n\psi$ is a Cauchy sequence in $\| \ \|_p$, so for every $\varepsilon > 0$ there is $N = N(\varepsilon, \psi)$ such that

$$\|\frac{1}{m}S_m\psi - \frac{1}{n}S_n\psi\|_p < \frac{\varepsilon}{2} \tag{3}$$

for all $m, n \geq N$.

Now if $\phi \in L^p(\mu)$ is unbounded, we want to show that $\frac{1}{n}S_n\phi$ is a Cauchy sequence in $\| \ \|_p$. Let $\varepsilon > 0$ be arbitrary, and take ψ bounded such that $\|\phi - \psi\|_p < \varepsilon/4$. Note that by T-invariance, $\|\frac{1}{n}S_n\phi - \frac{1}{n}S_n\psi\|_p \leq \|\phi - \psi\|_p$ for all $n \geq 1$. Therefore, using the triangle inequality and (3) above,

$$\begin{aligned}\|\frac{1}{m}S_m\phi - \frac{1}{n}S_n\phi\|_p &\leq \|\frac{1}{m}S_m\phi - \frac{1}{m}S_m\psi\|_p + \|\frac{1}{m}S_m\psi - \frac{1}{n}S_n\psi\|_p + \|\frac{1}{n}S_n\phi - \frac{1}{n}S_n\psi\|_p \\ &< \frac{\varepsilon}{4} + \frac{\varepsilon}{2} + \frac{\varepsilon}{4} = \varepsilon\end{aligned}$$

for all $m, n \geq N(\varepsilon, \psi)$. Hence $\frac{1}{n}S_n\phi$ is Cauchy in $\| \ \|_p$ and thus converges to $\phi^* = \lim_n \frac{1}{n}S_n\psi$. Since

$$\left|\frac{n+1}{n}\frac{1}{n+1}S_{n+1}\phi(x) - \frac{1}{n}S_n\phi \circ T(x)\right| = |\frac{1}{n}\phi(x)|$$

for all x, it follows by taking the limit $n \to \infty$ that $\phi^* = \phi^* \circ T$ μ-a.e. □

5 Absolute continuity and invariant densities

Definition 4. *A measure μ is called* **absolutely continuous** *w.r.t. the measure ν (notation: $\mu \ll \nu$) if $\nu(A) = 0$ implies $\mu(A) = 0$. If both $\mu \ll \nu$ and $\nu \ll \mu$, then μ and ν are called equivalent.*

Proposition 3. *Suppose that $\mu \ll \nu$ are both T-invariant probability measures, with a common σ-algebra $\mathcal{B}$ of measurable sets. If ν is ergodic, then $\mu = \nu$.*

Proof. First we show that μ is ergodic. Indeed, otherwise there is a T-invariant set A such that $\mu(A) > 0$ and $\mu(A^c) > 0$. By ergodicity of ν at least one of A or A^c must have ν-measure 0, but this would contradict that $\mu \ll \nu$.

Now let $A \in \mathcal{B}$ and let $Y \subset X$ be the set of ν-typical points. Then $\nu(Y^c) = 0$ and hence $\mu(Y^c) = 0$. Applying Birkhoff's Ergodic Theorem to μ and ν separately for $\psi = 1_A$ and some μ-typical $y \in Y$, we get

$$\mu(A) = \lim_n \frac{1}{n} \sum_{i=0}^{n-1} \psi \circ T(y) = \nu(A).$$

But $A \in \mathcal{B}$ was arbitrary, so $\mu = \nu$. $\square$

Theorem 8 (Radon-Nikodym). *If μ is a probability measure and $\mu \ll \nu$ then there is a function $h \in L^1(\nu)$ (called* **Radon-Nikodym derivative** *or* **density***) such that $\mu(A) = \int_A h(x)\ d\nu(x)$ for every measurable set A.*

Sometimes we use the notation: $h = \frac{d\mu}{d\nu}$.

Proposition 4. *Let $T : U \subset \mathbb{R}^n \to U$ be (piecewise) differentiable, and μ is absolutely continuous w.r.t. Lebesgue. Then μ is T-invariant if and only if its density $h = \frac{d\mu}{dx}$ satisfies*

$$h(x) = \sum_{T(y)=x} \frac{h(y)}{|\det DT(y)|} \tag{4}$$

for every x.

Proof. The T-invariance means that $d\mu(x) = d\mu(T^{-1}(x))$, but we need to be aware that T^{-1} is multivalued. So it is more careful to split the space U into pieces U_n such that the restrictions $T_n := T|U_n$ are diffeomorphic (onto their images) and write $y_n = T_n^{-1}(x) = T^{-1}(x) \cap U_n$. Then we obtain (using the change of coordinates)

$$\begin{aligned} h(x)\ dx &= d\mu(x) = d\mu(T^{-1}(x)) = \sum_n d\mu \circ T_n^{-1}(x) \\ &= \sum_n h(y_n) |\det(DT_n^{-1})(x)| dy_n = \sum_n \frac{h(y_n)}{\det |DT(y_n)|} dy_n, \end{aligned}$$

and the statement follows.

Conversely, if (4) holds, then the above computation gives $d\mu(x) = d\mu \circ T^{-1}(x)$, which is the required invariance. $\square$

Example: If $T : [0,1] \to [0,1]$ is (countably) piecewise linear, and each branch $T : I_n \to [0,1]$ (on which T is affine) is onto, then T preserves Lebesgue measure. Indeed, the intervals I_n have pairwise disjoint interiors, and their lengths add up to 1. If s_n is the slope of $T : I_n \to [0,1]$, then $s_n = 1/|I_n|$. Therefore $\sum_n \frac{1}{DT(y_n)} = \sum_n \frac{1}{s_n} = \sum_n |I_n| = 1$.

Example: The map $T : \mathbb{R} \setminus \{0\} \to \mathbb{R},\ T(x) = x - \frac{1}{x}$ is called the **Boole transformation**. It is 2-to-1; the two preimages of $x \in \mathbb{R}$ are $y_\pm = \frac{1}{2}(x \pm \sqrt{x^2+4})$. Clearly $T'(x) = 1 + \frac{1}{x^2}$. A tedious computation shows that

$$\frac{1}{|T'(y_-)|} + \frac{1}{|T'(y_+)|} = 1.$$

Indeed, $|T'(y_\pm)| = 1 + \frac{2}{x^2+2\pm x\sqrt{x^2+4}}$, $\ 1/|T'(y_\pm)| = \frac{x^2+2\pm x\sqrt{x^2+4}}{x^2+4\pm x\sqrt{x^2+4}}$, and

$$\begin{aligned}
\frac{1}{|T'(y_-)|} + \frac{1}{|T'(y_+)|} &= \frac{x^2+2-x\sqrt{x^2+4}}{x^2+4-x\sqrt{x^2+4}} + \frac{x^2+2+x\sqrt{x^2+4}}{x^2+4+x\sqrt{x^2+4}} \\
&= \frac{(x^2+2-x\sqrt{x^2+4})(x^2+4+x\sqrt{x^2+4})}{(x^2+4)^2 - x^2(x^2+4)} \\
&\qquad + \frac{(x^2+2+x\sqrt{x^2+4})(x^2+4-x\sqrt{x^2+4})}{(x^2+4)^2 - x^2(x^2+4)} \\
&= \frac{(x^2+2)^2 - x^2(x^2+4) + 2(x^2+2) - 2x\sqrt{x^2+4}}{4(x^2+4)} + \\
&\qquad \frac{(x^2+2)^2 - x^2(x^2+4) + 2(x^2+2) + 2x\sqrt{x^2+4}}{4(x^2+4)} \\
&= \frac{4(x^2+2)+8}{4(x^2+4)} = 1.
\end{aligned}$$

Therefore $h(x) \equiv 1$ is an invariant density, so Lebesgue measure is preserved.

Example: The **Gauß map** $G : (0,1] \to [0,1)$ is defined as $G(x) = \frac{1}{x} - \lfloor \frac{1}{x} \rfloor$. It has an invariant density $h(x) = \frac{1}{\log 2} \frac{1}{1+x}$. Here $\frac{1}{\log 2}$ is just the normalising factor (so that $\int_0^1 h(x)dx = 1$).

Let $I_n = (\frac{1}{n+1}, \frac{1}{n}]$ for $n = 1,2,3,\dots$ be the domains of the branches of G, and for $x \in (0,1)$, and $y_n := G^{-1}(x) \cap I_n = \frac{1}{x+n}$. Also $G'(y_n) = -\frac{1}{y_n^2}$. Therefore

$$\begin{aligned}
\sum_{n \geqslant 1} \frac{h(y_n)}{|G'(y_n)|} &= \frac{1}{\log 2} \sum_{n \geqslant 1} \frac{y_n^2}{1+y_n} = \frac{1}{\log 2} \sum_{n \geqslant 1} \frac{\frac{1}{(x+n)^2}}{1 + \frac{1}{x+n}} \\
&= \frac{1}{\log 2} \sum_{n \geqslant 1} \frac{1}{x+n} \frac{1}{x+n+1} \\
&= \frac{1}{\log 2} \sum_{n \geqslant 1} \frac{1}{x+n} - \frac{1}{x+n+1} \qquad \text{telescoping series} \\
&= \frac{1}{\log 2} \frac{1}{x+1} = h(x).
\end{aligned}$$

Exercise 1. *Show that for each integer $n \geqslant 2$, the interval map given by*

$$T_n(x) = \begin{cases} nx & \text{if } 0 \leqslant x \leqslant \frac{1}{n}, \\ \frac{1}{x} - \lfloor \frac{1}{x} \rfloor & \text{if } \frac{1}{n} < x \leqslant 1, \end{cases}$$

has invariant density $\frac{1}{\log 2} \frac{1}{1+x}$.

Theorem 9 (Folklore Theorem)**.** *If $T : \mathbb{S}^1 \to \mathbb{S}^1$ is a C^2 expanding circle map, then it preserves a measure μ equivalent to Lebesgue, and μ is ergodic.*

Expanding here means that there is $\lambda > 1$ such that $|T'(x)| \geqslant \lambda$ for all $x \in \mathbb{S}^1$. The above theorem can be proved in more generality, but in the stated version it conveys the ideas more clearly.

Proof. First some estimates on derivatives. Using the Mean Value Theorem twice, we obtain

$$\begin{aligned} \log \frac{|T'(x)|}{|T'(y)|} &= \log(1 + \frac{|T'(x)| - |T'(y)|}{|T'(y)|}) \leqslant \frac{|T'(x)| - |T'(y)|}{|T'(y)|} \\ &= \frac{|T''(\xi)| \cdot |x-y|}{|T'(y)|} = \frac{|T''(\xi)|}{|T'(y)|} \frac{|Tx - Ty|}{T'(\zeta)}. \end{aligned}$$

Since T is expanding, the denominators are $\geq \lambda$ and since T is C^2 on a compact space, also $|T''(\xi)|$ is bounded. Therefore there is some $K \leq \sup |T''(\xi)|/\lambda^2$ such that

$$\log \frac{|T'(x)|}{|T'(y)|} \leqslant K|T(x) - T(y)|.$$

The chain rule then gives:

$$\log \frac{|DT^n(x)|}{|DT^n(y)|} = \sum_{i=0}^{n-1} \log \frac{|T'(T^i x)|}{|T'(T^i y)|} \leqslant K \sum_{i=1}^{n} |T^i(x) - T^i(y)|.$$

Since T is a continuous expanding map of the circle, it wraps the circle d times around itself, and for each n, there are d^n pairwise disjoint intervals Z_n such that $T^n : Z_n \to \mathbb{S}^1$ is onto, with slope at least λ^n. If we take x, y above in one such Z_n, then $|x - y| \leqslant \lambda^{-n}|T^n(x) - T^n(y)|$ and in fact $|T^i(x) - T^i(y)| \leqslant \lambda^{-(n-i)}|T^n(x) - T^n(y)|$. Therefore we obtain

$$\log \frac{|DT^n(x)|}{|DT^n(y)|} = K \sum_{i=1}^{n} \lambda^{-(n-i)} |T^n(x) - T^n(y)| \leqslant \frac{K}{\lambda - 1} |T^n(x) - T^n(y)| \leqslant \log K'$$

for some $K' \in (1, \infty)$. This means that if $A \subset Z_n$ (so $T^n : A \to T^n(A)$ is a bijection), then

$$\frac{1}{K'} \frac{m(A)}{m(Z_n)} \leqslant \frac{m(T^n A)}{m(T^n Z_n)} = \frac{m(T^n A)}{m(\mathbb{S}^1)} \leqslant K' \frac{m(A)}{m(Z_n)}, \tag{5}$$

where m is Lebesgue measure.

Now we construct the T-invariant measure μ. Take $B \subset \mathcal{B}$ arbitrary, and set $\mu_n(B) = \frac{1}{n}\sum_{i=0}^{n-1} m(T^{-i}B)$. Then by (5),

$$\frac{1}{K'}m(B) \leqslant \mu_n(B) \leqslant K'm(B).$$

We can take a weak* limit of the μ_n's; call it μ. Then

$$\frac{1}{K'}m(B) \leqslant \mu(B) \leqslant K'm(B),$$

and therefore μ and m are equivalent. The T-invariance of μ proven in the same way as in Theorem 1.

Now for the ergodicity of μ, we need the Lebesgue Density Theorem, which says that if $m(A) > 0$, then for m-a.e. $x \in A$, the limit

$$\lim_{\varepsilon\to 0} \frac{m(A \cap B_\varepsilon(x))}{m(B_\varepsilon(x))} = 1,$$

where $B_\varepsilon(x)$ is the ε-balls around x. Points x with this property are called **(Lebesgue) density points** of A. (In fact, the above also holds, if $B_\varepsilon(x)$ is just a one-sided ε-neighbourhood of x.)

Assume by contradiction that μ is not ergodic. Take $A \in \mathcal{B}$ a T-invariant set such that $\mu(A) > 0$ and $\mu(A^c) > 0$. By equivalence of μ and m, also $\delta := m(A^c) > 0$. Let x be a density point of A, and Z_{nn} be a neighbourhood of x such that $T^n : Z_n \to \mathbb{S}^1$ is a bijection. As $n \to \infty$, $Z_n \to \{x\}$, and therefore we can choose n so large (hence Z_n so small) that

$$\frac{m(A \cap Z_n)}{m(Z_n)} > 1 - \delta/K'.$$

Therefore $\frac{m(A^c \cap Z_n)}{m(Z_n)} < \delta/K'$, and using (5),

$$\frac{m(T^n(A^c \cap Z_n))}{m(T^n(Z_n))} \leqslant K'\frac{m(A^c \cap Z_n)}{m(Z_n)} < K'\delta/K' = \delta.$$

Since $T^n : A^c \cap Z_n \to A^c$ is a bijection, and $m(T^n Z_n) = m(\mathbb{S}^1) = 1$, we get $\delta = m(A^c) < \delta$, a contraction. Therefore μ is ergodic. □

6 The Choquet Simplex and the Ergodic Decomposition

Throughout this section, let $T : X \to X$ a **continuous** transformation of a compact metric space. Recall that $\mathcal{M}(X)$ is the collection of probability measures defined on

X; we saw in (1) that it is compact in the weak* topology. In general, X carries many T-invariant measures. The set $\mathcal{M}(X,T) = \{\mu \in \mathcal{M}(X) : \mu \text{ is } T\text{-invariant}\}$ is called the **Choquet simplex** of T. Let $\mathcal{M}_{erg}(X,T)$ be the subset of $\mathcal{M}(X,T)$ of ergodic T-invariant measures.

Clearly $\mathcal{M}(X,T) = \{\mu\}$ if (X,T) is uniquely ergodic. The name "simplex" just reflects the convexity of $\mathcal{M}(X,T)$: if $\mu_1, \mu_2 \in \mathcal{M}(X,T)$, then also $\alpha\mu_1 + (1-\alpha)\mu_2 \in \mathcal{M}(X,T)$ for every $\alpha \in [0,1]$.

Lemma 3. *The Choquet simplex $\mathcal{M}(X,T)$ is a compact subset of $\mathcal{M}(X)$ w.r.t. weak* topology.*

Proof. Suppose $\{\mu_n\} \subset \mathcal{M}(X,T)$, then by the compactness of $\mathcal{M}(X)$, see (1), there is $\mu \in \mathcal{M}(X)$ and a subsequence $(n_i)_i$ such that for every continuous function $f : X \to \mathbb{R}$ such that $\int f d\mu_{n_i} \to \int f d\mu$ as $i \to \infty$. It remains to show that μ is T-invariant, but this simply follows from continuity of $f \circ T$ and

$$\int f \circ T \ d\mu = \lim_i \int f \circ T \ d\mu_{n_i} = \lim_i \int f \ d\mu_{n_i} = \int f \ d\mu.$$

□

Theorem 10. *The ergodic measures are exactly the extremal points of the Choquet simplex.*

Proof. First assume that μ is not ergodic. Hence there is a T-invariant set A such that $0 < \mu(A) < 1$. Define

$$\mu_1(B) = \frac{\mu(B \cap A)}{\mu(A)} \quad \text{and} \quad \mu_2(B) = \frac{\mu(B \setminus A)}{\mu(X \setminus A)}.$$

Then $\mu = \alpha\mu_1 + (1-\alpha)\mu_2$ for $\alpha = \mu(A) \in (0,1)$ so μ is not an extremal point.

Conversely, suppose that μ is ergodic, and $\mu = \alpha\mu_1 + (1-\alpha)\mu_2$ for some $\alpha \in (0,1)$ and $\mu_1, \mu_2 \in \mathcal{M}(X,T)$. Then $\mu_1 \ll \mu$ and also $\mu_2 \ll \mu$. Proposition 3 implies that $\mu_1 = \mu_2 = \mu$, so the convex combination is trivial and μ must be extremal. □

The following fundamental theorem implies that for checking the properties of any measure $\mu \in \mathcal{M}(X,T)$, it suffices to verify the properties for ergodic measures:

Theorem 11 (Ergodic Decomposition)**.** *For every $\mu \in \mathcal{M}(X,T)$, there is a measure ν on the spaces of ergodic measures such that $\nu(\mathcal{M}_{erg}(X,T)) = 1$ and*

$$\mu(B) = \int_{\mathcal{M}_{erg}(X,T)} m(B) \ d\nu(m)$$

for all Borel sets B.

Definition 5. *The simplex $\mathcal{M}(X,T)$ of T-invariant probability measures is called a* **Poulsen simplex** *if it is not degenerate, but the extremal points (i.e. $\mathcal{M}_{erg}(X,T)$ lie dense in $\mathcal{M}(X,T)$.*

This definition shows what a enormous and complicated space the Choquet simplex can be. And it is a reality for many dynamical systems, as we will demonstrate, as an example, for the doubling map.

Proposition 5. *The Choquet simplex of the doubling map $T : S^1 \to \mathbb{S}^1$, $x \mapsto 2x \bmod 1$, is a Poulsen smplex.*

Proof. First note that an equidistribution on periodic orbits is an ergodic measure. Therefore it suffices to show that the equidistributions lie dense in the Choquet simplex $\mathcal{M}(\mathbb{S}^1, T)$.

We claim that for every $\delta > 0$, there is $N \in \mathbb{N}$ such that for every $n \in \mathbb{N}$ and $x \in \mathbb{S}^1$ there is $y \in \mathbb{S}^1$ such that

- $|T^k(x) - T^k(y)| < \delta$ for all $0 \leq k < n$;
- $T^{n+N}(y) = y$.

To prove the claim. Take N so large that $2^{-N} < \delta$ and $n \in \mathbb{N}$, $x \in \mathbb{S}^1$ arbitrary. Take y in the same dyadic interval J of generation $n+N$ as x, so that $T^{n+N}(y) = y$. Since T^{n+N} maps J onto S^1 this is possible. Also $|T^k(x) - T^k(y)| \leq |T^k(J)| \leq |T^n(J)| = 2^{-N} < \delta$ for all $0 \leq k < n$.

With this claim, we argue as follows. Let $\mu \in \mathcal{M}$ and $m \in \mathbb{N}$ be arbitrary. Take x a typical point for μ, i.e., for every $f \in C(X)$, say with $\|f\|_\infty \leq m$, $\lim_n \frac{1}{n} \sum_{k=0}^{n-1} f \circ T^k(x) = \int_X f \, d\mu$. We can find a finite collection $f_j \subset C(X)$ such that for every $f \in C(X)$, $\|f\|_\infty$, there is f_j such that $| \int_X f_j \, d\mu - \int_X f \, d\mu| < 1/m$.

Since f_j is continuous, it is uniformly continuous on the compact space X. Hence we can find $\delta > 0$ such that $|x - y| < \delta$ implies $|f_j(x) - f_j(y)| < 1/m$ for all j. For this $\delta > 0$, take $N \in N$ as in the claim.

Take $n > Nm$ so large that for each f_j, $|\frac{1}{n} \sum_{k=0}^{n-1} f \circ T^k(x) - \int_X f \, d\mu| < \frac{1}{m}$.

Now find the $n + N$-periodic point y close to x as in the claim. Let $\nu = \nu_m$ be the equidistribution on the orbit of y. Then we can compute

$$\begin{aligned}
|\int_X f\,d\mu - \int_X f\,d\nu| &\le |\frac{1}{n}\sum_{k=0}^{n-1} f\circ T^k(x) - \frac{1}{n+N}\sum_{k=0}^{n+N-1} f\circ T^k(y)| + \frac{1}{m} \\
&\le |\frac{1}{n}\sum_{k=0}^{n-1} f_j\circ T^k(x) - \frac{1}{n+N}\sum_{k=0}^{n+N-1} f_j\circ T^k(y)| + \frac{2}{m} \\
&\le \frac{1}{n}|\sum_{k=0}^{n-1} f_j\circ T^k(x) - f_j\circ T^k(y)| + \frac{N}{n+N}\|f_j\|_\infty + \frac{3}{m} \\
&\le \frac{1}{n}\sum_{k=0}^{n-1}\frac{1}{m} + \frac{3}{m} = \frac{4}{m}.
\end{aligned}$$

In this we can produce such equidistribution ν_m for each $m \in \mathbb{N}$, and the sequence (ν_m) converges to μ in the weak* topology. □

7 Poincaré Recurrence

Theorem 12 (Poincaré's Recurrence Theorem)**.** *If (X, T, μ) is a measure preserving system with $\mu(X) = 1$, then for every measurable set $U \subset X$ of positive measure, μ-a.e. $x \in U$ returns to U,* i.e., *there is $n = n(x)$ such that $T^n(x) \in U$.*

Proof of Theorem 12. Let U be an arbitrary measurable set of positive measure. As μ is invariant, $\mu(T^{-i}(U)) = \mu(U) > 0$ for all $i \geqslant 0$. On the other hand, $1 = \mu(X) \geqslant \mu(\cup_i T^{-i}(U))$, so there must be overlap in the backward iterates of U, *i.e.,* there are $0 \leqslant i < j$ such that $\mu(T^{-i}(U) \cap T^{-j}(U)) > 0$. Take the j-th iterate and find $\mu(T^{j-i}(U) \cap U) \geqslant \mu(T^{-i}(U) \cap T^{-j}(U)) > 0$. This means that a positive measure part of the set U returns to itself after $n := j - i$ iterates.

For the part U' of U that didn't return after n steps, assuming U' has positive measure, we repeat the argument. That is, there is n' such that $\mu(T^{n'}(U') \cap U') > 0$ and then also $\mu(T^{n'}(U') \cap U) > 0$.

Repeating this argument, we can exhaust the set U up to a set of measure zero, and this proves the theorem. □

Definition 6. *(i) A system $(X, T, \mathcal{B}, \mu)$ is called* conservative *if for every set $A \in \mathcal{B}$ with $\mu(A) > 0$, there is $n \geqslant 1$ such that $\mu(T^n(A) \cap A) > 0$. The Poincaré Recurrence Theorem thus states that probability measure preserving systems are conservative.*
(ii) The system is called dissipative *otherwise, and it is called* totally dissipative *if for every set $A \in \mathcal{B}$, $\mu(T^n(A) \cap A) = 0$ for every $n \ge 1$.*
(iii) We call the transformation T **recurrent** *w.r.t. μ if $B \setminus \cup_{i\in\mathbb{N}} T^{-i}(B)$ has zero measure for every $B \in \mathcal{B}$. In fact, this is equivalent to μ being conservative.*

Define the first return time to a set Y as

$$\tau_Y = \min\{n \geq 1 : T^n(x) \in Y\}.$$

The next result quantifies the expected value of the first return time.

Lemma 4 (Kac Lemma)**.** *Let (X,T) preserve an ergodic measure μ. Take $Y \subset X$ measurable such that $0 < \mu(Y) \leqslant 1$, and let $\tau = \tau_Y : Y \to \mathbb{N}$ be the first return time to Y. Then*

$$\int \tau d\mu = \sum_{k\geqslant 1} k\mu(Y_k) = \mu(X)$$

for $Y_k := \{y \in Y : \tau(y) = k\}$.

Proof. First set $A = \{y \in Y : T^j(y) \notin Y \text{ for all } j \geq 1\}$. Then $T^j(A) \cap A = \emptyset$ for all $j \geq 1$, so by the Poincaré Recurrence Theorem if $\mu(X) = 1$, or by the assumed conservativity if μ is infinite, $\mu(A) = 0$. But then also $\bigcup_{j\geq 0} T^{-j}(A) = \{x \in X : T^j(x) \in Y \text{ finitely often}\}$ has zero measure. This shows that μ-a.e. $x \in X$ enters Y infinitely often.

Next define $L_0 = Y$, $L_1 = T^{-1}(Y) \setminus Y$ and recursively $L_{j+1} = T^{-j}(L_j) \setminus Y$. In other words:

$$L_j = \{x \in X : T^j(x) \in Y \text{ and } T^k(x) \notin Y \text{ for } 0 \leq k < j\}.$$

Clearly all the L_js are pairwise disjoint, and by the previous paragraph, $\sum_{j\geq 0} \mu(L_j) = \mu(X)$.

Furthermore, $T^{-1}(L_j)$ is the disjoint union of L_{j+1} and Y_{j+1} where we recall that $Y_{j+1} = \{y \in Y : \tau(y) = j+1\}$. By T-invariance of μ it follows that $\mu(L_j) = \mu(L_{j+1}) + \mu(Y_{j+1})$. Therefore

$$\begin{aligned}
\sum_{k=1}^{\infty} k\mu(Y_k) = \sum_{k=0}^{\infty} (k+1)\mu(Y_{k+1}) &= \sum_{k=0}^{\infty} (k+1)(\mu(L_k) - \mu(L_{k+1})) \\
&= \sum_{k=0}^{\infty} \mu(L_k) + \underbrace{k\mu(L_k) - (k+1)\mu(L_{k+1})}_{\text{telescopes to } 0} \\
&= \sum_{k=0}^{\infty} \mu(L_k) = \mu(X).
\end{aligned}$$

This proves Kac' Lemma. □

7.1 Induced transformations

Kac's Lemma effectively combines a measure preserving system (X, f) to the first return mapped to a subset $Y \subset X$.

Proposition 6. *Let $(X, \mathcal{B}, T, \mu)$ be a non-singular dynamical system and $Y \in \mathcal{B}$ a set with $\mu(Y) > 0$. Let $F = T^\tau$ be the first return map to Y.*

If μ is T-invariant, then $\nu(A) := \frac{1}{m(Y)}\mu(A \cap Y)$ is F-invariant. Conversely, if ν is F-invariant, and

$$\Lambda = \int_Y \tau(x) d\nu < \infty, \tag{6}$$

then

$$\mu(A) = \frac{1}{\Lambda} \sum_{j=1}^{\infty} \nu(T^{-j}(A) \cap \{y \in Y : \tau(y) \geq j\}) \tag{7}$$

is a T-invariant probability measure. Moreover μ is ergodic for T if and only if ν is ergodic for F.

Proof. Let $A \subset Y$ be measurable. We can write $T^{-1}(A)$ as disjoint union $F^{-1}(A) = \sqcup_{j \geq 1} Y_j \cap T^{-j}(A)$, where $Y_j = \{y \in Y : \tau(y) = j\}$. Using the notation of the previous proof, we compute

$$\begin{aligned} \mu(A) = \mu(L_0 \cap A) &= \mu(L_1 \cap T^{-1}(A)) + \mu(Y_1 \cap T^{-1}(A)) \\ &= \mu(L_2 \cap T^{-2}(A)) + \mu(Y_2 \cap T^{-2}(A)) + \mu(Y_1 \cap T^{-1}(A)) \\ &= \quad \vdots \qquad\qquad \vdots \\ &= \sum_{j \geq 1} \mu(Y_j \cap T^{-j}(A)) = \mu(F^{-1}(A)). \end{aligned}$$

After scaling by $1/\mu(Y)$, we get $\nu(A) = \nu(F^{-1}(A))$.

Conversely, note that $\mu(X) = \frac{1}{\Lambda} \sum_{j=1}^{\infty} \nu(\{y \in Y : \tau(y) \geq 1\}) = \frac{1}{\Lambda} \sum_{j=1}^{\infty} j\nu(\{y \in Y : \tau(y) = j\}) = \frac{1}{\Lambda} \int_Y \tau \, d\nu = 1$. For the invariance, we compute

$$\begin{aligned} \mu(T^{-1}(A)) &= \frac{1}{\Lambda} \sum_{j=1}^{\infty} \nu(T^{-(j+1)}(A) \cap \{\tau(y) \geq j\}) \\ &= \frac{1}{\Lambda} \sum_{j=1}^{\infty} \nu(T^{-(j+1)}(A) \cap \{\tau(y) \geq j+1\}) + \nu(T^{-(j+1)}(A) \cap \{\tau(y) = j\}) \\ &= \frac{1}{\Lambda} \sum_{j=1}^{\infty} \big(\nu(T^{-j}(A) \cap \{\tau(y) \geq j\}) + \nu(T^{-j}(T^{-1}(A)) \cap \{\tau(y) = j\})\big) \\ &\qquad - \frac{1}{\Lambda} \nu(T^{-1}(A) \cap \{\tau(y) \geq 1\}) \\ &= \mu(A) + \frac{1}{\Lambda} \big(\nu(F^{-1}(T^{-1}(A)) - \nu(T^{-1}(A))\big) = \mu(A), \end{aligned}$$

where the last equality is by F-invariance of ν.

Now for ergodicity, first assume that μ is ergodic and $A \subset Y$ is F-invariant. Then $A' = \cup_{j=0}^{\infty} T^{-j}(A)$ is T-invariant, so $\mu(A') \in \{0,1\}$. If $\mu(A') = 0$ then $\mu(A) = \nu(A) = 0$ and if $\mu(A') = 1$, then $\mu(A) = \mu(Y)$ and hence $\nu(A) = 1$. Finally, if ν is ergodic, and A' is T-invariant, then $A := A' \cap Y$ is F-invariant, and therefore $\nu(A) \in \{0,1\}$. Since T is non-singular, it follows from (7) that $\mu(A') \in \{0,1\}$. $\square$

As an illustration, we take the quadratic map $f(x) = 4x(1-x)$. It is not uniformly expanding, so we cannot apply the Folklore Theorem 9 to find an absolutely continuous probability measure μ. Therefore we take $Y = [1-p, p]$ for the fixed point $p = \frac{3}{4}$ of f. and consider the first return map $F : Y \to Y$. Note that the critical point $c = \frac{1}{2}$ (*i.e.,* the point where the derivative is zero) never returns to Y. Indeed, $f(c) = 1$ and $f^2(1) = 0$ is fixed under f. This is essential for F to have a chance to be uniformly expanding.

Without proofs, we mention the properties of F:

- F is defined for Lebesgue-a.e. $y \in Y$.
- If $y \in Y$ has return time $\tau(y) = n$, then there is a neighborhood U_x of x such that $F : U_x \to Y^\circ$ is a C^∞ diffeomorphism and $|F'| \geq 2$.
- F has infinitely many branches (so it is not piecewise C^2 in the strict sense), and F' is not bounded. However, there is a constant C such that

$$\frac{|F''(y)|}{|F'(y)|^2} \leq C \qquad \text{wherever defined.}$$

- The Lebesgue measure of $\{y \in Y : \tau(y) = n\}$ is exponentially small in n.

These conditions are sufficient to get the conclusion of the Folklore Theorem 9, so we have an F-invariant measure ν and in fact, its density $\frac{d\nu}{dx}$ is bounded and bounded away from zero. This means that $\nu(\{y \in Y : \tau(y) = n\})$ is exponentially small in n as well, so that the normalizing constant Λ from (6) is finite. Hence, we conclude that f preserves an ergodic absolutely continuous measure μ, satisfying the formula (7).

For the above example, it is not essential that f is a quadratic map; any C^2 unimodal map $f : [0,1] \to [0,1]$ with $f^2(c) = 0$ fixed and $f''(c) \neq 0$ can be treated in the same way. For the quadratic map, however, the density of μ is known precisely:

$$\frac{d\mu}{dx} = \frac{1}{\pi\sqrt{x(1-x)}}.$$

8 The Koopman operator

Given a probability measure preserving dynamical system $(X, \mathcal{B}, \mu, T)$, we can take the space of complex-valued square-integrable observables $L^2(\mu)$. This is a Hilbert space, equipped with inner product $\langle f, g\rangle = \int_X f(x) \cdot \overline{g(x)}\ d\mu$.

The Koopman operator $U_T : L^2(\mu) \to L^2(\mu)$ is defined as $U_T f = f \circ T$. By T-invariance of μ, it is a unitary operator. Indeed

$$\langle U_T f, U_T g\rangle = \int_X f \circ T(x) \cdot \overline{g \circ T(x)}\ d\mu = \int_X (f \cdot \overline{g}) \circ T(x)\ d\mu = \int_X f \cdot \overline{g}\ d\mu = \langle f, g\rangle,$$

and therefore $U_T^* U_T = U_T U_T^* = I$. This has several consequences, common to all unitary operators. First of all, the spectrum $\sigma(U_T)$ of U_T is a closed subset of the unit circle.

Secondly, we can give a (continuous) decomposition of U_T in orthogonal projections, called the **spectral decomposition**. For a fixed eigenfunction ψ (with eigenvalue $\lambda \in \mathbb{S}^1$, we let $\Pi_\lambda : L^2(\mu) \to L^2(\mu)$ be the orthogonal projection onto the span of ψ. More generally, if $S \subset \sigma(U_T)$, we define Π_S as the orthogonal projection on the largest closed subspace V such that $U_T|_V$ has spectrum contained in S. As any orthogonal projection, we have the properties:

- $\Pi_S^2 = \Pi_S$ (Π_S is idempotent);
- $\Pi_S^* = \Pi_S$ (Π_S is self-adjoint);
- $\Pi_S \Pi_{S'} = 0$ if $S \cap S' = \varnothing$;
- The kernel $\mathcal{N}(\Pi_S)$ equals the orthogonal complement, $V^\perp$, of V.

Theorem 13 (Spectral Decomposition of Unitary Operators)**.** *There is a measure ν_T on $\mathbb{S}^1$ such that*

$$U_T = \int_{\sigma(U_T)} \lambda\, \Pi_\lambda d\nu_T(\lambda),$$

and $\nu_T(\lambda) \neq 0$ if and only if λ is an eigenvalue of U_T. Using the above properties of orthogonal projections, we also get

$$U_T^n = \int_{\sigma(U_T)} \lambda^n \Pi_\lambda d\nu_T(\lambda).$$

9 The Perron-Frobenius operator

Definition 7. *The* **Perron-Frobenius operator** *of a transformation $T : X \to X$ is the dual of the Koopman operator:*

$$\int_X P_T f \cdot g\ d\mu = \int_X f \cdot U_T g\ d\mu = \int_X f \cdot g \circ T\ d\mu. \tag{8}$$

Note that, although U_T is independent of the measure, P_T is not. Often it will be important to specify the measure explicitly, and this measure need not be invariant.

The following basic properties are straighforward to check.

Proposition 7. *The Perron-Frobenius operator has the following properties:*

1. *P_T is linear;*
2. *P_T is positive: $f \geq 0$ implies $P_T f \geq 0$.*
3. *$\int P_T f \; d\mu = \int f \; d\mu$.*
4. *$P_{T^k} = (P_T)^k$.*

Lemma 5. *Let $T : [0,1] \to [0,1]$ be a piecewise C^1 interval map. Then the Perron-Frobenius operator P_T w.r.t. Lebesgue measure λ has the pointwise formula*

$$P_T f(x) = \sum_{y \in T^{-1}} \frac{f(y)}{|T'(y)|}. \tag{9}$$

Proof. Let $0 = a_0 < a_1 < \cdots < a_N = 1$ be such that T is C^1 monotone on each (a_{i-1}, a_i). Let $y_i = T^{-1}(x) \cap (a_{i-1}, a_i)$. We obtain

$$\begin{aligned}
(P_T f)(x) &= \frac{d}{dx}\int_0^x P_T f \; ds = \frac{d}{dx}\int_0^1 (P_T f) \cdot 1_{[0,x]} ds \\
&= \frac{d}{dx}\int_0^1 f \cdot 1_{[0,x]} \circ T \; ds = \frac{d}{dx}\int_{T^{-1}[0,x]} f \; ds \\
&= \sum_{\substack{T|_{(a_{i-1},a_i)} \text{ increasing} \\ x \in T((a_{i-1},a_i))}} \frac{d}{dx}\int_{a_{i-1}}^{y_i} f \; ds + \sum_{\substack{T|_{(a_{i-1},a_i)} \text{ decreasing} \\ x \in T((a_{i-1},a_i))}} \frac{d}{dx}\int_{y_i}^{a_i} f \; ds \\
&\quad + \sum_{T((a_{i-1},a_i)) \subset [0,x]} \frac{d}{dx}\int_{a_{i-1}}^{a_i} f \; ds \\
&= \sum_{\substack{T|_{(a_{i-1},a_i)} \text{ increasing} \\ x \in T((a_{i-1},a_i))}} \frac{f(y_i)}{T'(y_i)} + \sum_{\substack{T|_{(a_{i-1},a_i)} \text{ decreasing} \\ x \in T((a_{i-1},a_i))}} -\frac{f(y_i)}{T'(y_i)} + 0 \\
&= \sum_i \frac{f(y_i)}{|T'(y_i)|},
\end{aligned}$$

as required. □

There is also a Perron-Frobenius operator with respect to $\mu \ll \lambda$ instead of Lebesgue measure:

Lemma 6. *If $d\mu = h d\lambda$, then the operator*

$$P_{T,\mu} f = \frac{P_T(f \cdot h)}{h} \tag{10}$$

acts as the Perron-Frobenius operator on $(X, \mathcal{B}, T, \mu)$.

Viewed differently, if $h \geq 0$ is a fixed function of P_T (w.r.t. Lebesgue) then $d\mu = h\ d\lambda$ is an invariant measure. Conversely, if $P_{T,\mu}$ is the Perron-Frobenius w.r.t. an invariant measure, then the constant function 1 ois a fixed point of $P_{T,\mu}$.

Proof. Let A be any μ-measurable set and $f \in L^1([0,1], \mu)$. Then

$$\int_A P_{T,\mu} f d\mu = \int_A \frac{P_T(f \cdot h)}{h} h d\lambda = \int_A P_T(f \cdot h) d\lambda = \int_{T^{-1}A} f \cdot h d\lambda = \int_{T^{-1}A} f\ d\mu.$$

Because A is arbitrary, this proves the lemma. $\square$

Example 1. *Let $T : [0,1] \to [0,1]$ be given by $T(x) = \frac{1}{2}x$. Then $P_T 1 = 2 \cdot 1_{[0,\frac{1}{2}]}$, $P_T^2 1 = 4 \cdot 1_{[0,\frac{1}{4}]}$ and in general $P_T^n 1 = 2^n \cdot 1_{[0,2^{-n}]}$. Therefore $P_T^n 1$ tends to 0 on $(0,1]$ pointwise, which leads to no probability density. (In the sense of distributions, the limit is the Dirac measure δ_0.) We see here that iterating P_T is unstable if T is contracting (hence expanding in backward direction). Conversely, expanding maps have a stabilizing effect on the Perron-Frobenius operator. Let $T : [0,1] \to [0,1]$ be given by $T(x) = 2x$ (mod 1). Then by (9), $P_T f(x) = \frac{1}{2}\left(f(\frac{x}{2}) + f(\frac{1+x}{2})\right)$, and as we iterate further $P_T^k f$ converges uniformly to a constant function.*

10 The Lasota-Yorke inquality for BV

Definition 8. *Let $g : [a,b] \to \mathbb{R}$. The* variation *of g is defined to be*

$$Var_{[a,b]}\, g = \sup \sum_{i=1}^{n} |g(x_i) - g(x_{i-1})|, \tag{11}$$

where the supremum runs over all finite partitions generated by points $a = x_0 < x_1 < \cdots < x_n = b$. Note that Var is a seminorm ($Var\ f = Var(f + C)$ for every constant C).

The variation measures the oscillation of a function. Obviously Var is homogeneous and subadditive in the sense that

$$\mathrm{Var}_{[a,b]}\, t \cdot g = |t| \mathrm{Var}_{[a,b]} g \text{ for every } t \in \mathbb{R}. \tag{12}$$

$$\mathrm{Var}_{[a,b]}\,(g_1 + g_2) \leq \mathrm{Var}_{[a,b]} g_1 + \mathrm{Var}_{[a,b]} g_2. \tag{13}$$

Furthermore

$$\sup g - \inf g \leq \mathrm{Var}_{[a,b]}\, g, \tag{14}$$

and equality is assumed for monotone functions.

Definition 9. *The space*

$$BV([a,b]) = \{g : [a,b] \to \mathbb{R};\ Var_{[a,b]} g < \infty\}$$

equipped with the norm $\|g\|_{BV} = Var_{[a,b]} g + \int_a^b |g|\, dx$ *is called the space of functions of* bounded variation.

It is not hard to show that functions of bounded variation are integrable; in fact they are even Riemann integrable. On the other hand, any C^1 function on $[a,b]$ has bounded variation. If g_2 is monotone, then $\mathrm{Var}_{[a,b]} g_1 \circ g_2 \leq \mathrm{Var}_{[\inf g_2, \sup g_2]} g_1$. In a sense $\mathrm{BV}([a,b])$ is also closed under taking products. Suppose $g_1 \in \mathrm{BV}([a,b])$ and $g_2 \in C^1([a,b])$. Then

$$\mathrm{Var}_{[a,b]}\, g_1 g_2 \leq \sup |g_2| \mathrm{Var}_{[a,b]} g_1 + \int_a^b |g_1(s) g_2'(s)| ds. \tag{15}$$

Proof of (15)*:* Use the equality

$$\sum_{i=1}^n |a_i b_i - a_{i-1} b_{i-1}| = \sum_{i=1}^n |b_i(a_i - a_{i-1}) + a_{i-1}(b_i - b_{i-1})|$$

and the Mean Value Theorem to obtain

$$\begin{aligned}
\mathrm{Var}_{[a,b]} g_1 g_2 &= \sup \sum_{i=1}^n |g_1(x_i) g_2(x_i) - g_1(x_{i-1}) g_2(x_{i-1})| \\
&= \sup \sum_{i=1}^n \{|g_2(x_i)||g_1(x_i) - g_1(x_{i-1})| + |g_1(x_{i-1})||g_2(x_i) - g_2(x_{i-1})|\} \\
&\leq \sup |g_2| \mathrm{Var}_{[a,b]} g_1 + \sup \sum_{i=1}^n |g_1(x_{i-1}) g_2'(\xi_i)||x_i - x_{i-1}| \\
&\leq \sup |g_2| \mathrm{Var}_{[a,b]} g_1 + \int_a^b |g_1(s) g_2'(s)| ds,
\end{aligned}$$

because the second term is just the Riemann sum approximating the integral. □

In particular, taking $g_1 \equiv 1$, we obtain for C^1 functions

$$\mathrm{Var}_{[a,b]}\, g \leq \int_a^b |g'(s)| ds. \tag{16}$$

Let us now proceed to Lasota and Yorke's result:

Theorem 14 (Lasota-Yorke)**.** *Suppose that $T : [0,1] \to [0,1]$ is piecewise C^2 and piecewise expanding. Then T has an absolutely continuous invariant probability measure whose density has bounded variation.*

Proof. The main technical step is to establish the **Lasota-Yorke[3] inequality**: there are $\rho \in (0,1)$ and $L > 0$ such that

$$\mathrm{Var}_{[0,1]} P_T g \leq \rho \mathrm{Var}_{[0,1]} g + L\|g\|_{L^1}, \tag{17}$$

for all $f \in$ BV. By iteration, it follows that the sequence $\mathrm{Var}_{[0,1]} P_T^n 1 \leq \frac{L}{1-\rho}$ for every n, and because $\int_0^1 P_T^n 1 = 1$, the densities $P_T^n 1$ are also bounded. This is also true for the Césaro means $\{\frac{1}{N}\sum_{n=0}^{N-1} P_T^n\}$, and by Helly's Theorem (see e.g. [5, Theorem 2.3.9]) there must be a weak* accumulation point which is an invariant density with bounded variation.

To prove (17) we need another formula on variations. If $0 \leq a < b \leq 1$ and $g \in \mathrm{BV}([0,1])$, then

$$\begin{aligned} \mathrm{Var}_{[0,1]} g\ 1_{[a,b]} &\leq \mathrm{Var}_{[a,b]} g + |g(a)| + |g(b)| \\ &\leq \mathrm{Var}_{[a,b]} g + |g(a) - g(c)| + |g(b) - g(c)| + 2|g(c)| \\ &\leq 2\mathrm{Var}_{[a,b]} g + 2|g(c)|, \end{aligned}$$

for any $c \in [a,b]$. We can choose c such that $|g(c)| \leq \frac{1}{b-a}\int_a^b |g(s)|ds$, and therefore

$$\mathrm{Var}_{[0,1]} g\ 1_{[a,b]} \leq 2\mathrm{Var}_{[a,b]} g + \frac{2}{b-a}\int_a^b |g(s)|ds. \tag{18}$$

Let $\rho := \sup 2/|T'|$. By assumption $\rho \in (0,2)$, but by taking an iterate of T we can assume that $\tilde{\rho} < 1$. Take $0 = a_0 < a_1 < \cdots < a_N = 1$ such that T is C^2 expanding on each $[a_{i-1}, a_i]$. In particular, it follows that

$$\frac{|T''(x)|}{|T'(x)|} \leq K \text{ and } \left|\frac{d}{dx}\frac{1}{T'(x)}\right| \leq \frac{K}{|T'(x)|}, \tag{19}$$

for some constant K. (In the points a_i this holds for the one-sided derivatives.)

Let $g \in \mathrm{BV}([0,1])$ be a probability density. We calculate

$$(P_T g)(x) = \sum_{z \in T^{-1}x} \frac{g(z)}{|T'(z)|} = \sum_{i=1}^N \frac{g(T^{-1}(x) \cap [a_{i-1}, a_i])}{|T'(T^{-1}(x) \cap [a_{i-1}, a_i])|},$$

where $T^{-1}(x) \cap [a_{i-1}, a_i]$ indicates the appropriate branch of the inverse T^{-1}. To compute the variation we have to take sums of densities $\frac{g}{|T'|}$ defined on disjoint intervals $[a_{i-1}, a_i]$.

[3]The attribution is correct, but was preceded by Doeblin & Fortet and Ionescu-Tulcea & Marinescu, who used it for more general spaces than BV.

To extend these densities to $[0,1]$, we need to add indicator functions. Then, using inequality (18),

$$\begin{aligned}
\mathrm{Var}_{[0,1]} P_T g &= \sum_{i=1}^{N} \mathrm{Var}_{[0,1]} \frac{g}{|T'|}\ 1_{[a_{i-1},a_i]} \\
&\leq 2\sum_{i=1}^{N} \mathrm{Var}_{[a_{i-1},a_i]} \frac{g}{|T'|} + 2\sum_{i=1}^{N} \frac{1}{a_i - a_{i-1}} \int_{a_{i-1}}^{a_i} \frac{|g(s|)}{|T'(s)|} ds
\end{aligned}$$

Using the product formulas (15) and (19) we obtain

$$\begin{aligned}
2\sum_{i=1}^{N} \mathrm{Var}_{[a_{i-1},a_i]} \frac{g}{|T'|} &\leq \rho \sum_{i=1}^{N} \mathrm{Var}_{[a_{i-1},a_i]} g + 2\sum_{i=1}^{N} \int_{a_{i-1}}^{a_i} |g(s)| \left| \frac{d}{ds} \frac{1}{T'(s)} \right| ds \\
&\leq \rho \mathrm{Var}_{[0,1]} g + 2K \sum_{i=1}^{N} \int_{a_{i-1}}^{a_i} \frac{|g(s)|}{|T'(s)|} ds.
\end{aligned}$$

Therefore

$$\mathrm{Var}_{[0,1]} P_T g \leq \rho \mathrm{Var}_{[0,1]} g + 2\sum_{i=1}^{N} [K + \frac{1}{a_i - a_{i-1}}] \int_{a_{i-1}}^{a_i} \frac{|g(s)|}{|T'(s)|} ds.$$

Taking $L = \rho(K + \max_i \frac{1}{a_i - a_{i-1}})$, we obtain the Lasota-Yorke inequality (17). $\square$

Remark 4. *The proof works exclusively with densities of bounded variation. Therefore, the result can be extended immediately to: For any function $g \in BV([0,1])$, we have*

$$\frac{1}{N} \sum_{n=0}^{N-1} P_T^n g \to g^*.$$

The convergence is in $L^1(\lambda)$ and $Var_{[0,1]} g^* \leq \frac{L\|g\|_1}{1-\rho}$. By (14), g^* is a bounded density, and the convergence is actually uniform.

Remark 5. *Lasota and Yorke [16] state that the result can be easily extended to expanding maps with countably many branches $T : I_i \to T(I_i)$ where $[0,1] = \cup_i I_i$ modulo nullsets. In the above proof this would cause L to be infinite. This can be mended by assuming that T has the Markov property and every branch has a definite height,* i.e., *there exists $\eta > 0$ such that $|T(I_i)| \geq \eta$ for all i. Indeed, we can estimate the problematic term*

$$\begin{aligned}
\sum_i \frac{1}{|I_i|} \int_{I_i} \frac{|g(s)|}{|T'(s)|} ds &\leq \sup|g| \sum_i \frac{1}{|T'(\xi_i)|} \quad \textit{(for some } \xi_i \in I_i\textit{)} \\
&\leq \sup|g|\, K \sum_i \frac{|I_i|}{|T(I_i)|} \\
&\leq \sup|g|\, \frac{K}{\eta},
\end{aligned}$$

where it is assumed that the distortion $T|_{I_i}$ is bounded by K for all i. We were more precise on these extra assumptions in the Folklore Theorem 9.

11 Bernoulli shifts

Let (Σ, σ, μ) be a Bernoulli shift, say with alphabet $\mathcal{A} = \{1, 2, \dots, N\}$. Here $\Sigma = \mathcal{A}^{\mathbb{Z}}$ (two-sided) or $\Sigma = \mathcal{A}^{\mathbb{N}\cup\{0\}}$ (one-sided), and μ is a stationary product measure with probability vector $(p_1, \dots, p_N)$. Write

$$Z_{[k+1,k+N]}(a_1 \dots a_N) = \{x \in \Sigma : x_{k+1} \dots x_{k+N} = a_1 \dots a_N\}$$

for the cylinder set of length N. If $C = Z_{[k+1,k+R]}$ and $C' = Z_{[l+1,l+S]}$ are two cylinders fixing coordinates on disjoint integer intervals (*i.e.*, $[k+1, k+R] \cap [l+1, l+S] = \varnothing$), then clearly $\mu(C \cap C') = \mu(C)\mu(C')$. This just reflects the independence of disjoint events in a sequence of Bernoulli trials.

Definition 10. *Two measure preserving dynamical systems $(X, \mathcal{B}, T, \mu)$ and $(Y, \mathcal{C}, S, \nu)$ are called* **isomorphic** *if there are $X' \in \mathcal{B}$, $Y' \in \mathcal{C}$ and $\phi : Y' \to X'$ such that*

- $\mu(X') = 1$, $\nu(Y') = 1$;
- $\phi : Y' \to X'$ *is a bi-measurable bijection;*
- ϕ *is measure preserving:* $\nu(\phi^{-1}(B)) = \mu(B)$ *for all* $B \in \mathcal{B}$.
- $\phi \circ S = T \circ \phi$.

Example 2. *The doubling map $T : [0,1] \to [0,1]$ with Lebesgue measure is isomorphic t the one-sided $(\frac{1}{2}, \frac{1}{2})$-Bernoulli shift $(X, \mathcal{B}, \sigma, \mu)$. The isomorphisim is the coding map $\psi : Y' \to X'$, where $Y' = [0,1] \setminus \{$ dyadic rationals in $(0,1)\}$ because these dyadic rationals map to $\frac{1}{2}$ und some iterate of T, and at $\frac{1}{2}$ the coding map is not well defined. Note that $X' = \{0,1\}^{\mathbb{N}} \setminus \{v10^\infty, v01^\infty : v$ is a finite word in the alphabet $\{0,1\}\}$.*

Example 3. *Let $(p_1, \dots, p_N)$ be some probability vector with all $p_i > 0$. Then the one-sided $(p_1, \dots, p_N)$-Bernoulli shift is isomorphic to $([0,1], \mathcal{B}, T, Leb)$ where $T : [0,1] \to [0,1]$ has N linear branches of slope $1/p_i$. The one-sided $(p_1, \dots, p_N)$-Bernoulli shift is* **also** *isomorphic to $([0,1], \mathcal{B}, S, \nu)$ where $S(x) = Nx \pmod 1$. But here ν is another measure that gives $[\frac{i-1}{N}, \frac{i}{N}]$ the mass p_i, and $[\frac{i-1}{N} + \frac{j-1}{N^2}, \frac{i-1}{N} + \frac{j}{N^2}]$ the mass $p_i p_j$, etc.*

Clearly invertible systems cannot be isomorphic to non-invertible systems. But there is a construction to make a non-invertible system invertible, namely by passing to the natural extension.

Definition 11. *Let $(X, \mathcal{B}, \mu, T)$ be a measure preserving dynamical system. A system $(Y, \mathcal{C}, S, \nu)$ is a* **natural extension** *of $(X, \mathcal{B}, \mu, T)$ if there are $X' \in \mathcal{B}$, $Y' \in \mathcal{C}$ and $\phi : Y' \to X'$ such that*

- $\mu(X') = 1$, $\nu(Y') = 1$;

- *$S : Y' \to Y'$ is* **invertible***;*
- *$\phi : Y' \to X'$ is a measurable surjection;*
- *ϕ is measure preserving: $\nu(\phi^{-1}(B)) = \mu(B)$ for all $B \in \mathcal{B}$;*
- *$\phi \circ S = T \circ \phi$.*

Any two natural extensions can be shown to be isomorphic, so it makes sense to speak of **the** natural extension. Sometimes natural extensions have explicit formulas (such as the baker transformation being the natural extension of the doubling map). There is also a general construction: Set

$$Y = \{(x_i)_{i \geqslant 0} : T(x_{i+1}) = x_i \in X \text{ for all } i \geqslant 0\}$$

with $S(x_0, x_1, \dots) = T(x_0), x_0, x_1, \dots$. Then S is invertible (with the left shift $\sigma = S^{-1}$) and

$$\nu(A_0, A_1, A_2, \dots) = \inf_i \mu(A_i) \quad \text{for } (A_0, A_1, A_2 \dots) \subset S,$$

is S-invariant. Now defining $\phi(x_0, x_1, x_2, \dots) := x_0$ makes the diagram commute: $T \circ \phi = \phi \circ S$. Also ϕ is measure preserving because, for each $A \in \mathcal{B}$,

$$\phi^{-1}(A) = (A, T^{-1}(A), T^{-2}(A), T^{-3}(A), \dots)$$

and clearly $\nu(A, T^{-1}(A), T^{-2}(A), T^{-3}(A), \dots) = \mu(A)$ because $\mu(T^{-i}(A)) = \mu(A)$ for every i by T-invariance of μ.

Definition 12. *Let $(X, \mathcal{B}, \mu, T)$ be a measure preserving dynamical system.*

1. *If T is invertible, then the system is called* **Bernoulli** *if it is isomorphic to a Bernoulli shift.*
2. *If T is non-invertible, then the system is called* **one-sided Bernoulli** *if it is isomorphic to a one-sided Bernoulli shift.*
3. *If T is non-invertible, then the system is called* **Bernoulli** *if its natural extension is isomorphic to a one-sided Bernoulli shift.*

The third Bernoulli property is quite general, even though the isomorphism ϕ may be very difficult to find explicitly. Expanding circle maps that satisfy the conditions of Theorem 9 are also Bernoulli, *i.e.,* have a Bernoulli natural extension, see [17]. Being **one-sided Bernoulli**, on the other hand quite, is special. If $T : [0,1] \to [0,1]$ has N linear surjective branches I_i, $i = 1, \dots, N$, then Lebesgue measure m is invariant, and $([0,1], \mathcal{B}, m, T)$ is isomorphic to the one-sided Bernoulli system with probability vector $(|I_1|, \dots, |I_N|)$. If T is piecewise C^2 but not piecewise linear, then it has to be C^2-conjugate to a piecewise linear expanding map to be one-sided Bernoulli, see [7].

12 Mixing and weak mixing

Whereas Bernoulli trials are totally independent, mixing refers to an **asymptotic independence**:

Definition 13. *A probability measure preserving dynamical systems* $(X, \mathcal{B}, \mu, T)$ *is* **mixing** *(or* **strong mixing***) if*

$$\mu(T^{-n}(A) \cap B) \to \mu(A)\mu(B) \text{ as } n \to \infty \tag{20}$$

for every $A, B \in \mathcal{B}$*.*

Proposition 8. *A probability preserving dynamical systems* $(X, \mathcal{B}, T, \mu)$ *is mixing if and only if*

$$\int_X f \circ T^n(x) \cdot \overline{g(x)} \, d\mu \to \int_X f(x) \, d\mu \cdot \int_X \overline{g(x)} \, d\mu \text{ as } n \to \infty \tag{21}$$

for all $f, g \in L^2(\mu)$*, or written in the notation of the Koopman operator* $U_T f = f \circ T$ *and inner product* $\langle f, g\rangle = \int_X f(x) \cdot \overline{g(x)} \, d\mu$*:*

$$\langle U_T^n f, g\rangle \to \langle f, 1\rangle\langle 1, g\rangle \text{ as } n \to \infty. \tag{22}$$

Proof. The "if"-direction follows by taking indicator functions $f = 1_A$ and $g = 1_B$. For the "only if"-direction, general $f, g \in L^2(\mu)$ can be approximated by linear combinations of indicator functions. □

Definition 14. *A probability measure preserving dynamical systems* $(X, \mathcal{B}, \mu, T)$ *is* **weak mixing** *if in average*

$$\frac{1}{n}\sum_{i=0}^{n-1} |\mu(T^{-i}(A) \cap B) - \mu(A)\mu(B)| \to 0 \text{ as } n \to \infty \tag{23}$$

for every $A, B \in \mathcal{B}$*.*

We can express ergodicity in analogy of (20) and (23):

Lemma 7. *A probability preserving dynamical system* $(X, \mathcal{B}, T, \mu)$ *is ergodic if and only if*

$$\frac{1}{n}\sum_{i=0}^{n-1} \mu(T^{-i}(A) \cap B) - \mu(A)\mu(B) \to 0 \text{ as } n \to \infty,$$

for all $A, B \in \mathcal{B}$*. (Compared to* (23)*, note the absence of absolute value bars.)*

Proof. Assume that T is ergodic, so by Birkhoff's Ergodic Theorem $\frac{1}{n}\sum_{i=0}^{n-1} 1_A \circ T^i(x) \to \mu(A)$ for μ-a.e. x. Multiplying by 1_B gives

$$\frac{1}{n}\sum_{i=0}^{n-1} 1_A \circ T^i(x) 1_B(x) \to \mu(A)1_B(x) \quad \mu\text{-a.e.}$$

Integrating over x (using the Dominated Convergence Theorem to swap limit and integral), gives $\lim_n \frac{1}{n}\sum_{i=0}^{n-1} \int_X 1_A \circ T^i(x) 1_B(x)\, d\mu = \mu(A)\mu(B)$.

Conversely, assume that $A = T^{-1}A$ and take $B = A$. Then we obtain $\mu(A) = \frac{1}{n}\sum_{i=0}^{n-1} \mu(T^{-i}(A)) \to \mu(A)^2$, hence $\mu(A) \in \{0, 1\}$. □

Theorem 15. *We have the implications:*

Bernoulli ⇒ mixing ⇒ weak mixing ⇒ ergodic ⇒ recurrent.

None of the reverse implications holds in general.

Proof. Bernoulli ⇒ mixing holds for any pair of cylinder sets C, C' because $\mu(\sigma^{-n}(C) \cap C) = \mu(C)\mu(C')$ for n sufficiently large. The property carries over to all measurable sets by the Kolmogorov Extension Theorem.

Mixing ⇒ weak mixing is immediate from the definition.

Weak mixing ⇒ ergodic: Let $A = T^{-1}(A)$ be a measurable T-invariant set. Then by weak mixing $\mu(A) = \frac{1}{n}\sum_{i=0}^{n-1} \mu(T^{-i}(A) \cap A) \to \mu(A)\mu(A) = \mu(A^2)$. This means that $\mu(A) = 0$ or 1.

Ergodic ⇒ recurrent. If $B \in \mathcal{B}$ has positive measure, then $A := \cup_{i \in \mathbb{N}} T^{-i}(B)$ is T-invariant up to a set of measure 0, see the Poincaré Recurrence Theorem. By ergodicity, $\mu(A) = 1$, and this is the definition of recurrence, see Definition 6. □

We say that a subset $E \subset \mathbb{N} \cup \{0\}$ has **density** zero if $\lim_n \frac{1}{n}\#(E \cap \{0, \dots, n-1\}) = 0$.

Lemma 8. *Let $(a_i)_{i \geqslant 0}$ be a bounded non-negative sequence of real numbers. Then $\lim_n \frac{1}{n}\sum_{i=0}^{n-1} a_i = 0$ if and only if there is a sequence E of zero density in $\mathbb{N} \cup \{0\}$ such that $\lim_{E \not\ni n \to \infty} a_n = 0$.*

Proof. ⇐: Assume that $\lim_{E \not\ni n \to \infty} a_n = 0$ and for $\varepsilon > 0$, take N such that $a_n < \varepsilon$ for all $E \not\ni n \geqslant N$. Also let $A = \sup a_n$. Then

$$\begin{aligned} 0 &\leqslant \frac{1}{n}\sum_{i=0}^{n-1} a_i = \frac{1}{n}\sum_{E \not\ni i=0}^{n-1} a_i + \frac{1}{n}\sum_{E \ni i=0}^{n-1} a_i \\ &\leqslant \frac{NA + (n-N)\varepsilon}{n} + A\frac{1}{n}\#(E \cap \{0, \dots, n-1\}) \to \varepsilon, \end{aligned}$$

as $n \to \infty$. Since $\varepsilon > 0$ is arbitrary, $\lim_n \frac{1}{n} \sum_{i=0}^{n-1} a_i = 0$.

$\Rightarrow$: Let $E_m = \{n : a_n \geqslant \frac{1}{m}\}$. Then clearly $E_1 \subset E_2 \subset E_3 \subset \dots$ and each E_m has density 0 because

$$0 = m \cdot \lim_n \frac{1}{n} \sum_{i=0}^{n-1} a_i \geqslant \lim_n \frac{1}{n} \sum_{i=0}^{n-1} 1_{E_m}(i) = \lim_n \frac{1}{n} \#(E_m \cap \{0, \dots n-1\}).$$

Now take $0 = N_0 < N_1 < N_2 < \dots$ such that $\frac{1}{n}\#(E_m \cap \{0, \dots, n-1\}) < \frac{1}{m}$ for every $n \geqslant N_{m-1}$. Let $E = \cup_m (E_m \cap \{N_{m-1}, \dots, N_m - 1\})$.

Then, taking $m = m(n)$ maximal such that $N_{m-1} < n$,

$$\begin{aligned} &\frac{1}{n} \# (E \cap \{0, \dots, n-1\}) \\ &\quad\leqslant \frac{1}{n}\#(E_{m-1} \cap \{0, \dots, N_{m-1} - 1\}) + \frac{1}{n}\#(E_m \cap \{N_{m-1}, \dots, n-1\}) \\ &\quad\leqslant \frac{1}{N_{m-1}}\#(E_{m-1} \cap \{0, \dots, N_{m-1} - 1\}) + \frac{1}{n}\#(E_m \cap \{0, \dots, n-1\}) \\ &\quad\leqslant \frac{1}{m-1} + \frac{1}{m} \to 0 \end{aligned}$$

as $n \to \infty$. □

Corollary 2. *For a non-negative sequence $(a_n)_{n \geqslant 0}$ of real numbers, $\lim_n \frac{1}{n} \sum_{i=0}^{n-1} a_i = 0$ if and only if $\lim_n \frac{1}{n} \sum_{i=0}^{n-1} a_i^2 = 0$.*

Proof. By the previous lemma, $\lim_n \frac{1}{n} \sum_{i=0}^{n-1} a_i = 0$ if and only if $\lim_{E \not\ni n \to \infty} a_n = 0$ for a set E of zero density. But the latter is clearly equivalent to $\lim_{E \not\ni n \to \infty} a_n^2 = 0$ for the same set E. Applying the lemma again, we have $\lim_n \frac{1}{n} \sum_{i=0}^{n-1} a_i^2 = 0$. □

Example 4. *Let $R_\alpha : \mathbb{S}^1 \to \mathbb{S}^1$ be an irrational circle rotation; it preserves Lebesgue measure. We claim that R_α is not mixing or weak mixing, but it is ergodic. To see why R_α is not mixing, take an interval A of length $\frac{1}{4}$. There are infinitely many n such that $R_\alpha^{-n}(A) \cap A = \varnothing$, so $\liminf_n \mu(R^{-n}(A) \cap A) = 0 \neq (\frac{1}{4})^2$. However, R_α has a non-constant eigenfunction $\psi : \mathbb{S}^1 \to \mathbb{C}$ defined as $\psi(x) = e^{2\pi i x}$ because $\psi \circ R_\alpha(x) = e^{2\pi i (x+\alpha)} = e^{2\pi i \alpha}\psi(x)$. Therefore R_α is not weak mixing, see Theorem 16 below. To prove ergodicity, we show that every T-invariant function $\psi \in L^2(m)$ must be constant. Indeed, write $\psi(x) = \sum_{n \in \mathbb{Z}} a_n e^{2\pi i n x}$ as a Fourier series. The T-invariance implies that $a_n e^{2\pi i n \alpha} = a_n$ for all $n \in \mathbb{Z}$. Since $\alpha \notin \mathbb{Q}$, this means that $a_n = 0$ for all $n \neq 0$, so $\psi(x) \equiv a_0$ is indeed constant.*

Theorem 16. *Let $(X, \mathcal{B}, \mu, T)$ be a probability measure preserving dynamical system. Then the following are equivalent:*

1. *$(X, \mathcal{B}, \mu, T)$ is weak mixing;*

2. $\lim_n \frac{1}{n}\sum_{i=0}^{n-1} |\langle f \circ T^i, g\rangle - \langle f, 1\rangle\langle 1, g\rangle| = 0$ *for all $L^2(\mu)$ functions f, g;*

3. $\lim_{E \not\ni n \to \infty} \mu(T^{-n}A \cap B) = \mu(A)\mu(B)$ *for all $A, B \in \mathcal{B}$ and a subset E of zero density;*

4. $T \times T$ is weak mixing;

5. $T \times S$ is ergodic on (X, Y) for every ergodic system $(Y, \mathcal{C}, \nu, S)$;

6. $T \times T$ is ergodic;

7. The Koopman operator U_T has no measurable eigenfunctions other than constants.

Proof. 2. $\Rightarrow$ 1. Take $f = 1_A$, $g = 1_B$.
1. $\Leftrightarrow$ 3. Use Lemma 8 for $a_i = |\mu(T^{-i}(A) \cap B) - \mu(A)\mu(B)|$.
3. $\Rightarrow$ 4. For every $A, B, C, D \in \mathcal{B}$, there are subsets E_1 and E_2 of $\mathbb{N}$ of zero density such that

$$\lim_{E_1 \not\ni n \to \infty} \mu(T^{-n}(A) \cap B) - \mu(A)\mu(B)| = \lim_{E_2 \not\ni n \to \infty} \mu(T^{-n}(C) \cap D) - \mu(C)\mu(D)| = 0.$$

The union $E = E_1 \cup E_2$ still has density 0, and

$$\begin{aligned} 0 \leqslant \lim_{E \not\ni n \to \infty} &| \ \mu \times \mu \ ((T \times T)^{-n}(A \times C) \cap (B \times D)) - \mu \times \mu(A \times B) \cdot \mu \times \mu(C \times D)| \\ = &\quad \lim_{E \not\ni n \to \infty} |\mu(T^{-n}(A) \cap B) \cdot \mu(T^{-n}(C) \cap D) - \mu(A)\mu(B)\mu(C)\mu(D)| \\ \leqslant &\quad \lim_{E \not\ni n \to \infty} \mu(T^{-n}(A) \cap B) \cdot |\mu(T^{-n}(C) \cap D) - \mu(C)\mu(D)| \\ &\qquad + \lim_{E \not\ni n \to \infty} \mu(C)\mu(D) \cdot |\mu(T^{-n}(A) \cap B) - \mu(A)\mu(B)| = 0. \end{aligned}$$

4. $\Rightarrow$ 5. If $T \times T$ is weakly mixing, then so is T itself. Suppose $(Y, \mathcal{C}, \nu, S)$ is an ergodic system, then, for $A, B \in \mathcal{B}$ and $C, D \in \mathcal{C}$ we have

$$\begin{aligned} &\frac{1}{n}\sum_{i=0}^{n-1} \mu \ (T^{-i}(A) \cap B)\nu(S^{-i}(C) \cap D) \\ &\quad = \frac{1}{n}\sum_{i=0}^{n-1} \mu(A)\mu(B)\nu(S^{-i}(C) \cap D) \\ &\qquad + \frac{1}{n}\sum_{i=0}^{n-1}(\mu(T^{-i}(A) \cap B) - \mu(A)\mu(B))\nu(S^{-i}(C) \cap D). \end{aligned}$$

By ergodicity of S (see Lemma 7), $\frac{1}{n}\sum_{i=0}^{n-1} \nu(S^{-i}(C) \cap D) \to \mu(C)\mu(D)$, so the first term in the above expression tends to $\mu(A)\mu(B)\mu(C)\mu(D)$. The second term is majorised by $\frac{1}{n}\sum_{i=0}^{n-1} |\mu(T^{-i}(A) \cap B) - \mu(A)\mu(B)|$, which tends to 0 because T is weak mixing.

5. $\Rightarrow$ 6. By assumption $T \times S$ is ergodic for the trivial map $S : \{0\} \to \{0\}$. Therefore T itself is ergodic, and hence $T \times T$ is ergodic.

6. $\Rightarrow$ 7. Suppose f is an eigenfunction with eigenvalue λ. The Koopman operator is an isometry (by T-invariance of the measure), so $\langle f, f\rangle = \langle U_T f, U_T f\rangle = \langle \lambda f, \lambda f\rangle = |\lambda|^2\langle f, f\rangle$, and $|\lambda| = 1$. Write $\psi(x, y) = f(x)\overline{f}(y)$. Then

$$\psi \circ (T \times T)(x, y) = \psi(Tx, Ty) = f(Tx)\overline{f(Ty)} = |\lambda|^2\psi(x, y) = \psi(x, y),$$

so ψ is $T \times T$-invariant. By ergodicity of $T \times T$, ψ must be constant $\mu \times \mu$-a.e. But then also f must be constant μ-a.e.

7. $\Rightarrow$ 2. This is the hardest step; it relies on spectral theory of unitary operators. If ψ is an eigenfunction of U_T, then by assumption, ψ is constant, so the eigenvalue is 1. Let $V = \text{span}\ (\psi)$ and Π_1 is the orthogonal projection onto V; clearly $V^\perp = \{f \in L^2(\mu) : \int f\ d\mu = 0\}$. One can derive that the spectral measure ν_T cannot have any atoms, except possibly at Π_1.

Now take $f \in V^\perp$ and $g \in L^2(\mu)$ arbitrary. Using the Spectral Theorem 13, we have

$$\begin{aligned}
\frac{1}{n}\sum_{i=0}^{n-1} |\langle U_T^i f, g\rangle|^2 &= \frac{1}{n}\sum_{i=0}^{n-1} \left| \int_{\sigma(U_T)} \lambda^i \langle \Pi_\lambda f, g\rangle\ d\nu_T(\lambda) \right|^2 \\
&= \frac{1}{n}\sum_{i=0}^{n-1} \int_{\sigma(U_T)} \lambda^i \langle \Pi_\lambda f, g\rangle\ d\nu_T(\lambda) \overline{\int_{\sigma(U_T)} \kappa^i \langle \Pi_\kappa f, g\rangle\ d\nu_T(\kappa)} \\
&= \frac{1}{n}\sum_{i=0}^{n-1} \int\int_{\sigma(U_T)\times\sigma(U_T)} \lambda^i\overline{\kappa}^i\ \langle \Pi_\lambda f, g\rangle \overline{\langle \Pi_\kappa f, g\rangle}\ d\nu_T(\lambda)\ d\nu_T(\kappa) \\
&= \int\int_{\sigma(U_T)\times\sigma(U_T)} \frac{1}{n}\sum_{i=0}^{n-1} \lambda^i\overline{\kappa}^i\ \langle \Pi_\lambda f, g\rangle \overline{\langle \Pi_\kappa f, g\rangle}\ d\nu_T(\lambda)\ d\nu_T(\kappa) \\
&= \int\int_{\sigma(U_T)\times\sigma(U_T)} \frac{1}{n}\frac{1-(\lambda\overline{\kappa})^n}{1-\lambda\overline{\kappa}}\ \langle \Pi_\lambda f, g\rangle \overline{\langle \Pi_\kappa f, g\rangle}\ d\nu_T(\lambda)\ d\nu_T(\kappa),
\end{aligned}$$

where in the final line we used that the diagonal $\{\lambda = \kappa\}$ has $\nu_T \times \nu_T$-measure zero, because ν is non-atomic (except possibly the atom Π_1 at $\lambda = 1$, but then $\Pi_1 f = 0$). Now $\frac{1}{n}\frac{1-(\lambda\overline{\kappa})^n}{1-\lambda\overline{\kappa}}$ is bounded (use l'Hôpital's rule) and tends to 0 for $\lambda \neq \kappa$, so by the Bounded Convergence Theorem, we have

$$\lim_n \frac{1}{n}\sum_{i=0}^{n-1} |\langle U_T^i f, g\rangle|^2 = 0.$$

Using Corollary 2, we derive that also $\lim_n \frac{1}{n}\sum_{i=0}^{n-1} |\langle U_T^i f, g\rangle| = 0$ (*i.e.,* without the square). Finally, if $f \in L^2(\mu)$ is arbitrary, then $f - \langle f, 1\rangle \in V^\perp$. We find

$$
\begin{aligned}
0 &= \lim_n \frac{1}{n} \sum_{i=0}^{n-1} |\langle U_T^i(f - \langle f, 1\rangle), g\rangle| \\
&= \lim_n \frac{1}{n} \sum_{i=0}^{n-1} |\langle U_T^i f - \langle f, 1\rangle, g\rangle| \\
&= \lim_n \frac{1}{n} \sum_{i=0}^{n-1} |\langle U_T^i f, g\rangle - \langle f, 1\rangle\langle 1, g\rangle|
\end{aligned}
$$

and so property 2. is verified. □

13 Cutting and Stacking

The purpose of **cutting and stacking** is to create invertible maps of the interval that preserve Lebesgue measure, and have further good properties such as "unique ergodicity", "not weak mixing", or rather the opposite "weak mixing but not strong mixing". Famous examples due to Kakutani and to Chacon achieve this, and we will present them here.

The procedure is as follows:

- Cut the unit interval into several intervals, say $A, B, C, \dots$ (these will become the **stacks**), and a remaining interval S.
- Cut each interval into parts (a fixed finite number for each stack), and also cut of some intervals from S.
- Pile the parts of the stacks and the cut-off pieces of S on top of the stacks, according to some fixed rule. By choosing the parts in the previous step of the correct size, we can ensure that all intervals in each separate stack have the same size; they can therefore be neatly aligned vertically.
- Map every point on a level of a stack directly to the level above. Then every point has a well-defined image (except for points at the top levels in a stack and points in the remaindeer of S), and also a well-defined preimage (except for points at a bottom level in a stack and points in the remainder of S). Where defined, Lebesgue measure is preserved.
- Repeat the process, now slicing vertically through whole stacks and stacking whole stacks on top of other stacks, possibly putting some intervals of S in between. Wherever the map was defined at a previous step, the definition remains the same.

- Keep repeating. Eventually, the measure of points where the map is not defined tends to zero. In the end, assuming that the interval S will be entirely spent, there will only be one point for each stack without image and one points in each stack without preimage. We can take an arbitrary bijection between them to define the map everywhere.

- The resulting transformation of the interval is invertible and preserves Lebesgue measure. The number of stacks used is called the **rank** of the transformation.

Example 5 (Kakutani). *Take one stack, so start with $A = [0,1]$. Cut it in half and put the right half on top of the left half. Repeat this procedure. Let us call the result limit map $T : [0,1] \to [0,1]$ the Kakutani map. The resulting formula is:*

$$T(x) = \begin{cases} x + \frac{1}{2} & \textit{if } x \in [0, \frac{1}{2}); \\ x - \frac{1}{4} & \textit{if } x \in [\frac{1}{2}, \frac{3}{4}); \\ x - \frac{3}{4} + \frac{1}{8} & \textit{if } x \in [\frac{3}{4}, \frac{7}{8}); \\ \vdots & \vdots \\ x - (1 - \frac{1}{2^n}) + \frac{1}{2^{n+1}} & \textit{if } x \in [1 - \frac{1}{2^n}, 1 - \frac{1}{2^{n+1}}), n \geqslant 1, \end{cases}$$

see Figure 1. If $x \in [0,1)$ is written in base 2, i.e.,

$$x = 0.b_1 b_2 b_3 \dots \qquad b_i \in \{0,1\}, \quad x = \sum_i b_i 2^{-i},$$

then T acts as the **adding machine** *or* **odometer***: add 0.1 with carry. That is, if $k = \min\{i \geqslant 1 : b_i = 0\}$, then $T(0.b_1 b_2 b_3 \dots) = 0.001 b_{k+1} b_{k+2} \dots$. If $k = \infty$, so $x = 0.111111\dots$, then $T(x) = 0.0000\dots$.*

Proposition 9. *The Kakutani map $T : [0,1] \to [0,1]$ of cutting and stacking is uniquely ergodic, but not weakly mixing.*

Proof. The map T permutes the dyadic intervals cyclically. For example $T((0, \frac{1}{2})) = (\frac{1}{2}, 1))$ and $T((\frac{1}{2}, 1)) = (0, \frac{1}{2})$. Therefore, $f(x) = 1_{(0,\frac{1}{2})} - 1_{(\frac{1}{2},1)}$ is an eigenfunction for eigenvalue -1. Using four intervals, we can construct (complex-valued) eigenfunctions for eigenvalues $\pm i$. In generality, all the numbers $e^{2\pi i m/2^n}$, $m, n \in \mathbb{N}$ are eigenvalues, and the corresponding eigenfunctions span $L^2(m)$. This property is called **pure point spectrum**. In any case, T is not weakly mixing.

Now for unique ergodicity, we use the fact again that T permutes the dyadic intervals cyclically. Call these intervals $D_{j,N} = [\frac{j}{2^N}, \frac{j+1}{2^N})$ for $N \in \mathbb{N}$ and $j = \{0, 1, \dots, 2^N - 1\}$, and if $x \in [0,1)$, we indicate the dyadic interval containing it by $D_{j,N}(x)$. Let

$$\begin{cases} \overline{f}_N(x) = \sup_{t \in D_{j,N}(x)} f(t), \\ \underline{f}_N(x) = \inf_{t \in D_{j,N}(x)} f(t), \end{cases}$$

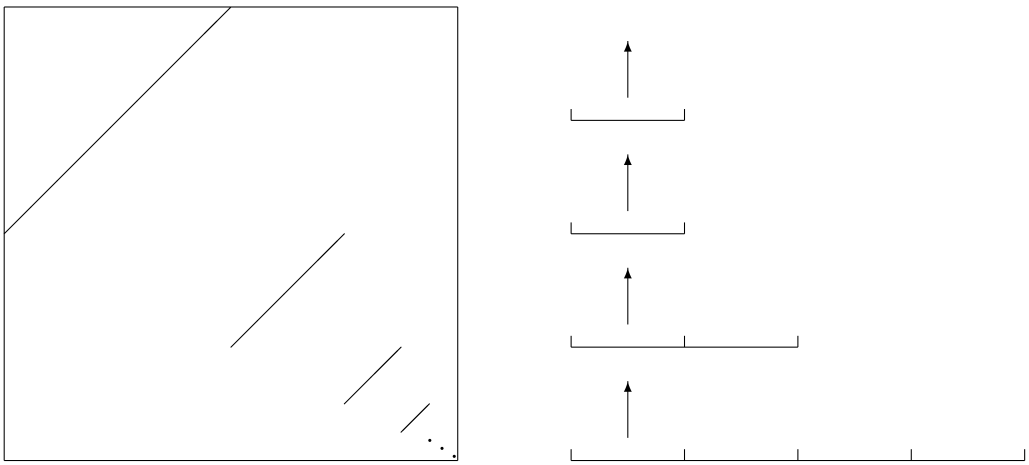

Figure 1: The Kakutani map $T : [0,1] \to [0,1]$ resulting from cutting and stacking.

be step-functions that we can use to compute the Riemann integral of f. That is:

$$\int \overline{f}_N(s)ds := \frac{1}{2^N} \sum_{j=0}^{2^N-1} \sup_{t \in D_{j,N}} f(t) \geqslant \int f(s)ds \geqslant \int \underline{f}_N(s)ds := \frac{1}{2^N} \sum_{j=0}^{2^N-1} \inf_{t \in D_{j,N}} f(t).$$

For continuous (or more generally Riemann integrable) functions, $\int \overline{f}_N dx - \int \underline{f}_N dx \to 0$ as $N \to \infty$, and their common limit is called the Riemann integral of f.

According to Theorem 3, we need to show that $\frac{1}{n}\sum_{i=0}^{N-1} f \circ T^i(x)$ converges uniformly to a constant (for each continuous function f) to show that T is uniquely ergodic, *i.e.,* Lebesgue measure is the unique invariant measure.

Let $f : [0,1] \to \mathbb{R}$ be continuous and $\varepsilon > 0$ be given. By uniform continuity, we can find N such that $\max_j(\sup_{t \in D_{j,N}} f(t) - \inf_{t \in D_{j,N}} f(t)) < \varepsilon$. Write $n = m2^N + r$. Any orbit x will visit all intervals $D_{j,N}$ cyclically before returning close to itself, and hence visit each $D_{j,N}$ exactly m times in the first $m2^N$ iterates. Therefore

$$\begin{aligned}
\frac{1}{n}\sum_{i=0}^{n-1} f \circ T^i(x) &\leqslant \frac{1}{m2^N + r}\left(\sum_{j=0}^{2^N-1} m \sup_{t \in D_{j,N}} f(t) + r\|f\|_\infty\right) \\
&\leqslant \frac{1}{2^N}\sum_{j=0}^{2^N-1} \sup_{t \in D_{j,N}} f(t) + \frac{r\|f\|_\infty}{m2^N + r} \\
&= \int \overline{f}_N(s)ds + \frac{r\|f\|_\infty}{m2^N + r} \to \int \overline{f}_N(s)ds,
\end{aligned}$$

as $m \to \infty$. A similar computation gives $\frac{1}{n}\sum_{i=0}^{n-1} f \circ T^i(x) \geqslant \int \underline{f}_N(x)dx$. As $\varepsilon \to 0$ (and hence $N \to \infty$), we get convergence to the integral $\int f(s)ds$, independently of the initial point x. $\square$

Example 6 (Chacon). *Take one stack and one stack: $A_0 = [0, \frac{2}{9})$ and $S = [\frac{2}{3}, 1)$. Cut A_0 is three equal parts and cut $[\frac{2}{3}, \frac{8}{9})$ from S. Pile the middle interval $[\frac{2}{9}, \frac{4}{9})$ on the left, then the cut-off piece $[\frac{2}{3}, \frac{8}{9})$ of S, and then remaining interval $[\frac{4}{9}, \frac{2}{3})$. The stack can now be coded upward as $A_1 = A_0A_0SA_0$.*

Repeat this procedure: cut the stack vertically in three stacks (of width $\frac{2}{27}$), cut an interval $[\frac{8}{9}, \frac{26}{27})$ from S, and pile them on top of one another: middle stack on left, then the cut-off piece of S, and then the remaining third of the stack. The stack can now be coded upward as $A_2 = A_1A_1SA_1$.

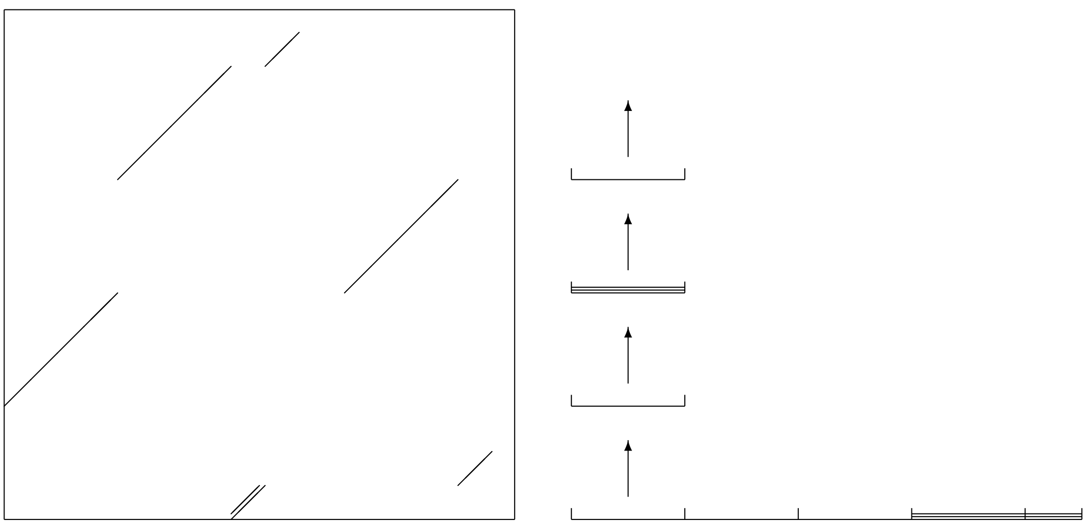

Figure 2: The Chacon map $T : [0,1] \to [0,1]$ resulting from cutting and stacking.

Proposition 10. *The Chacon map $T : [0,1] \to [0,1]$ of cutting and stacking is uniquely ergodic, weakly mixing but not strongly mixing.*

Sketch of Proof. First some observations on the symbolic pattern that emerges of the Chacon cutting and stacking. When stacking intervals, their labels follow the following pattern

$$\underbrace{\underbrace{\underbrace{A_0A_0SA_0}_{A_1}\underbrace{A_0A_0SA_0}_{A_1}\,S\,\underbrace{A_0A_0SA_0}_{A_1}}_{A_2}\underbrace{\underbrace{A_0A_0SA_0}_{A_1}\underbrace{A_0A_0SA_0}_{A_1}\,S\,\underbrace{A_0A_0SA_0}_{A_1}}_{A_2}\,S\,\underbrace{\underbrace{A_0A_0SA_0}_{A_1}\underbrace{A_0A_0SA_0}_{A_1}\,S\,\underbrace{A_0A_0SA_0}_{A_1}}_{A_2}}_{A_3}$$

This pattern is the same at every level; we could have started with A_n, grouped together as $A_{n+1} = A_nA_nSA_n$, etc. At step n in the construction of the tower, the width of the stack is $w_n = \frac{2}{3}(3^{-(n+1)}$ and the length of the the word A_n is $l_n = \frac{1}{2}(3^{n+1} - 1)$.

The frequency of each block $\sigma^k(A_n)$ is almost the same in every block huge block B, regardless where taken in the infinite string. This observation leads to unique ergodicity (similar although a bit more involved as in the case of the Kakutani map), but we will skip the details.

Instead, we focus on the weak mixing. Clearly the word A_n appears in triples, and also as $A_nA_nA_nSA_nA_nA_n$. To explain the idea behind the proof, pretend that an eigenfunction (with eigenvalue $e^{2\pi i\lambda}$) were constant on any set E whose code is A_n (or $\sigma^k A_n$ for some $0 \leqslant k < l_n$, where σ denotes the left-shift). Such sets E are intervals of width w_n. Then

$$f \circ T^{l_n}|_E = e^{2\pi i\lambda l_n} f|_E \quad \text{and} \quad f \circ T^{2l_n+1}|_E = e^{2\pi i\lambda l_n} f|_E.$$

This gives $1 = e^{2\pi i\lambda l_n} = e^{2\pi i\lambda(2l_n+1)}$, so $\lambda = 0$, and the eigenvalue is 1 after all.

The rigorous argument is as follows. Suppose that $f(x) = e^{2\pi i\vartheta(x)}$ were an eigenfunction for eigenvalue $e^{2\pi i\lambda}$ and a measurable function $\vartheta : \mathbb{S}^1 \to \mathbb{R}$. By Lusin's Theorem, we can find a subset $F \subset \mathbb{S}^1$ of Lebesgue measure $\geqslant 1 - \varepsilon$ such that ϑ is uniformly continuous on F. Choose $\varepsilon > 0$ arbitrary, and take N so large that the variation of ϑ is less that ε on any set of the form $E \cap F$, where points in E have code starting as $\sigma^k(A_N)$, $0 \leqslant k < l_N$. Sets of this type fill a set E^* with mass at least half of the unit interval.

Because of the frequent occurrence of $A_NA_NA_NSA_NA_NA_N$, a definite proportion of E^* is covered by set E with the property that such that $T^{2l_N+1} \cap T^{l_N} E \cap E \neq \varnothing$, because they have codes of length l_N that reappear after both l_N and $2l_N + 1$ shifts. For x in this intersection,

$$\begin{cases} \vartheta \circ T^{2l_N+1}(x) = (l_N + 1)\lambda + \vartheta \circ T^{l_N}(x) \pmod 1 \\ \vartheta \circ T^{l_N}(x) = l_N\lambda + \vartheta(x) \pmod 1 \end{cases}$$

where all three point $x, T^{l_N}(x), T^{2l_N+1}(x)$ belong to the same copy E. Subtracting the two equations gives

$$\lambda \bmod 1 = \vartheta \circ T^{2l_N+1}(x) - \vartheta \circ T^{l_N}(x) + \vartheta(x) - \vartheta \circ T^{l_N}(x) \leqslant 2\varepsilon.$$

But ε is arbitrary, so $\lambda = 0 \bmod 1$ and the eigenvalue is 1.

Now for the strong mixing, consider once more the sets $E = E_{k,n}$ of points whose codes starts as the k-th cyclic permutation of A_n for some $0 \leqslant k < l_n$, that is: the first l_n symbols of $\sigma^k(A_nA_n)$. Their measure is $\mu(E) = w_n$, and for different k, they are disjoint. Furthermore, the only l_n-block appearing are cyclic permutations of A_n or cyclic permutations with pieces of S inserted somewhere. At least half of these appearances are of the first type, so $\mu(\cup_{k=0}^{l_n-1} E_{k,n}) \geqslant \frac{1}{2}$ for each n.

The basic idea is that $\mu(E \cap T^{-l_n} E) \geqslant \frac{1}{3}\mu(E)$ because at least a third of the appearances of A_n is followed by another A_n. But $\frac{1}{3}\mu(E) \gg \mu(E)^2$, as one would expect for mixing. Of course, mixing only says that $\lim_l \mu(Y \cap T^{-l}(E)) = \mu(Y)^2$ only for sets Y not depending on l.

However, let $Y_m = [m/8, (m+1)/8] \subset [0,1]$, $m = 0, \dots, 7$ be the eight dyadic intervals of length 1/8. For each n, at least one Y_m is covered for at least half by sets E of the above type, say a set $Z \subset Y_m$ of measure $\mu(Z) \geqslant \frac{1}{2}\mu(Y_m))$ such that $Z \subset \cup_k E_{k,n}$. That means that

$$\mu(Y_m \cap T^{-l_n}(Y_m)) \geqslant \mu(Z \cap T^{-l_n}(Z)) \geqslant \frac{1}{3}\mu(Z) \geqslant \frac{1}{6}\mu(Y_m) > \mu(Y_m)^2.$$

Let Y be one of the Y_m's for which the above holds for infinitely many n. Then $\limsup_n \mu(Y_m \cap T^{-l_n}(Y_m)) > \mu(Y)^2$, contradicting strong mixing. $\square$

14 Toral automorphisms

The best known example of a toral automorphism (that is, an invertible linear map on the torus $\mathbb{T}^n = \mathbb{S}^1 \times \cdots \times \mathbb{S}^1$) is the Arnol'd cat map. This map $T_C : \mathbb{T}^2 \to \mathbb{T}^2$ is defined as

$$T_C(x,y) = C\begin{pmatrix} x \\ y \end{pmatrix} \pmod 1 \qquad \text{for the matrix } C = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}.$$

The name come from the illustration in Arnol'd's book [3] showing how the head of a cat, drawn on a torus, is distorted by the action of the map[4]. Properties of T_C are:

- C preserves the integer lattice, so T_C is well-defined an continuous.
- $\det(C) = 1$, so Lebesgue measure m is preserved (both by C and T_C). Also C and T_C are invertible, and C^{-1} is still an integer matrix.
- The eigenvalues of C are $\lambda_\pm = (3 \pm \sqrt{5})/2$, and the corresponding eigenspaces $E_\pm$ are spanned $(-1, (\sqrt{5}+1)/2)^T$ and $(1, (\sqrt{5}-1)/2)^T$. These are orthogonal (naturally, since C is symmetric), and have irrational slopes, so they wrap densely in the torus.
- Every rational point in $\mathbb{T}^2$ is periodic under T (as their denominators cannot increase, so T acts here as an invertible map on a finite set). This gives many invaraint measures: the equidistribution on each periodic orbit. Therefore T_C is not uniquely ergodic.

The properties are common to all maps T_A, provided they satisfy the following definition.

Definition 15. *A toral automorphism $T : \mathbb{T}^d \to \mathbb{T}^d$ is an invertible linear map on the (d-dimensional) torus $\mathbb{T}^d$. Each such T is of the form $T_A(x) = Ax \pmod 1$, where the matrix A satisfies:*

[4] Arnol'd didn't seem to like cats, but see the applet https://www.jasondavies.com/catmap/ how the cat survives

- *A is an integer matrix with* $\det(A) = \pm 1$*;*
- *the eigenvalues of A are not on the unit circle; this property is called* **hyperbolicity**.

Somewhat easier to treat that the cat map is T_A for $A = \left(\begin{smallmatrix} 1 & 1 \\ 1 & 0 \end{smallmatrix}\right)$, which is an orientation reversing matrix with $A^2 = C$. The map T_A has a **Markov partition**, that is a partition $\{R_i\}_{i=1}^N$ for sets such that

1. The R_i have disjoint interiors and $\cup_i R_i = \mathbb{T}^d$;
2. If $T_A(R_i) \cap R_j \neq \varnothing$, then $T_A(R_i)$ stretches across R_j in the unstable direction (*i.e.*, the direction spanned by the unstable eigenspaces of A).
3. If $T_A^{-1}(R_i) \cap R_j \neq \varnothing$, then $T_A^{-1}(R_i)$ stretches across R_j in the stable direction (*i.e.*, the direction spanned by the stable eigenspaces of A).

In fact, every hyperbolic toral automorphism has a Markov partition, but in general they are fiendishly difficult to find explicitly. In the case of A, a Markov partition of three rectangles R_i for $i = 1, 2, 3$ can be constructed, see Figure 3.

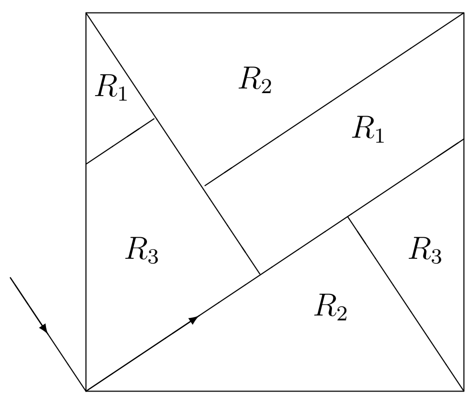

Figure 3: The Markov partition for the toral automorphism T_A. The arrows indicate the stable and unstable directions at $(0, 0)$.

The corresponding transition matrix is

$$B = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \quad \text{where } B_{ij} = \begin{cases} 1 & \text{if } T_A(R_i) \cap R_j \neq \varnothing \\ 0 & \text{if } T_A(R_i) \cap R_j = \varnothing. \end{cases}$$

Note that the characteristic polynomial of B is

$$\det(B - \lambda I) = -\lambda^3 + 2\lambda + 1 = -(\lambda + 1)(\lambda^2 - \lambda - 1) = -(\lambda + 1)\det(A - \lambda I).$$

so B has the eigenvalues of A (no coincidence!), together with $\lambda = -1$. The transition matrix B generates a **subshift of finite type**:

$$\Sigma_B = \{(x_i)_{i\in\mathbb{Z}} : x_i \in \{1,2,3\}, B_{x_i x_{i+1}} = 1 \ \forall\ i \in \mathbb{Z}\},$$

equipped with the left-shift σ. That is, Σ_B contains only sequences in which each $x_i x_{i+1}$ indicate transitions from Markov partition elements that are allowed by the map T_A.

It can be shown that $(\mathbb{T}^d, \mathcal{B}, T, Leb)$ is isomorphic to the shift space $(\Sigma_B, \mathcal{C}, \sigma, \mu)$ where

$$\mu([x_k x_{k+1} \dots x_n]) = m_{x_k} \Pi_{x_k x_{k+1}} \Pi_{x_{k+1} x_{k+2}} \dots \Pi_{x_{n-1} x_n},$$

for $m_i = Leb(R_i)$, $i = 1, \dots, d\}$, and weighted transition matrix Π where

$$\Pi_{ij} = \frac{Leb(T_A(R_i) \cap R_j)}{Leb(R_i)} \text{ is the relative mass that } T_A \text{ transports from } R_i \text{ to } R_j.$$

Finally $\mathcal{C}$ the σ-algebra of set generated by allowed cylinder sets.

Theorem 17. *For every hyperbolic toral automorphism, Lebesgue measure is ergodic and mixing.*

Proof. We only give the proof for dimension 2. The higher dimensional case goes similarly. Consider the Fourier modes (also called **characters**)

$$\chi_{(m,n)} : \mathbb{T}^2 \to \mathbb{C}, \qquad \chi_{(m,n)}(x,y) = e^{2\pi i(mx+ny)}.$$

These form an orthogonal system (w.r.t. $\langle \varphi, \psi \rangle = \int \varphi \overline{\psi} d\lambda$), spanning $L^2(\lambda)$ for Lebesgue measure λ. We have

$$U_{T_A}\chi_{(m,n)}(x,y) = \chi_{(m,n)} \circ T_A(x,y) = \chi_{m,n}(x,y) = e^{2\pi i(am+cn)x+(bm+dn)y)} = \chi_{A^t(m,n)}(x,y).$$

In other words, U_{T_A} maps the character with index (m,n) to the character with index $A^t(m,n)$, where A^t is the transpose matrix.

For the proof of ergodicity, assume that φ is a T_A-invariant L^2-function. Write it as Fourier series:

$$\varphi(x,y) = \sum_{m,n\in\mathbb{Z}} \varphi_{(m,n)}\chi_{(m,n)}(x,y),$$

where the Fourier coefficients $\varphi_{m,n} \to 0$ as $|m| + |n| \to \infty$ By T_A-invariance, we have

$$\varphi(x,y) = \varphi \circ T_A(x,y) = \sum_{m,n\in\mathbb{Z}} \varphi_{(m,n)}\chi_{A^t(m,n)}(x,y),$$

and hence $\varphi_{(m,n)} = \varphi_{A^t(m,n)}$ for all m,n. For $(m,n) = (0,0)$ this is not a problem, but this only produces constant functions. If $(m,n) \neq (0,0)$, then the A^t-orbit of (m,n), so infinitely many equal Fourier coefficients

$$\varphi_{(m,n)} = \varphi_{A^t(m,n)} = \varphi_{(A^t)^2(m,n)} = \varphi_{(A^t)^3(m,n)} = \varphi_{(A^t)^4(m,n)} \cdots$$

As the Fourier coefficients converge to zero as $|m| + |n| \to \infty$, they all must be equal to zero, and hence φ is a constant function. This proves ergodicity.

For the proof of mixing, we need a lemma, which we give without proof.

Lemma 9. *A transformation (X, T, μ) is mixing if and only if for all φ, ψ in a complete orthogonal system spanning $L^2(\mu)$, we have*

$$\int_X \varphi \circ T^N(x)\overline{\psi(x)}\, d\mu \to \int_X \varphi(x)\, d\mu \cdot \int_X \overline{\psi(x)}\, d\mu$$

as $N \to \infty$.

To use this lemma on $\varphi = \chi_{(m,n)}$ and $\psi = \chi_{(k,l)}$, we compute

$$\int_X \chi_{(m,n)} \circ T^N(x)\overline{\chi_{(k,l)}(x)}\, d\lambda = \int_X \chi_{(A^t)^N(m,n)}\overline{\chi_{(k,l)}(x)}\, d\lambda.$$

If $(m,n) = (0,0)$, then $(A^t)^N(m,n) = (0,0) = (m,n)$ for all N. Hence, the integral is non-zero only if $(k,l) = (0,0)$, but then the integral equals 1, which is the same as $\int_X \chi_{(0,0)}\, d\lambda \int_X \overline{\chi_{(0,0)}(x)}\, d\lambda$. If $(k,l) = (0,0)$, then the integral is zero, but so is $\int_X \chi_{(0,0)}\, d\lambda \int_X \overline{\chi_{(0,0)}(x)}\, d\lambda$.

If $(m,n) \neq (0,0)$, then, regardless what (k,l) is, there is N such that $(A^t)^M(m,n) \neq (k,l)$ for all $M \geqslant N$. Therefore

$$\int_X \chi_{(m,n)} \circ T^M(x)\overline{\chi_{(k,l)}(x)}\, d\lambda = 0 = \int_X \chi_{(m,n)}\, d\lambda \int_X \overline{\chi_{(k,l)}(x)}\, d\lambda.$$

The lemma therefore guarantees mixing. □

15 Topological entropy and topological pressure

Topological entropy was first defined in 1965 by Adler et al. [1], but the form that Bowen [4] and Dinaburg [10] redressed it in is commonly used nowadays.

We will start by start giving the original definition, because the idea of joints of covers easily relates to joints of partitions as used in measure-theoretic entropy. After that, we will give Bowen's approach, since it readily generalises to topological pressure as well.

15.1 The original definition

Let (X, d, T) be a continuous map on compact metric space (X, d). We say that $\mathcal{U} = \{U_i\}$ is an *open ε-cover* if all U_i are open sets of diamter $\leqslant \varepsilon$ and $X \subset \bigcup_i U_i$. Naturally, compactness of X guarantees that for every open cover, we can select a finite subcover. Thus, let $\mathcal{N}(\mathcal{U})$ the the minimal possible cardinality of subcovers of $\mathcal{U}$. We say that $\mathcal{U}$ *refines* $\mathcal{V}$ (notation $\mathcal{U} \succeq \mathcal{V}$) if evey $U \in \mathcal{U}$ is contained in a $V \in \mathcal{V}$. If $\mathcal{U} \succeq \mathcal{V}$ then $\mathcal{N}(\mathcal{U}) \geq \mathcal{N}(\mathcal{V})$.

Given two cover $\mathcal{U}$ and $\mathcal{V}$, the joint

$$\mathcal{U} \vee \mathcal{V} := \{U \cap V : U \in \mathcal{U}, V \in \mathcal{V}\}$$

is an open cover again, and one can verify that $\mathcal{N}(\mathcal{U} \vee \mathcal{V}) \leqslant \mathcal{N}(\mathcal{U})\mathcal{N}(\mathcal{V})$. Since T is continuous, $T^{-1}(\mathcal{U})$ is an open cover as well, although in this case it need not be an ε-cover; However, $\mathcal{U} \vee T^{-1}(\mathcal{U})$ is an ε-cover, and it refines $T^{-1}(\mathcal{U})$.

Define the *topological entropy* as

$$h_{top}(T) = \lim_{\varepsilon \to 0} \sup_{\mathcal{U}} \lim_{n} \frac{1}{n} \log \mathcal{N}(\mathcal{U}^n) \qquad \text{for } \mathcal{U}^n := \bigvee_{i=0}^{n-1} T^{-i}(\mathcal{U}), \tag{24}$$

where the supremum is taken over all open ε-covers $\mathcal{U}$. Because $\mathcal{N}(\mathcal{U} \vee \mathcal{V}) \leqslant \mathcal{N}(\mathcal{U})\mathcal{N}(\mathcal{V})$, the sequence $\log \mathcal{N}(\mathcal{U}^n)$ is subadditive, so the limit $\lim_n \frac{1}{n} \log \mathcal{N}(\mathcal{U}^n)$ exists. We have the following properties:

Lemma 10.

- *$h_{top}(T^k) = k h_{top}(T)$ for $k \geqslant 0$. If T is invertible, then also $h_{top}(T^{-1}) = h_{top}(T)$.*
- *If (Y, S) is semiconjugate to (X, T), then $h_{top}(S) \leqslant h_{top}(T)$. In particular, conjugate systems (on compact spaces!) have the same entropy.*

Proof. □

15.2 Topological entropy of interval maps

If $X = [0, 1]$ with the usual Euclidean metric, then there are various shortcuts to compute the entropy of a continuous map $T : [0, 1] \to [0, 1]$. Let us call any maximal interval on which T is monotone a *lap*; the number of laps is denoted as $\ell(T)$. Also, the *variation* of T is defined as

$$Var(T) = \sup_{0 \leqslant x_0 < \dots\, x_N \leqslant N} \sum_{i=1}^{N} |T(x_i) - T(x_{i-1})|,$$

where the supremum runs over all finite collections of points in $[0, 1]$. The following result is due to Misurewicz & Szlenk [19].

Proposition 11. *Let $T : [0,1] \to [0,1]$ have finitely many laps. Then*

$$\begin{aligned} h_{top}(T) &= \lim_{n\to\infty} \frac{1}{n} \log \ell(T^n) \\ &= \limsup_{n\to\infty} \frac{1}{n} \log \#\{\textit{clusters of } n\textit{-periodic points}\} \\ &= \max\{0, \lim_{n\to\infty} \frac{1}{n} \log Var(T^n)\}. \end{aligned}$$

where two n-periodic points are in the same cluster if they belong to the same lap of T^n.

Remark 6. *The identity map has one branch, consisting of (uncountaly many) fixed point, that form one cluster. The map $x \mapsto x + (x/10)^2 \sin(\pi/x) \mod 1$ has also one branch, but with countably many fixed point, forming one cluster. For an expanding map, every branch can contain only one fixed point.*

Proof. Since the variation of a monotone function is given by $\sup T - \inf T$, and due to the definition of "cluster" of n-periodic points, the inequalities

$$\#\{\text{clusters of } n\text{-periodic points}\}, \mathrm{Var}(T^n) \leqslant \ell(T^n)$$

are immediate. For a lap I of T^n, let $\gamma := |T^n(I)|$ be its *height*. We state without proof (cf. [6, Chapter 9]):

$$\begin{array}{c} \text{For every } \delta > 0, \text{ there is } \gamma > 0 \text{ such that} \\ \#\{J : J \text{ is a lap of } T^n,\ |T^n(J)| > \gamma\} \geq 1 - \delta)^n \ell(T_n). \end{array} \tag{25}$$

This means that $Var(T^n) \geqslant \gamma(1-\delta)^n \ell(T^n)$, and therefore

$$-2\delta + \lim_n \frac{1}{n} \log \ell(T^n) \leqslant \lim_n \frac{1}{n} \log \mathrm{Var}(T^n) \leqslant \lim_n \frac{1}{n} \log \ell(T^n).$$

Since δ is arbitrary, both above quantities are all equal.

Making the further assumption (without proof[5]) that there is $K = K(\gamma)$ such that $\cup_{i=0}^K T^i(J) = X$ for every interval of length $|J| \geqslant \gamma$, we also find that

$$\#\{\text{clusters of } n+i\text{-periodic points}, 0 \leqslant i \leqslant K\} \geqslant (1-\delta)^n \ell(T^n).$$

This implies that

$$-2\delta + \lim_n \frac{1}{n} \log \ell(T^n) \leqslant \limsup_n \frac{1}{n} \max_{0 \leqslant i \leqslant K} \log \#\{\text{clusters of } n+i\text{-periodic points}\}$$

so also $\lim_n \frac{1}{n} \log \ell(T^n) = \limsup_{n\to\infty} \frac{1}{n} \log \#\{\text{clusters of } n\text{-periodic points}\}$

If $\varepsilon > 0$ is so small that the width of every lap is greater than 2ε, then for every ε-cover $\mathcal{U}$, every subcover of $\mathcal{U}^n$ has at least one element in each lap of T^n. Therefore $\ell(T^n) \leqslant \mathcal{N}(\mathcal{U}^n)$ for every ε-cover, so $\lim_n \frac{1}{n} \log \ell(T^n) \leqslant h_{top}(T)$. □

[5] In fact, it is not entirely true if T has an invariant subset attracting an open neighbourhood. But it suffices to restrict T to its *nonwandering set*, that is, the set $\Omega(T) = \{x \in X : x \in \cup_{n\geqslant 1} T^n(U))$ for every neighbourhood $U \ni x\}$, because $h_{top}(T) = h_{top}(T|_{\Omega(T)})$.

15.3 Bowen's approach

Let T be map of a compact metric space (X, d). If my eyesight is not so good, I cannot distinguish two points $x, y \in X$ if they are at a distance $d(x, y) < \varepsilon$ from one another. I may still be able to distinguish there orbits, if $d(T^k x, T^k y) > \varepsilon$ for some $k \geqslant 0$. Hence, if I'm willing to wait $n-1$ iterations, I can distinguish x and y if

$$d_n(x, y) := \max\{d(T^k x, T^k y) : 0 \leqslant k < n\} > \varepsilon.$$

If this holds, then x and y are said to be (n, ε)**-separated**. Among all the subsets of X of which all points are mutually (n, ε)-separated, choose one, say $E_n(\varepsilon)$, of maximal cardinality. Then $s_n(\varepsilon) := \#E_n(\varepsilon)$ is the maximal number of n-orbits I can distinguish with ε-poor eyesight.

The **topological entropy** is defined as the limit (as $\varepsilon \to 0$) of the exponential growth-rate of $s_n(\varepsilon)$:

$$h_{top}(T) = \lim_{\varepsilon \to 0} \limsup_{n \to \infty} \frac{1}{n} \log s_n(\varepsilon). \tag{26}$$

Note that $s_n(\varepsilon_1) \geqslant s_n(\varepsilon_2)$ if $\varepsilon_1 \leqslant \varepsilon_2$, so $\limsup_n \frac{1}{n} \log s_n(\varepsilon)$ is a decreasing function in ε, and the limit as $\varepsilon \to 0$ indeed exists.

Instead of (n, ε)-separated sets, we can also work with (n, ε)**-spanning** sets, that is, sets that contain, for every $x \in X$, a y such that $d_n(x, y) \leqslant \varepsilon$. Note that, due to its maximality, $E_n(\varepsilon)$ is always (n, ε)-spanning, and no proper subset of $E_n(\varepsilon)$ is (n, ε)-spanning. Each $y \in E_n(\varepsilon)$ must have a point of an $(n, \varepsilon/2)$-spanning set within an $\varepsilon/2$-ball (in d_n-metric) around it, and by the triangle inequality, this $\varepsilon/2$-ball is disjoint from $\varepsilon/2$-ball centred around all other points in $E_n(\varepsilon)$. Therefore, if $r_n(\varepsilon)$ denotes the minimal cardinality among all (n, ε)-spanning sets, then

$$r_n(\varepsilon) \leqslant s_n(\varepsilon) \leqslant r_n(\varepsilon/2). \tag{27}$$

Thus we can equally well define

$$h_{top}(T) = \lim_{\varepsilon \to 0} \limsup_{n \to \infty} \frac{1}{n} \log r_n(\varepsilon). \tag{28}$$

Examples: Consider the β-transformation $T_\beta : [0,) \to [0, 1)$, $x \mapsto \beta x \pmod 1$ for some $\beta > 1$. Take $\varepsilon < 1/(2\beta^2)$, and $G_n = \{\frac{k}{\beta^n} \ : \ 0 \leqslant k < \beta^n\}$. Then G_n is (n, ε)-separating, so $s_n(\varepsilon) \geqslant \beta^n$. On the other hand, $G'_n = \{\frac{2k\varepsilon}{\beta^n} \ : \ 0 \leqslant k < \beta^n/(2\varepsilon)\}$ is (n, ε)-spanning, so $r_n(\varepsilon) \leqslant \beta^n/(2\varepsilon)$. Therefore

$$\log \beta = \limsup_n \frac{1}{n} \log \beta^n \leqslant h_{top}(T_\beta) \leqslant \limsup_n \log \beta^n/(2\varepsilon) = \log \beta.$$

Circle rotations, or in general isometries, T have zero topological entropy. Indeed, if $E(\varepsilon)$ is an ε-separated set (or ε-spanning set), it will also be (n, ε)-separated (or

(n,ε)-spanning) for every $n \geqslant 1$. Hence $s_n(\varepsilon)$ and $r_n(\varepsilon)$ are bounded in n, and their exponential growth rates are equal to zero.

Finally, let (X,σ) be the full shifts on N symbols. Let $\varepsilon > 0$ be arbitrary, and take m such that $2^{-m} < \varepsilon$. If we select a point from each $n+m$-cylinder, this gives an (n,ε)-spanning set, whereas selecting a point from each n-cylinder gives an (n,ε)-separated set. Therefore

$$\begin{aligned}\log N = \limsup_n \frac{1}{n} \log N^n &\leqslant \limsup_n \frac{1}{n} \log s_n(\varepsilon) \leqslant h_{top}(T_\beta) \\ &\leqslant \limsup_n \frac{1}{n} \log r_n(\varepsilon) \leqslant \limsup_n \log N^{n+m} = \log N.\end{aligned}$$

Proposition 12. *For a continuous map T on a compact metric space (X,d), the three definitions* (24)*,* (26) *and* (28) *give the same outcome.*

Proof. The equality of the limits (26) and (28) follows directly from (27).

If $\mathcal{U}$ is an ε-cover, every $A \in \mathcal{U}^n$ can contain at most one point in an (n,ε)-separated set, so $s(n,\varepsilon) < \mathcal{N}(\mathcal{U}^n)$, whence $\limsup_n \frac{1}{n} \log s(n,\varepsilon) \leqslant \lim_n \frac{1}{n} \log \mathcal{N}(\mathcal{U}^n)$.

Finally, in a compact metric space, every open cover $\mathcal{U}$ has a number (called its *Lebesgue number*) such that for every $x \in X$, there is $U \in \mathcal{U}$ such that $B_\delta(x) \subset U$. Clearly $\delta < \varepsilon$ if $\mathcal{U}$ is an ε-cover. Now if an open ε-cover $\mathcal{U}$ has Lebesgue number δ, and E is an (n,δ)-spanning set of cardinality $\#E = r(n,\delta)$, then $X \subset \cup_{x\in E} \cap_{i=0}^{n-1} T^{-i}(B_\delta(T^i x))$. Since each $B_\delta(T^i(x))$ is contained in some $U \in \mathcal{U}$, we have $\mathcal{N}(\mathcal{U}^n) \leqslant r(n,\delta)$. Since $\delta \to 0$ as $\varepsilon \to 0$, also

$$\lim_{\varepsilon\to 0} \lim_n \frac{1}{n} \log \mathcal{N}(\mathcal{U}^n) \leqslant \lim_{\delta\to 0} \limsup_n \frac{1}{n} \log r(n,\delta).$$

□

15.4 Topological pressure

The topological pressure $P_{top}(T,\psi)$ combines entropy with a potential function $\psi : X \to \mathbb{R}$. By definition, $h_{top}(T) = P_{top}(T,\psi)$ if $\psi(x) \equiv 0$. Denote the n-th ergodic sum of ψ by

$$S_n\psi(x) = \sum_{k=0}^{n-1} \psi \circ T^k(x).$$

Next set

$$\begin{cases} K_n(T,\psi,\varepsilon) = \sup\{\sum_{x\in E} e^{S_n\psi(x)} \ : \ E \text{ is } (n,\varepsilon)\text{-separated}\}, \\ L_n(T,\psi,\varepsilon) = \inf\{\sum_{x\in E} e^{S_n\psi(x)} \ : \ E \text{ is } (n,\varepsilon)\text{-spanning}\}. \end{cases} \tag{29}$$

For reasonable choices of potentials, the quantities $\lim_{\varepsilon \to 0} \limsup_{n\to\infty} \frac{1}{n} \log K_n(T, \psi, \varepsilon)$ and $\lim_{\varepsilon \to 0} \limsup_{n\to\infty} \frac{1}{n} \log L_n(T, \psi, \varepsilon)$ are the same, and this quantity is called the **topological pressure**. To give an example of an unreasonable potential, take X_0 be a dense T-invariant subset of X such that $X \setminus X_0$ is also dense. Let

$$\psi(x) = \begin{cases} 100 & \text{if } x \in X_0, \\ 0 & \text{if } x \notin X_0. \end{cases}$$

Then $L_n(T, \psi, \varepsilon) = r_n(\varepsilon)$ whilst $K_n(T, \psi, \varepsilon) = e^{100n} s_n(\varepsilon)$, and their exponential growth rates differ by a factor 100. Hence, some amount of continuity of ψ is necessary to make it work.

Lemma 11. *If $\varepsilon > 0$ is such that $d(x, y) < \varepsilon$ implies that $|\psi(x) - \psi(y)| < \delta/2$, then*

$$e^{-n\delta} K_n(T, \psi, \varepsilon) \leqslant L_n(T, \psi, \varepsilon/2) \leqslant K_n(T, \psi, \varepsilon/2).$$

Exercise 2. *Prove Lemma 11. In fact, the second inequality holds regardless of what ψ is.*

Theorem 18. *If $T : X \to X$ and $\psi : X \to \mathbb{R}$ are continuous on a compact metric space, then the topological pressure is well-defined by*

$$P_{top}(T, \psi) := \lim_{\varepsilon \to 0} \limsup_{n\to\infty} \frac{1}{n} \log K_n(T, \psi, \varepsilon) = \lim_{\varepsilon \to 0} \limsup_{n\to\infty} \frac{1}{n} \log L_n(T, \psi, \varepsilon).$$

Exercise 3. *Show that $P_{top}(T^R, S_R\psi) = R \cdot P_{top}(T, \psi)$.*

16 Measure-theoretic entropy

Entropy is a measure for the complexity of a dynamical system (X, T). In the previous sections, we related this (or rather topological entropy) to the exponential growth rate of the cardinality of $\mathcal{P}_n = \bigvee_{k=0}^{n-1} T^{-k}\mathcal{P}$ for some partition of the space X. In this section, we look at the measure theoretic entropy $h_\mu(T)$ of an T-invariant measure μ, and this amounts to, instead of just counting $\mathcal{P}_n$, taking a particular weighted sum of the elements $Z_n \in \mathcal{P}_n$. However, if the mass of μ is equally distributed over the all the $Z_n \in \mathcal{P}_n$, then the outcome of this sum is largest; then μ would be the measure of maximal entropy. In "good" systems (X, T) is indeed the supremum over the measure theoretic entropies of all the T-invariant probability measures. This is called the **Variational Principle**:

$$h_{top}(T) = \sup\{h_\mu(T) \ : \ \mu \text{ is } T\text{-invariant probability measure}\}. \tag{30}$$

In this section, rather than presenting more philosophy what entropy should signify, let us first give the mathematical definition.

Define

$$\varphi : [0,1] \to \mathbb{R} \quad \varphi(x) = -x \log x$$

with $\varphi(0) := \lim_{x \downarrow 0} \varphi(x) = 0$. Clearly $\varphi'(x) = -(1+\log x)$ so $\varphi(x)$ assume its maximum at $1/e$ and $\varphi(1/e) = 1/e$. Also $\varphi''(x) = -1/x < 0$, so that φ is **strictly concave**:

$$\alpha\varphi(x) + \beta\varphi(y) \leqslant \varphi(\alpha x + \beta y) \qquad \text{for all } \alpha + \beta = 1, \alpha, \beta \geqslant 0, \tag{31}$$

with equality if and only if $x = y$.

Theorem 19 (Jensen's Inequality)**.** *For every strictly concave function $f : [0,\infty) \to \mathbb{R}$, and all $\alpha_i > 0$, $\sum_{i=1}^n \alpha_i = 1$ and $x_i \in [0,\infty)$ we have*

$$\sum_{i=1}^n \alpha_i f(x_i) \leqslant f(\sum_{i=1}^n \alpha_i x_i), \tag{32}$$

with equality if and only if all the x_i are the same.

Proof. We prove this by induction on n. For $n = 2$ it is simply (31). So assume that (32) holds for some n, and we treat the case $n+1$. Assume $\alpha_i > 0$ and $\sum_{i=1}^{n+1} \alpha_i = 1$ and write $B = \sum_{i=1}^n \alpha_i$.

$$\begin{aligned} f(\sum_{i=1}^{n+1} \alpha_i x_i) &= f(B \sum_{i=1}^n \frac{\alpha_i}{B} x_i + \alpha_{n+1} x_{n+1}) \\ &\geqslant Bf(\sum_{i=1}^n \frac{\alpha_i}{B} x_i) + \alpha_{n+1} f(x_{n+1}) \qquad \text{by (31)} \\ &\geqslant B \sum_{i=1}^n \frac{\alpha_i}{B} f(x_i) + \alpha_{n+1} f(x_{n+1}) \qquad \text{by (32) for } n \\ &= \sum_{i=1}^{n+1} \alpha_i f(x_i) \end{aligned}$$

as required. Equality also carries over by induction, because if x_i are all equal for $1 \leqslant i \leqslant n$, (31) only preserves equality if $x_{n+1} = \sum_{i=1}^n \frac{\alpha_i}{B} x_i = x_1$. □

Applying Jensen's inequality to $\varphi(x) = -x \log x$, we obtain:

Corollary 3. *For $p_1 + \cdots + p_n = 1$, $p_i > 0$, then $\sum_{i=1}^n \varphi(p_i) \leqslant \log n$ with equality if and only if all p_i are equal,* i.e., $p_i \equiv \frac{1}{n}$.

Proof. Take $\alpha_i = \frac{1}{n}$, then by Theorem 19,

$$\frac{1}{n} \sum_{i=1}^n \varphi(p_i) = \sum_{i=1}^n \alpha_i \varphi(p_i) \leqslant \varphi(\sum_{i=1}^n \frac{1}{n} p_i) = \varphi(\frac{1}{n}) = \frac{1}{n} \log n.$$

Now multiply by n. □

Corollary 4. *For real numbers a_i and $p_1 + \cdots + p_n = 1$, $p_i > 0$, $\sum_{i=1}^n p_i(a_i - \log p_i) \leqslant \log \sum_{i=1}^n e^{a_i}$ with equality if and only if $p_i = e^{a_i} / \sum_{i=1}^n e^{a_i}$ for each i.*

Proof. Write $\mathcal{Z} = \sum_{i=1}^n e^{a_i}$. Put $\alpha_i = e^{a_i}/\mathcal{Z}$ (so $\sum_{i=1}^n \alpha_i = 1$) and $x_i = p_i\mathcal{Z}/e^{a_i}$ (so $\sum_{i=1}^n \alpha_i x_i = 1$) in Theorem 19. Then

$$\begin{aligned}
\sum_{i=1}^n p_i(a_i - \log \mathcal{Z} - \log p_i) &= -\sum_{i=1}^n \frac{e^{a_i}}{\mathcal{Z}} \left(\frac{p_i\mathcal{Z}}{e^{a_i}} \log \frac{p_i\mathcal{Z}}{e^{a_i}} \right) \\
&\leqslant -\sum_{i=1}^n \frac{e^{a_i}}{\mathcal{Z}} \frac{p_i\mathcal{Z}}{e^{a_i}} \log \sum_{i=1}^n \frac{e^{a_i}}{\mathcal{Z}} \frac{p_i\mathcal{Z}}{e^{a_i}} = \varphi(1) = 0.
\end{aligned}$$

Rearranging gives $\sum_{i=1}^n p_i(a_i - \log p_i) \leqslant \log \mathcal{Z}$, with equality only if $x_i = p_i\mathcal{Z}/e^{a_i}$ are all the same. But as $\sum_{i=1}^n \alpha_i x_i = 1$ and also $\sum_{i=1}^n \alpha_i = 1$, this means that $x_i = 1$, *i.e.*, $p_i = e^{a_i}/\mathcal{Z}$. □

Exercise 4. *Reprove Corollaries 3 and 4 using Lagrange multipliers.*

Given a finite partition $\mathcal{P}$ of a probability space (X, μ), let

$$H_\mu(\mathcal{P}) = \sum_{P \in \mathcal{P}} \varphi(\mu(P)) = -\sum_{P \in \mathcal{P}} \mu(P) \log(\mu(P)), \tag{33}$$

where we can ignore the partition elements with $\mu(P) = 0$ because $\varphi(0) = 0$. For a T-invariant probability measure μ on $(X, \mathcal{B}, T)$, and a partition $\mathcal{P}$, define the **entropy of μ w.r.t. $\mathcal{P}$** as

$$h_\mu(T, \mathcal{P}) = \lim_{n \to \infty} \frac{1}{n} H_\mu(\bigvee_{k=0}^{n-1} T^{-k}\mathcal{P}). \tag{34}$$

Finally, the **measure theoretic entropy** of μ is

$$h_\mu(T) = \sup\{h_\mu(T, \mathcal{P}) \ : \ \mathcal{P} \text{ is a finite partition of } X\}. \tag{35}$$

Naturally, this raises the questions:

Does the limit exist in (34)?
How can one possibly consider **all** partitions of X?

We come to this later; first we want to argue that entropy is a characteristic of a measure preserving system. That is, two measure preserving systems $(X, \mathcal{B}, T, \mu)$ and $(Y, \mathcal{C}, S, \nu)$ that are **isomorphic**, *i.e.*, there are full-measured sets $X_0 \subset X$, $Y_0 \subset Y$ and a bi-measurable invertible measure-preserving map $\pi : X_0 \to Y_0$ (called **isomorphism**) such that the diagram

$$\begin{array}{ccc}
(X_0, \mathcal{B}, \mu) & \xrightarrow{T} & (X_0, \mathcal{B}, \mu) \\
\pi \downarrow & & \downarrow \pi \\
(Y_0, \mathcal{C}, \nu) & \xrightarrow{S} & (Y_0, \mathcal{C}, \nu)
\end{array}$$

commutes, then $h_\mu(T) = h_\nu(S)$. This holds, because the bi-measurable measure-preserving map π preserves all the quantities involved in (33)-(35), including the class of partitions for both systems.

A major class of systems where this is very important are the Bernoulli shifts. These are the standard probability space to measure a sequence of i.i.d. events each with outcomes in $\{0, \dots, N-1\}$ with probabilities $p_0, \dots, p_{N-1}$ respectively. That is: $X = \{0, \dots, N-1\}^{\mathbb{N}_0}$ or $\{0, \dots, N-1\}^{\mathbb{Z}}$, σ is the left-shift, and μ the Bernoulli measure that assigns to every cylinder set $[x_m \dots x_n]$ the mass

$$\mu([x_m \dots x_n]) = \prod_{k=m}^{n} \rho(x_k) \qquad \text{where } \rho(x_k) = p_i \text{ if } x_k = i.$$

For such a Bernoulli shift, the entropy is

$$h_\mu(\sigma) = -\sum_i p_i \log p_i, \tag{36}$$

so two Bernoulli shifts (X, p, μ_p) and $(X', p', \mu_{p'})$ can only be isomorphic if $-\sum_i p_i \log p_i = -\sum_i p'_i \log(p'_i)$. The famous theorem of Ornstein showed that entropy is a complete invariant for Bernoulli shifts:

Theorem 20 (Ornstein 1974 [21], cf. page 105 of [24])**.** *Two two-sided Bernoulli shifts (X, p, μ_p) and $(X', p', \mu_{p'})$ are isomorphic if and only if $-\sum_i p_i \log p_i = -\sum_i p'_i \log p'_i$.*

The isomorphism between these Bernoulli shifts is usually extremely complicated. A more (although still complicated) way of constructing these isomorphisms was given by Keane & Smorodinski in 1979, see [13].

Exercise 5. *Conclude that the Bernoulli shift $\mu_{(\frac14,\frac14,\frac14,\frac14)}$ is isomorphic to $\mu_{(\frac18,\frac18,\frac18,\frac18,\frac12)}$, but that no Bernoulli measure on four symbols can be isomorhic to $\mu_{(\frac15,\frac15,\frac15,\frac15,\frac15)}$*

For one-sided Bernoulli shifts, Ornstein's theorem does not hold. If the number of symbols are different, then the one-sided Bernoulli shifts can definitely not be isomorphic.

Let us go back to the definition of entropy, and try to answer the outstanding questions.

Definition 16. *We call a real sequence $(a_n)_{n \geqslant 1}$* **subadditive** *if*

$$a_{m+n} \leqslant a_m + a_n \qquad \textit{for all } m, n \in \mathbb{N}.$$

Theorem 21. *If $(a_n)_{n \geqslant 1}$ is subadditive, then $\lim_n \frac{a_n}{n} = \inf_{r \geqslant 1} \frac{a_r}{r}$.*

Proof. Every integer n can be written uniquely as $n = i \cdot r + j$ for $0 \leqslant j < r$. Therefore

$$\limsup_{n \to \infty} \frac{a_n}{n} = \limsup_{i \to \infty} \frac{a_{i \cdot r + j}}{i \cdot r + j} \leqslant \limsup_{i \to \infty} \frac{i a_r + a_j}{i \cdot r + j} = \frac{a_r}{r}.$$

This holds for all $r \in \mathbb{N}$, so we obtain

$$\inf_r \frac{a_r}{r} \leqslant \liminf_n \frac{a_n}{n} \leqslant \limsup_n \frac{a_n}{n} \leqslant \inf_r \frac{a_r}{r},$$

as required. □

16.1 Conditional Entropy

Definition 17. *Motivated by the conditional measure $\mu(P|Q) = \frac{\mu(P\cap Q)}{\mu(Q)}$, we define* **conditional entropy** *of a measure μ as*

$$H_\mu(\mathcal{P}|\mathcal{Q}) = -\sum_j \mu(Q_j) \sum_i \frac{\mu(P_i \cap Q_j)}{\mu(Q_j)} \log \frac{\mu(P_i \cap Q_j)}{\mu(Q_j)}, \tag{37}$$

where i runs over all elements $P_i \in \mathcal{P}$ and j runs over all elements $Q_j \in \mathcal{Q}$.

Before trying to interpret this notion, let us first list some properties that follow directly from the definition and Jensen's inequality:

Proposition 13. *Given measures μ, μ_i and two partitions $\mathcal{P}$ and $\mathcal{Q}$, we have*

1. $H_\mu(\mathcal{P} \vee \mathcal{Q}) \leqslant H_\mu(\mathcal{P}) + H_\mu(\mathcal{Q})$;
2. $H_\mu(\mathcal{Q}) = H_\mu(\mathcal{P}) + H_\mu(\mathcal{Q} \mid \mathcal{P})$, *whence* $h_\mu(T, \mathcal{Q}) = h_\mu(T, \mathcal{P}) + H_\mu(\mathcal{Q} \mid \mathcal{P})$.
3. $\sum_{i=1}^n p_i H_{\mu_i}(\mathcal{P}) \leqslant H_{\sum_{i=1}^n p_i \mu_i}(\mathcal{P})$ *whenever* $\sum_{i=1}^n p_1 = 1$, $p_i \geqslant 0$,

Proof. Direct computation gives

$$\begin{aligned}
H_\mu(\mathcal{P} \vee \mathcal{Q}) &= -\sum_{P\in\mathcal{P}} \sum_{Q\in\mathcal{Q}} \mu(P\cap Q) \log \mu(P \cap Q) \\
&= -\sum_{P\in\mathcal{P}} \sum_{Q\in\mathcal{Q}} \mu(P\cap Q) \log \frac{\mu(P\cap Q)}{\mu(P)} - \sum_{P\in\mathcal{P}} \sum_{Q\in\mathcal{Q}} \mu(P\cap Q) \log \mu(P) \\
&= H_\mu(\mathcal{Q} \mid \mathcal{P}) - \sum_{P\in\mathcal{P}} \mu(P) \log \mu(P) = H_\mu(\mathcal{Q} \mid \mathcal{P}) + H_\mu(\mathcal{P}),
\end{aligned}$$

and this proves the first part of 2. The second part of 2. then follows from the definition. Using Jensen's inequality, we get

$$\begin{aligned}
H_\mu(\mathcal{Q} \mid \mathcal{P}) &= -\sum_{P\in\mathcal{P}} \sum_{Q\in\mathcal{Q}} \mu(P) \frac{\mu(P\cap Q)}{\mu(P)} \log \frac{\mu(P\cap Q)}{\mu(P)} \\
&= \sum_{Q\in\mathcal{Q}} \sum_{P\in\mathcal{P}} \mu(P) \varphi(\frac{\mu(P\cap Q)}{\mu(P)}) \\
&\leq \sum_{Q\in\mathcal{Q}} \varphi(\sum_{P\in\mathcal{P}} \mu(P) \frac{\mu(P\cap Q)}{\mu(P)}) = \sum_{Q\in\mathcal{Q}} \varphi(\mu(Q)) = H_\mu(\mathcal{Q}).
\end{aligned}$$

Together with 2. we obtain $H_\mu(\mathcal{P} \vee \mathcal{Q}) \leqslant H_\mu(\mathcal{P}) + H_\mu(\mathcal{Q})$. $\square$

Subadditivity is the key to the convergence in (34). Call $a_n = H_\mu(\bigvee_{k=0}^{n-1} T^{-k}\mathcal{P})$. Then

$$\begin{aligned}
a_{m+n} &= H_\mu(\bigvee_{k=0}^{m+n-1} T^{-k}\mathcal{P}) && \text{use Proposition 13, part 1.} \\
&\leqslant H_\mu(\bigvee_{k=0}^{m-1} T^{-k}\mathcal{P}) + H_\mu(\bigvee_{k=m}^{m+n-1} T^{-k}\mathcal{P}) && \text{use } T\text{-invariance of } \mu \\
&= H_\mu(\bigvee_{k=0}^{m-1} T^{-k}\mathcal{P}) + H_\mu(\bigvee_{k=0}^{n-1} T^{-k}\mathcal{P}) \\
&= a_m + a_n.
\end{aligned}$$

Therefore $H_\mu(\bigvee_{k=0}^{n-1} T^{-k}\mathcal{P})$ is subadditive, and the existence of the limit of $\frac{1}{n}H_\mu(\bigvee_{k=0}^{n-1} T^{-k}\mathcal{P})$ follows.

Proposition 14. *Entropy has the following properties:*

1. *The identity map has entropy* 0*;*
2. $h_\mu(T^R) = R \cdot h_\mu(T)$ *and for invertible systems* $h_\mu(T^{-R}) = R \cdot h_\mu(T)$.

Proof. Statement 1. follows simply because $\bigvee_{k=0}^{n-1} T^{-k}\mathcal{P} = \mathcal{P}$ if T is the identity map, so the cardinality of $\bigvee_{k=0}^{n-1} T^{-k}\mathcal{P}$ doesn't increase with n.

For statement 2. set $\mathcal{Q} = \bigvee_{j=0}^{R-1} T^{-j}\mathcal{P}$. Then for $R \geqslant 1$,

$$\begin{aligned}
R \cdot h_\mu(T, \mathcal{P}) &= \lim_{n\to\infty} R \cdot \frac{1}{nR} H_\mu(\bigvee_{j=0}^{nR-1} T^{-j}\mathcal{P}) \\
&= \lim_{n\to\infty} \frac{1}{n} H_\mu(\bigvee_{j=0}^{n-1} (T^R)^{-j}\mathcal{Q}) \\
&= h_\mu(T^R, \mathcal{Q}).
\end{aligned}$$

Taking the supremum over all $\mathcal{P}$ or $\mathcal{Q}$ has the same effect. $\square$

To give some more intuition about condition entropies, observe that for some arbitrary partition $\mathcal{P}$ of X, the definition of condition entropy (37) gives

$$H(\mathcal{P}|\{\varnothing, X\}) = H(\mathcal{P}) \quad \text{and} \quad H(\mathcal{P}|\mathcal{P}) = 0.$$

In general, the finer the partition w.r.t. which we take conditional entropy, the smaller the entropy. That is, if $\mathcal{P}$ and $\mathcal{Q}, \mathcal{Q}'$ are partitions of X such that $\mathcal{Q}$ refines $\mathcal{Q}'$, then

$$H(\mathcal{Q}|\mathcal{P}) \geq H(\mathcal{Q}'|\mathcal{P}) \quad \text{but} \quad H(\mathcal{P}|\mathcal{Q}) \leq H(\mathcal{P}|\mathcal{Q}').$$

Corollary 5. *If $\mathcal{P}$ is a finite (or countable) partition, then* $\lim_{n\to\infty} H(\mathcal{P}|\bigvee_{j=1}^{n-1} T^{-j}\mathcal{P}) = h(\mathcal{P},T)$.

Proof. Using Proposition 13 part 2. and invariance of the measure repeatedly, we find

$$\begin{aligned}
H_\mu(\bigvee_{j=0}^{n-1} T^{-j}\mathcal{P}) &= H_\mu(\bigvee_{j=1}^{n-1} T^{-j}\mathcal{P}\vee\mathcal{P}) = H_\mu(\mathcal{P}|\bigvee_{j=1}^{n-1} T^{-j}\mathcal{P}) + H_\mu(\bigvee_{j=1}^{n-1} T^{-j}\mathcal{P}) \\
&= H_\mu(\mathcal{P}|\bigvee_{j=1}^{n-1} T^{-j}\mathcal{P}) + H_\mu(\bigvee_{j=0}^{n-2} T^{-j}\mathcal{P}) \\
&= H_\mu(\mathcal{P}|\bigvee_{j=1}^{n-1} T^{-j}\mathcal{P}) + H_\mu(\mathcal{P}|\bigvee_{j=1}^{n-2} T^{-j}\mathcal{P}) + H_\mu(\bigvee_{j=0}^{n-3} T^{-j}\mathcal{P}) \\
&\quad\vdots \qquad\qquad\qquad \vdots \\
&= \sum_{k=0}^{n-1} H_\mu(\mathcal{P}|\bigvee_{j=1}^{k} T^{-j}\mathcal{P}).
\end{aligned}$$

Since $\bigvee_{j=1}^{k}\mathcal{P}$ refines $\bigvee_{j=1}^{k'}\mathcal{P}$ if $k\geq k'$, the summands $H_\mu(\mathcal{P}|\bigvee_{j=1}^{k-1} T^{-j}\mathcal{P})$ are decreasing in k. Now divide by n and take $n\to\infty$:

$$h(\mathcal{P},T) = \lim_{n\to\infty}\frac{1}{n}H_\mu(\bigvee_{j=0}^{n-1} T^{-j}\mathcal{P}) = \lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1} H_\mu(\mathcal{P}|\bigvee_{j=1}^{k-1} T^{-j}\mathcal{P}) = \lim_{n\to\infty} H_\mu(\mathcal{P}|\bigvee_{j=1}^{n-1} T^{-j}\mathcal{P})$$

as required. □

Also, if $\mathcal{P}$ and $\mathcal{Q}$ are almost the same, in the sense that the measures of the symmetric difference $\mu(P\triangle Q)$ is small for all $P\in\mathcal{P}$, $Q\in\mathcal{Q}$, then $H(\mathcal{P}|\mathcal{Q})$ is small too. The following lemma quantifies this.

Lemma 12. *For every $\varepsilon>0$ there is $\delta>0$ such that if $\mathcal{P}=\{P_1,\dots,P_r\}$ and $\mathcal{Q}=\{Q_1,\dots,Q_r\}$ are two finite partitions with $\sum_{i=1}^{r}\mu(P_i\triangle Q_i)<\delta$, then $H(\mathcal{P}|\mathcal{Q})<\varepsilon/2$ and $H(\mathcal{Q}|\mathcal{P})<\varepsilon/2$ (so that $H(\mathcal{P}|\mathcal{Q})+H(\mathcal{Q}|\mathcal{P})<\varepsilon/2$).*

Proof. Let $\varepsilon>0$ be arbitrary and choose $\delta\in(0,\frac{1}{4})$ such that $-r(r-1)\delta\log\delta-(1-\delta)\log(1-\delta)<\frac{\varepsilon}{2}$. Let $\mathcal{A}=\{P_i\cap Q_j\}_{i\neq j}\cup(\cup_{i=1}^{r}P_i\cap Q_i)$. Then $\mathcal{P}\vee\mathcal{Q}=\mathcal{Q}\vee\mathcal{A}$ and $P_i\cap Q_j\subset\cup_{k=1}^{r}P_k\triangle Q_k$. Therefore (using the assumption of the lemma)

$$\mu(P_i\cap Q_j)<\delta \quad i\neq j \qquad\text{and}\qquad \mu(\cup_{i=1}^{r}P_i\cap Q_i)>1-\delta.$$

We compute $H(\mathcal{A})=-r(r-1)\delta\log\delta-(1-\delta)\log(1-\delta)<\varepsilon/2$. Finally, $H(\mathcal{Q})+H(\mathcal{P}|\mathcal{Q})=H(\mathcal{P}\vee\mathcal{Q})=H(\mathcal{Q}\vee\mathcal{A})\leq H(\mathcal{Q})+H(\mathcal{A})$ so that $H(\mathcal{P}|\mathcal{Q})\leq\varepsilon/2$. The symmetric statement $H(\mathcal{Q}|\mathcal{P})\leq\varepsilon/2$ follows likewise. □

16.2 Generators and the Kolmogorov-Sinaĭ Theorem

The next theorem is the key to really computing entropy, as it shows that a single well-chosen partition $\mathcal{P}$ suffices to compute the entropy as $h_\mu(T) = h_\mu(T, \mathcal{P})$.

Theorem 22 (Kolmogorov-Sinaĭ)**.** *Let $(X, \mathcal{B}, T, \mu)$ be a measure-preserving dynamical system. If partition $\mathcal{P}$ is such that*

$$\begin{cases} \bigvee_{j=0}^{\infty} T^{-k}\mathcal{P} \text{ generates } \mathcal{B} & \text{if } T \text{ is non-invertible,} \\ \bigvee_{j=-\infty}^{\infty} T^{-k}\mathcal{P} \text{ generates } \mathcal{B} & \text{if } T \text{ is invertible,} \end{cases}$$

then $h_\mu(T) = h_\mu(T, \mathcal{P})$.

We haven't explained properly what "generates $\mathcal{B}$ means, but the idea you should have in mind is that (up to measure 0), every two points in X should be in different elements of $\bigvee_{k=0}^{n-1} T^{-k}\mathcal{P}$ (if T is non-invertible), or of $\bigvee_{k=-n}^{n-1} T^{-k}\mathcal{P}$ (if T is invertible) for some sufficiently large n. The partition $\mathcal{B} = \{X\}$ fails miserably here, because $\bigvee_{j=-n}^{n} T^{-k}\mathcal{P} = \mathcal{P}$ for all n and no two points are ever separated in $\mathcal{P}$. A more subtle example can be created for the doubling map $T_2 : \mathbb{S}^1 \to \mathbb{S}^1$, $T_2(x) = 2x \pmod 1$. The partition $\mathcal{P} = \{[0, \frac{1}{2}), [\frac{1}{2}, 1)\}$. is separating every two points, because if $x \neq y$, say $2^{-(n+1)} < |x - y| \leqslant 2^{-n}$, then there is $k \leqslant n$ such that $T_2^k x$ and $T_2^k y$ belong to different partition elements.

On the other hand, $\mathcal{Q} = \{[\frac{1}{4}, \frac{3}{4}), [0, \frac{1}{4}) \cup [\frac{3}{4}, 1)\}$ does **not** separate points. Indeed, if $y = 1 - x$, then $T_2^k(y) = 1 - T_2^k(x)$ for all $k \geqslant 0$, so x and y belong to the same partition element, $T_2^k(y)$ and $T_2^k(x)$ will also belong to the same partition element!

In this case, $\mathcal{P}$ can be used to compute $h_\mu(T)$, while $\mathcal{Q}$ in principle cannot (although here, for all Bernoulli measure $\mu = \mu_{p,1-p}$, we have $h_\mu(T_2) = h_\mu(T, \mathcal{P}) = h_\mu(T, \mathcal{Q})$).

The existence of finite generating partition is guaranteed by a theorem due to Krieger [15].

Theorem 23. *Let $(X, \mathcal{B}, \mu)$ be a Lebesgue space (*i.e., *it is isomorphic to* $([0, 1], Leb) \sqcup$ *countable set). If T is an invertible measure-preserving transformation, then there is a finite generator $\mathcal{A} = \{A_1, \ldots, A_n\}$ and $e^{h_\mu(T)} \leq n \leq e^{h_\mu(T)} + 1$.*

We will not prove Krieger's Theorem here, but we will prove Theorem 22.

Proof of Theorem 22. Let $\mathcal{A}$ be the generating partition. Then $h_\mu(T, \mathcal{A}) \leqslant h_\mu(T)$ because the right hand side is the supremum over all partitions. For the other inequality, take an arbitrary finite partition $\mathcal{P}$. By Proposition 13, part 2, we have

$$h_\mu(T, \mathcal{P}) \leq h_\mu(T, \bigvee_{i=-k}^{k-1} T^{-i}\mathcal{A}) + H(\mathcal{P} \mid \bigvee_{i=-k}^{k-1} T^{-i}\mathcal{A}).$$

Since T is invertible and preserves μ we have

$$\begin{aligned}
h_\mu(T, \bigvee_{i=-k}^{k-1} T^{-i}\mathcal{A}) &= h_\mu(T, \bigvee_{i=0}^{2k-1} T^{-i}\mathcal{A}) \\
&= \lim_{n\to\infty} \frac{1}{n} H_\mu(\bigvee_{j=0}^{n-1} T^{-j}(\bigvee_{i=-k}^{k-1} T^{-i}\mathcal{A})) \\
&= \lim_{n\to\infty} \frac{1}{n} H_\mu(\bigvee_{j=0}^{n+2k-1} T^{-j}\mathcal{A}) \\
&= \lim_{n\to\infty} \frac{n+2k}{n} \lim_{n\to\infty} \frac{1}{n+2k} H_\mu(\bigvee_{j=0}^{n+2k-1} T^{-j}\mathcal{A}) = h_\mu(T, \mathcal{A}).
\end{aligned}$$

Since $\mathcal{P} = \{P_1, \dots, P_r\}$ is finite and $\mathcal{A}$ is generating, for any $\varepsilon >$, we can choose k sufficiently large and a finite partition $\{A_1, \dots, A_r\} \subset \bigvee_{i=-k}^{k-1} T^{-i}\mathcal{A}$ such that Lemma 12 applies. This togther gives

$$h_\mu(T, \mathcal{P}) \le h_\mu(T, \mathcal{A}) + \varepsilon/2.$$

Since ε was arbitrary, we have the required inequality $h_\mu(T, \mathcal{P}) \le h_\mu(T, \mathcal{A})$ and the theorem follows. $\square$

We finish this section with computing the entropy for a Bernoulli shift on two symbols, *i.e.,* we will prove (36) for two-letter alphabets and any probability $\mu([0]) =: p \in [0, 1]$. The space is thus $X = \{0, 1\}^{\mathbb{N}_0}$ and each $x \in X$ represents an infinite sequence of coin-flips with an unfair coin that gives head probability p (if head has the symbol 0). Recall from probability theory

$$\mathbb{P}(k \text{ heads in } n \text{ flips}) = \binom{n}{k} p^k (1-p)^{n-k},$$

so by full probability:

$$\sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} = 1.$$

Here $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ are the binomial coefficients, and we can compute

$$\begin{cases} k\binom{n}{k} = \frac{n!}{(k-1)!(n-k)!} = n\frac{(n-1)!}{(k-1)!(n-k)!} = n\binom{n-1}{k-1} \\ (n-k)\binom{n}{k} = \frac{n!}{(k)!(n-k-1)!} = n\frac{(n-1)!}{k!(n-k-1)!} = n\binom{n-1}{k} \end{cases} \tag{38}$$

This gives all the ingredients necessary for the computation.

$$
\begin{aligned}
H_\mu(\bigvee_{k=0}^{n-1} \sigma^{-k}\mathcal{P}) &= - \sum_{x_0,\dots,x_{n-1}=0}^{1} \mu([x_0,\dots,x_{n-1}]) \log \mu([x_0,\dots,x_{n-1}]) \\
&= - \sum_{x_0,\dots,x_{n-1}=0}^{1} \prod_{j=0}^{n-1} \rho(x_j) \log \prod_{j=0}^{n-1} \rho(x_j) \\
&= - \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} \log \left(p^k (1-p)^{n-k} \right) \\
&= - \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} k \log p \\
&\qquad - \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} (n-k) \log(1-p)
\end{aligned}
$$

In the first sum, the term $k=0$ gives zero, as does the term $k=n$ for the second sum. Thus we leave out these terms and rearrange by (38):

$$
\begin{aligned}
&= -p \log p \sum_{k=1}^{n} k \binom{n-1}{k} p^{k-1} (1-p)^{n-k} \\
&\qquad -(1-p)\log(1-p) \sum_{k=0}^{n-1} (n-k) \binom{n}{k} p^k (1-p)^{n-k-1} \\
&= -p \log p \sum_{k=1}^{n} n \binom{n-1}{k-1} p^{k-1} (1-p)^{n-k} \\
&\qquad -(1-p)\log(1-p) \sum_{k=0}^{n-1} n \binom{n-1}{k} p^k (1-p)^{n-k-1} \\
&= n \left(-p \log p - (1-p) \log(1-p) \right).
\end{aligned}
$$

The partition $\mathcal{P} = \{[0],[1]\}$ is generating, so by Theorem 22,

$$
h_\mu(\sigma) = h_\mu(\sigma, \mathcal{P}) = \lim_n \frac{1}{n} H_\mu(\bigvee_{k=0}^{n-1} \sigma^{-k}\mathcal{P}) = -p\log p - (1-p)\log(1-p)
$$

as required.

17 The Variational Principle

The Variational Principle claims that topological entropy (or pressure) is achieved by taking the supremum of the measure-theoretic entropies over all invariant probability measures. But in the course of these notes, topological entropy has seen various definitions. Even $\sup\{h_\mu(T) : \mu$ is a T-invariant probability measure$\}$ is sometimes used as definition of topological entropy. So it is time to be more definite.

We will do this by immediately passing to topological pressure, which we will base on the definition in terms of (n,δ)-spanning sets and/or (n,ε)-separated sets. Topological entropy then simply emerges as $h_{top}(T) = P_{top}(T, 0)$.

Theorem 24 (The Variational Principle)**.** *Let (X, d) be a compact metric space, $T : X \to X$ a continuous map and $\psi : X \to \mathbb{R}$ as continuous potential. Then*

$$P_{top}(T,\psi) = \sup\{h_\mu(T) + \int_X \psi \, d\mu \ : \ \mu \text{ is a } T\text{-invariant probability measure}\}. \quad (39)$$

Remark 7. *By the ergodic decomposition, every T-invariant probability measure can be written as convex combination (sometimes in the form of an integral) of ergodic T-invariant probability measures. Therefore, it suffices to take the supremum over all ergodic T-invariant probability measures.*

Proof. First we show that for every T-invariant probability measure, $h_\mu(T) + \int_X \psi \, d\mu \leqslant P_{top}(T,\psi)$. Let $\mathcal{P} = \{P_0, \dots, P_{N-1}\}$ be an arbitrary partition with $N \geqslant 2$ (if $\mathcal{P} = \{X\}$, then $h_\mu(T,\mathcal{P}) = 0$ and there is not much to prove). Let $\eta > 0$ be arbitrary, and choose $\varepsilon > 0$ so that $\varepsilon N \log N < \eta$.

By "regularity of μ", there are compact sets $Q_i \subset P_i$ such that $\mu(P_i \setminus Q_i) < \varepsilon$ for each $0 \leqslant i < N$. Take $Q_N = X \setminus \cup_{i=0}^{N-1} Q_i$. Then $\mathcal{Q} = \{Q_0, \dots, Q_N\}$ is a new partition of X, with $\mu(Q_N) \leqslant N\varepsilon$. Furthermore

$$\frac{\mu(P_i \cap Q_j)}{\mu(Q_j)} = \begin{cases} 0 & \text{if } i \neq j < N, \\ 1 & \text{if } i = j < N. \end{cases}$$

whereas $\sum_{i=0}^{N-1} \frac{\mu(P_i \cap Q_N)}{\mu(Q_N)} = 1$. Therefore the conditional entropy

$$\begin{aligned} H_\mu(\mathcal{P}|\mathcal{Q}) &= \sum_{j=0}^{N} \sum_{i=0}^{N-1} \mu(Q_j) \underbrace{\varphi\left(\frac{\mu(P_i \cap Q_j)}{\mu(Q_j)}\right)}_{=\ 0 \text{ if } j<N} \\ &= -\mu(Q_N) \sum_{i=0}^{N-1} \frac{\mu(P_i \cap Q_N)}{\mu(Q_N)} \log(\frac{\mu(P_i \cap Q_N)}{\mu(Q_N)}) \\ &\leqslant \mu(Q_N) \log N \qquad \text{by Corollary 3} \\ &\leqslant \varepsilon N \log N < \eta. \end{aligned}$$

Choose $0 < \delta < \frac{1}{2} \min_{0 \leqslant i < j < N} d(Q_i, Q_j)$ so that

$$d(x, y) < \delta \text{ implies } |\psi(x) - \psi(y)| < \varepsilon. \quad (40)$$

Here we use uniform continuity of ψ on the compact space X. Fix n and let $E_n(\delta)$ be an (n, δ)-spanning set. For $Z \in \mathcal{Q}_n := \bigvee_{k=0}^{n-1} T^{-k}\mathcal{Q}$, let $\alpha(Z) = \sup\{S_n\psi(x) \,:\, x \in Z\}$. For each such Z, also choose $x_Z \in \overline{Z}$ such that $S_n\psi(x_Z) = \alpha(Z)$ (again we use continuity of ψ here), and $y_Z \in E_n(\delta)$ such that $d_n(x_Z, y_Z) < \delta$. Hence

$$\alpha(Z) - n\varepsilon \leqslant S_n\psi(y_Z) \leqslant \alpha(Z) + n\varepsilon.$$

This gives

$$H_\mu(\mathcal{Q}_n) + \int_X S_n\psi \; d\mu \leqslant \sum_{Z\in\mathcal{Q}_n} \mu(Z)(\alpha(Z) - \log \mu(Z)) \leqslant \log \sum_{Z\in\mathcal{Q}_n} e^{\alpha(Z)} \tag{41}$$

by Corollary 4.

Each δ-ball intersects the closure of at most two elements of $\mathcal{Q}$. Hence, for each $y \in E_n(\delta)$, the cardinality $\#\{Z \in \mathcal{Q}_n : y_Z = y\} \leqslant 2^n$. Therefore

$$\sum_{Z\in\mathcal{Q}_n} e^{\alpha(Z)-n\varepsilon} \leqslant \sum_{Z\in\mathcal{Q}_n} e^{S_n\psi(y_Z)} \leqslant 2^n \sum_{y\in E_n(\delta)} e^{S_n\psi(y)}.$$

Take the logarithm and rearrange to

$$\log \sum_{Z\in\mathcal{Q}_n} e^{\alpha(Z)} \leqslant n(\varepsilon + \log 2) + \log \sum_{y\in E_n(\delta)} e^{S_n\varphi(y)}.$$

By T-invariance of μ we have $\int S_n\psi \; d\mu = n \int \psi \; d\mu$. Therefore

$$\begin{aligned} \frac{1}{n}H_\mu(\mathcal{Q}_n) + \int_X \psi \; d\mu &\leqslant \frac{1}{n}H_\mu(\mathcal{Q}_n) + \frac{1}{n}\int_X S_n\psi \; d\mu \\ &\leqslant \frac{1}{n}\log \sum_{Z\in\mathcal{Q}_n} e^{\alpha(Z)} \\ &\leqslant \varepsilon + \log 2 + \frac{1}{n}\log \sum_{y\in E_n(\delta)} e^{S_n\varphi(y)}. \end{aligned}$$

Taking the limit $n \to \infty$ and recalling that $E_n(\delta)$ is an arbitrary (n, δ)-spanning set, gives

$$H_\mu(T, \mathcal{Q}) + \int_X \psi \; d\mu \leqslant \varepsilon + \log 2 + P_{top}(T, \psi).$$

By Proposition 13, part 2., and recalling that $\varepsilon < \eta$, we get

$$h_\mu(T, \mathcal{P}) + \int_X \psi \; d\mu = H_\mu(T, \mathcal{Q}) + H_\mu(\mathcal{P}|\mathcal{Q}) + \int_X \psi \; d\mu \leqslant 2\eta + \log 2 + P_{top}(T, \psi).$$

We can apply the same reasoning to T^R and $S_R\psi$ instead of T and ψ. This gives

$$\begin{aligned} R \cdot \left(h_\mu(T, \mathcal{P}) + \int_X \psi \; d\mu \right) &= h_\mu(T^R, \mathcal{P}) + \int_X S_R\psi \; d\mu \\ &\leqslant 2\eta + \log 2 + P_{top}(T^R, S_R\psi) \\ &= 2\eta + \log 2 + R \cdot P_{top}(T, \psi). \end{aligned}$$

Divide by R and take $R \to \infty$ to find $h_\mu(T, \mathcal{P}) + \int_X \psi \; d\mu \leqslant P_{top}(T, \psi)$. Finally take the supremum over all partitions $\mathcal{P}$.

Now the other direction, we will work with (n, ε)-separated sets. After choosing $\varepsilon > 0$ arbitrary, we need to find a T-invariant probability measure μ such that

$$h_\mu(T) + \int_X \psi \; d\mu \geqslant \limsup_{n\to\infty} \frac{1}{n} \log K_n(T, \psi, \varepsilon) := P(T, \psi, \varepsilon).$$

Let $E_n(\varepsilon)$ be an (n, ε)-separated set such that

$$\log \sum_{y \in E_n(\varepsilon)} e^{S_n\psi(y)} \geqslant \log K_n(T, \psi, \varepsilon) - 1. \tag{42}$$

Define Δ_n as weighted sum of Dirac measures:

$$\Delta_n = \frac{1}{\mathcal{Z}} \sum_{y \in E_n(\varepsilon)} e^{S_n\psi(y)} \delta_y,$$

where $\mathcal{Z}_n = \sum_{y \in E_n(\varepsilon)} e^{S_n\psi(y)}$ is the normalising constant. Take a new probability measure

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

Therefore

$$\begin{aligned}
\int_X \psi \; d\mu_n &= \frac{1}{n} \sum_{k=0}^{n-1} \int_X \psi \; d(\Delta_n \circ T^{-k}) = \frac{1}{n} \sum_{k=0}^{n-1} \sum_{y \in E_n(\varepsilon)} \psi \circ T^k(y) \frac{1}{\mathcal{Z}} e^{S_n\psi(y)} \\
&= \frac{1}{n} \sum_{y \in E_n(\varepsilon)} S_n\psi(y) \frac{1}{\mathcal{Z}} e^{S_n\psi(y)} = \frac{1}{n} \int_X S_n\psi \; d\Delta_n.
\end{aligned} \tag{43}$$

Since the space of probability measures on X is compact in the weak* topology, we can find a sequence $(n_j)_{j\geqslant 1}$ such that for every continuous function $f : X \to \mathbb{R}$

$$\int_X f \; d\mu_{n_j} \to \int_X f \; d\mu \quad \text{ as } j \to \infty.$$

Choose a partition $\mathcal{P} = \{P_0, \dots, P_{N-1}\}$ with $\text{diam}(P_i) < \varepsilon$ and $\mu(\partial P_i) = 0$ for all $0 \leqslant i < N$. Since $Z \in \mathcal{P}_n := \bigvee_{k=0}^{n-1} T^{-k}\mathcal{P}$ contains at most one element of an (n, ε)-separated set, we have

$$\begin{aligned}
H_{\Delta_n}(\mathcal{P}_n) + \int_X S_n\psi \; d\Delta_n &= \sum_{y \in E_n(\varepsilon)} \Delta_n(\{y\}) \left(S_n\psi(y) - \log \Delta_n(\{y\})\right) \\
&= \log \sum_{y \in E_n(\varepsilon)} e^{S_n\psi(y)} = \log \mathcal{Z}_n.
\end{aligned}$$

by Corollary 4 (the part that gives equality).

Take $0 < q < n$ arbitrary, and for $0 \leqslant j < q$, let we split

$$\begin{aligned}
\bigvee_{k=0}^{n-1} T^{-k}\mathcal{P} &= \left(\bigvee_{r=0}^{a_j-1}\bigvee_{i=0}^{q-1} T^{-(rq+j+i)}\mathcal{P}\right) \vee \bigvee_{l\in V_j} T^{-l}\mathcal{P} \\
&= \left(\bigvee_{r=0}^{a_j-1} T^{-(rq+j)} \bigvee_{i=0}^{q-1} T^{-i}\mathcal{P}\right) \vee \bigvee_{l\in V_j} T^{-l}\mathcal{P},
\end{aligned}$$

where $V_j := \{0, 1, \ldots, j-1\} \cup \{a_j q + j, a_j q + j + 1, \ldots, n-1\}$ has at most $2q$ elements. Therefore, for j fixed, and using Proposition 13, part 2.,

$$\begin{aligned}
\log \mathcal{Z}_n &= H_{\Delta_n}(\mathcal{P}_n) + \int_X S_n\psi \; d\Delta_n \\
&\leqslant \sum_{r=0}^{a_j-1} H_{\Delta_n}(T^{-(rq+j)} \bigvee_{i=0}^{q-1} T^{-i}\mathcal{P}) + H_{\Delta_n}(\bigvee_{l\in V_j} T^{-l}\mathcal{P}) + \int_X S_n\psi \; d\Delta_n \\
&\leqslant \sum_{r=0}^{a_j-1} H_{\Delta_n\circ T^{-(rq+j)}}(\bigvee_{i=0}^{q-1} T^{-i}\mathcal{P}) + 2q\log N + \int_X S_n\psi \; d\Delta_n,
\end{aligned}$$

because $\bigvee_{l\in V_j} T^{-l}\mathcal{P}$ has at most N^{2q} elements and using Corollary 3. Summing the above inequality over $j = 0, \ldots, q-1$, gives

$$\begin{aligned}
q\log \mathcal{Z}_n &\leq \sum_{j=0}^{q-1}\sum_{r=0}^{a_j-1} H_{\Delta_n\circ T^{-rq+j}}(\bigvee_{i=0}^{q-1} T^{-i}\mathcal{P}) + 2q^2\log N + q\int_X S_n\psi \; d\Delta_n \\
&\leqslant n\sum_{k=0}^{n-1} \frac{1}{n} H_{\Delta_n\circ T^{-k}}(\bigvee_{i=0}^{q-1} T^{-i}\mathcal{P}) + 2q^2\log N + q\int_X S_n\psi \; d\Delta_n.
\end{aligned}$$

Proposition 13, part 3., allows us to swap the weighted average and the operation H:

$$q\log \mathcal{Z}_n \leqslant nH_{\mu_n}(\bigvee_{i=0}^{q-1} T^{-i}\mathcal{P}) + 2q^2\log N + q\int_X S_n\psi \; d\Delta_n.$$

Dividing by n and recalling (42) for the left hand side, and (43) to replace Δ_n by μ_n, we find

$$\frac{q}{n}\log K_n(T,\psi,\varepsilon) - \frac{q}{n} \leqslant H_{\mu_n}(\bigvee_{i=0}^{q-1} T^{-i}\mathcal{P}) + \frac{2q^2}{n}\log N + q\int_X \psi \; d\mu_n.$$

Because $\mu(\partial P_i) = 0$ for all i, we can replace n by n_j and take the weak limit as $j \to \infty$. This gives

$$qP(T,\psi,\varepsilon) \leqslant H_\mu(\bigvee_{i=0}^{q-1} T^{-i}\mathcal{P}) + q\int_X \psi \; d\mu.$$

Finally divide by q and let $q \to \infty$:

$$P(T, \psi, \varepsilon) \leqslant h_\mu(T) + \int_X \psi \, d\mu.$$

This concludes the proof. $\square$

18 Measures of maximal entropy

As we remarked before, the variational principle applied to $\psi = 0$ gives

$$h_{top}(f) = \sup\{h_\mu(f) : \mu \text{ is an ergodic } f\text{-invariant probability measure}\}.$$

If $h_\mu(f) = h_{top}(f)$, then we call μ a **measure of maximal entropy**, and if there is a unique measure of maximal entropy, then we call the system (X, f) **intrinsically ergodic**. Clearly, uniquely ergodic systems of finite entropy are intrinsically ergodic. In fact, intrinsic ergodicity is very common. Every β-transformation is intrinsically ergodic; more generally, every piecewise monotone piecewise continuous topologically transitive interval map is. We will not prove this in these notes, but in the rest of this section, we discuss the situationofor subshifts of finite type.

18.1 Subshifts of finite type

To each directed graph $(G, \to)$, say with vertices $\{1, \dots, N\}$ one can assign a *transition matrix* $A = (a_{i,j})_{i,j=1}^N$ where for each i, j, $A_{i,j}$ counts the number of edges from vertex i to vertex j. We call G *irreducible* if there exists a path (of some length) from each vertex to each vertex. It is called *aperiodic* if for each i, j there is $m \in \mathbb{N}$ such that there is a path from i to j of length n for every $n \geqslant m$. In terms of the transmatrix, this translates to: A is irreducible if for every i, j there is n such that $A^n_{i,j} > 0$, and A is aperiodic if in addition there is n such that $A^n_{i,j} > 0$ for all i, j.

The set of (bi)infinite strings

$$\Sigma_A = \{(x_i)_{i \in \mathbb{Z}} : x_i \in \{1, \dots, N\}, A_{x_i, x_{i+1}} > 0 \text{ for all } i \in \mathbb{Z}\}$$

is shift-invariant and closed in the standard product topology of $\{1, \dots, N\}^{\mathbb{Z}}$. Hence it is a *subshift*. It is called *subshift of finite type (SFT)* because of the finite collection of fobidden words (namely the pairs ij such that $A_{i,j} = 0$) that fully determines Σ_A.

It is easy to see that the *word-complexity*

$$\begin{aligned} p_n(\Sigma_A) &:= \#\{x_0 \dots x_{n-1} \text{ subword appearing in } \Sigma_A\} \\ &= \#\{\text{paths of length } n-1 \text{ in } G\} = \sum_{i,j} A^n_{i,j}. \end{aligned}$$

Because the partition into n-cylinders forms an open 2^{-n}-cover of Σ_A, we can derive

$$h_{top}(\sigma|_{\Sigma_A}) = \lim_{n\to\infty} \frac{1}{n} \log p_n(\Sigma_A) = \log \lambda,$$

where λ is the leading eigenvalue of the transition matrix A. That A has a unique, positive, leading eigenvalue follows from the following theorem.

Theorem 25 (Perron-Frobenius)**.** *Let A be a nonnegative $N \times N$-matrix such that $A^n > 0$ for some $m \in \mathbb{N}$. Then A has a unique (up to scaling) eigenvector with all entries > 0. The corresponding eigenvalue is positive, has multiplicity one, and is larger than the absolute value of every other eigenvalues of A.*

Proof. Let $C = \mathbb{R}^N_{\geqslant 0}$ be the one-sided cone of nonnegative vectors. Since A is nonnegative, $AC \subset C$, and because $A^m > 0$, $A^m C \subset C^o \cup \{0\}$, by which we mean that every nonzero vector in C is mapped into the interior of C by A^m. Define the simplex $S = \{x \in C : \|x\| = 1\}$ spanned by the unit vectors e_1, and let the map $f : S \to S$ be defined by $f(x) = Ax/\|Ax\|$. Since $A^n > 0$, it is impossible that $Ax = 0$ for $x \in S$, so f is well-defined. Although nonlinear, the map f is convex, meaning that it sends convex subsets of S to convex subsets, and extremal points to extremal points. Applying this to $\Pi_n := \cap_{k=0}^n f^k(S)$, we conclude that $(\Pi_n)_{n\geq 0}$ is a nested sequence of convex sets with $f^n(e_i)$, $i = 1, \dots N$ as extremal points. This carries over to the limit $\Pi := \bigcap_n \Pi_n$ as well; note that Π is contained in the interior of S because $A^n > 0$. We can select a subsequence (n_j) such that $f^{n_j}(e_i) \to p_i$ are the extremal points of Π. This is a finite set, invariant under f, so there is M such that each p_i is fixed by f^M and therefore an eigenvector of A^M associated to a positive eigenvalue. By reordering the p_i, we can assume that the corresponding eigenvalues of A^M are $\lambda_1 \geqslant \lambda_2 \geqslant \dots \geqslant \lambda_N$.

1. If $\lambda_2 = \lambda_1$ and $p_1 \neq p_2$, then we can find $v = \alpha_1 p_1 + \alpha_2 p_2 \in \partial C$. This is also an eigenvector of A^M, so $A^{kM} v \in \partial C$ for all k, but this contradicts that $A^m C \in C^o$.

2. If $\lambda_2 < \lambda_1$, then take $v = p_2 - \varepsilon p_1 \in C$ (for $\varepsilon > 0$ sufficiently small), and note that $A^{kM} v = \lambda_2^k p_2 - \varepsilon \lambda_1^k p_1$ cannot be contained in C for all k. This contradicts again the invariance of C. Hence, all p_i coincide, and it is the unique fixed point of f.

3. To show that λ_1 has multiplicity one, assume by contradiction that there is a generalised eigenvector $v \in S$ with $A^m v = \lambda_1^m v + p_1$. Then also $A^{kM} v = \lambda^k v + k\lambda_1^{(k-1)} p_1$. Take $w = p_1 - \varepsilon v \in C$ for some small $\varepsilon > 0$. Then $A^{kM} w = \lambda_1^{k-1}(\lambda_1 - \varepsilon k) p_1 - \varepsilon \lambda_1^k v$ which cannot be contained in C for large k. This again contradicts that $A^m C \subset C$ for all m.

4. Finally, suppose that μ is some eigenvalue, not necessarily associated with an eigenvector in S, such that $|\mu| > \lambda_1$. There is a A^M-invariant subspace V (possibly of dimension two if $\mu \in \mathbb{C} \setminus \mathbb{R}$) such that $A^M : V \to V$ is the composition of an

isometry and a dilatation by a factor $|\mu|$. In particular, there is a subsequence (k_j) such that $|\mu|^{-k_j} A^{k_j M} v \to v$ for every $v \in V$. Take $v \in V$ so that $w := v + p_1 \in \partial C$. If $|\mu| = \lambda_1$, then $|\mu|^{-k_j} A^{k_j M} w \to w$, contradicting that $A^m C \subset C^o$ for al m. If $|\mu| = \lambda_1$, then $|\mu|^{-k_j} A^{k_j M} w \to v$, again contradicting that $A^m C \subset C^o \cup \{0\}$. Hence all other eigenvectors of A^M are strictly smaller than λ_1.

The proof now follows by taking $\lambda = \lambda_1^{1/M}$. □

18.2 Parry measure

For the full shift (Σ, σ) with $\Sigma = \{0, \dots, N-1\}^{\mathbb{N}_0}$ or $\Sigma = \{0, \dots, N-1\}^{\mathbb{Z}}$, we have $h_{top}(\sigma) = \log N$, and the $(\frac{1}{N}, \dots, \frac{1}{N})$-Bernoulli measure μ indeed achieves this maximum: $h_\mu(\sigma) = h_{top}(\sigma)$. Hence μ is a (and in this case unique) **measure of maximal entropy**. The intuition to have here is that for a measure to achieve maximal entropy, it should distribute its mass as evenly over the space as possible. But how does this work for subshifts, where it is not immediately obvious how to distribute mass evenly?

For subshifts of finite type, Parry [22] demonstrated how to construct the measure of maximal entropy, which is now called after him. Let (Σ_A, σ) be a subshift of finite type on alphabet $\{0, \dots, N-1\}$ with transition matrix $A = (A_{i,j})_{i,j=0}^{N-1}$, so $x = (x_n) \in \Sigma_n$ if and only if $A_{x_n, x_{n+1}} = 1$ for all n. Let us assume that A is aperiodic and irreducible. Then by the Perron-Frobenius Theorem for matrices, there is a unique real eigenvalue, of multiplicity one, which is larger in absolute value than every other eigenvalue, and $h_{top}(\sigma) = \log \lambda$. Furthermore, by irreducibility of A, the left and right eigenvectors $u = (u_0, \dots, u_{N-1})$ and $v = (v_0, \dots, v_{N-1})^T$ associated to λ are unique up to a multiplicative factor, and they can be chosen to be strictly positive. We will scale them such that

$$\sum_{i=0}^{N-1} u_i v_i = 1.$$

Now define the **Parry measure** by

$$\begin{aligned} p_i &:= u_i v_i = \mu([i]), \\ p_{i,j} &:= \frac{A_{i,j} v_j}{\lambda v_i} = \mu([ij] \mid [i]), \end{aligned}$$

so $p_{i,j}$ indicates the conditional probability that $x_{n+1} = j$ knowing that $x_n = i$. Therefore $\mu([ij]) = \mu([i])\mu([ij] \mid [i]) = p_i p_{i,j}$. It is stationary (*i.e.,* shift-invariant) but not quite a product measure: $\mu([i_m \dots i_n]) = p_{i_m} \cdot p_{i_m, i_{m+1}} \cdots p_{i_{n-1}, i_n}$.

Theorem 26. *The Parry measure μ is the unique measure of maximal entropy for a subshift of finite type with aperiodic irreducible transition matrix.*

Proof. In this proof, we will only show that $h_\mu(\sigma) = h_{top}(\sigma) = \log \lambda$, and skip the (more complicated) uniqueness part.

The definitions of the masses of 1-cylinders and 2-cylinders are compatible, because (since v is a right eigenvector)

$$\sum_{j=0}^{N-1} \mu([ij]) = \sum_{j=0}^{N-1} p_i p_{i,j} = p_i \sum_{j=0}^{N-1} \frac{A_{i,j} v_j}{\lambda v_i} = p_i \frac{\lambda v_i}{\lambda v_i} = p_i = \mu([i]).$$

Summing over i, we get $\sum_{i=0}^{N-1} \mu([i]) = \sum_{i=0}^{N-1} u_i v_i = 1$, due to our scaling.

To show that μ is shift-invariant, we take any cylinder set $Z = [i_m \dots i_n]$ and compute

$$\begin{aligned}
\mu(\sigma^{-1}Z) &= \sum_{i=0}^{N-1} \mu([i i_m \dots i_n]) = \sum_{i=0}^{N-1} \frac{p_i p_{i,i_m}}{p_{i_m}} \mu([i_m \dots i_n]) \\
&= \mu([i_m \dots i_n]) \sum_{i=0}^{N-1} \frac{u_i v_i \ A_{i,i_m} v_{i_m}}{\lambda v_i \ u_{i_m} v_{i_m}} \\
&= \mu(Z) \sum_{i=0}^{N-1} \frac{u_i A_{i,i_m}}{\lambda u_{i_m}} = \mu(Z) \frac{\lambda u_{i_m}}{\lambda u_{i_m}} = \mu(Z).
\end{aligned}$$

This invariance carries over to all sets in the σ-algebra $\mathcal{B}$ generated by the cylinder sets.

Based on the interpretation of conditional probabilities, the identity

$$\sum_{\substack{i_{m+1},\dots,i_n=0 \\ A_{i_k,i_{k+1}}=1}}^{N-1} p_{i_m} p_{i_m,i_{m+1}} \cdots p_{i_{n-1},i_n} = p_{i_m} \text{ and } \sum_{\substack{i_m,\dots,i_{n-1}=0 \\ A_{i_k,i_{k+1}}=1}}^{N-1} p_{i_m} p_{i_m,i_{m+1}} \cdots p_{i_{n-1},i_n} = p_{i_n} \quad (44)$$

follows because the left hand side indicates the total probability of starting in state i_m and reach some state after $n-m$ steps, respectively start at some state and reach state n after $n-m$ steps.

To compute $h_\mu(\sigma)$, we will confine ourselves to the partition $\mathcal{P}$ of 1-cylinder sets; this partition is generating, so this restriction is justified by Theorem 22.

$$\begin{aligned}
H_\mu(\bigvee_{k=0}^{n-1} \sigma^{-k}\mathcal{P}) &= -\sum_{\substack{i_0,\dots,i_{n-1}=0 \\ A_{i_k,i_{k+1}}=1}}^{N-1} \mu([i_0 \dots i_{n-1}]) \log \mu([i_0 \dots i_{n-1}]) \\
&= -\sum_{\substack{i_0,\dots,i_{n-1}=0 \\ A_{i_k,i_{k+1}}=1}}^{N-1} p_{i_0} p_{i_0,i_1} \cdots p_{i_{n-1},i_n} \left(\log p_{i_0} + \log p_{i_0,i_1} + \dots + \log p_{i_{n-2},i_{n-1}}\right) \\
&= -\sum_{i_0=0}^{N-1} p_{i_0} \log p_{i_0} - (n-1) \sum_{i,j=0}^{N-1} p_i p_{i,j} \log p_{i,j},
\end{aligned}$$

by (44) used repeatedly. Hence

$$\begin{aligned} h_\mu(\sigma) &= \lim_{n\to\infty} \frac{1}{n} H_\mu(\bigvee_{k=0}^{n-1} \sigma^{-k}\mathcal{P}) \\ &= -\sum_{i,j=0}^{N-1} p_i p_{i,j} \log p_{i,j} \\ &= -\sum_{i,j=0}^{N-1} \frac{u_i A_{i,j} v_j}{\lambda} \left(\log A_{i,j} + \log v_j - \log v_i - \log \lambda\right). \end{aligned}$$

The first term in the brackets is zero because $A_{i,j} \in \{0, 1\}$. The second term (summing first over i) simplifies to

$$-\sum_{j=0}^{N-1} \frac{\lambda u_j v_j}{\lambda} \log v_j = -\sum_{j=0}^{N-1} u_j v_j \log v_j,$$

whereas the third term (summing first over j) simplifies to

$$\sum_{i=0}^{N-1} \frac{u_i \lambda v_i}{\lambda} \log v_i = \sum_{i=0}^{N-1} u_i v_i \log v_i.$$

Hence these two terms cancel each other. The remaining term is

$$\sum_{i,j=0}^{N-1} \frac{u_i A_{i,j} v_j}{\lambda} \log \lambda = \sum_{i=0}^{N-1} \frac{u_i \lambda v_i}{\lambda} \log \lambda = \sum_{i=0}^{N-1} u_i v_i \log \lambda = \log \lambda.$$

□

Remark 8. *There are systems without maximising measure, for example among the "shifts of finite type" on infinite alphabets. To give an example (without proof, see [8]), if $\mathbb{N}$ is the alphabet, and the infinite transition matrix $A = (A_{i,j})_{i,j\in\mathbb{N}}$ is given by*

$$A_{i,j} = \begin{cases} 1 & \text{if } j \geqslant i-1, \\ 0 & \text{if } j < i-1, \end{cases}$$

then $h_{top}(\sigma) = \log 4$, but there is no measure of maximal entropy.

Exercise 6. *Find the maximal measure for the Fibonacci subshift of finite type. What is the limit frequency of the symbol zero in μ-typical sequences x?*

19 The Shannon-McMillan-Breiman Theorem

Before starting with this theorem, we make a small digression into conditional expectations. Given a measure preserving system $(X, \mathcal{B}, \mu, T)$, some measurable function $f : X \to \mathbb{R}$ and some σ-algebra $\mathcal{C}$ (possibly $\mathcal{C} = \mathcal{B}$, possibly $\mathcal{C}$ coarser than $\mathcal{B}$), we can define the **conditional expectation** $\mathbb{E}_\mu(f|\mathcal{C})$ as the unique $\mathcal{C}$-measurable function $\bar{f}$ such that

$$\int_C \bar{f}\, d\mu = \int_C f\, d\mu$$

for all $C \in \mathcal{C}$. Recall that $\mathcal{C}$-measurable means that $\bar{f}^{-1}([t, \infty)) \in \mathcal{C}$ for all $t \in \mathbb{R}$, and therefore $\bar{f}$ must be constant on all atoms of $\mathcal{C}$, i.e., on all sets $A \in \mathcal{C}$ such that $A' \in \mathcal{C}$ can only be a subset of A if $\mu(A') = 0$ or $\mu(A \setminus A') = 0$. Note that conditional expectation is a function, and (unlike expectation or conditional probability) not a number. It is the function $\bar{f}$ such that for each atom A,

$$\bar{f}(x) = \frac{1}{\mu(A)} \int_A f\, d\mu \quad \text{for } \mu\text{-a.e. } x \in A. \tag{45}$$

The finer the σ-algebra $\mathcal{C}$, the more $\bar{f}$ looks like f. This is expressed in the following version of the

Theorem 27 (Martingale Convergence Theorem). *If $(\mathcal{C})_n$ is a sequence of σ-algebras such that $\mathcal{C}_{n+1}$ refines $\mathcal{C}_n$ and $\mathcal{C} = \lim_{n\to\infty} \mathcal{C}_n := \bigvee_{n=1}^{\infty} \mathcal{C}_n$, then*

$$\mathbb{E}_\mu(f|\mathcal{C}_n) \to \mathbb{E}_\mu(f|\mathcal{C}) \quad \mu\text{-a.e. as } n \to \infty.$$

Proof. For the proof of this, see e.g. [**?**] □

The main result of this section says that, given a partition $\mathcal{P}$, the n-cylinder of μ-a.e. point x scales as $\mu(\mathcal{P}_n(x)) \sim e^{-nh(\mathcal{P},T)}$.

Theorem 28 (Shannon-McMillan-Breiman). *Let $(X, \mathcal{B}, \mu, T)$ be a measure-preserving transformation and $\mathcal{P}$ a (countable or finite) partition with $H(\mathcal{P}) < \infty$ Let $\mathcal{P}_n = \bigvee_{j=0}^{n-1} T^{-j}(\mathcal{P})$ and $\mathcal{P}_n(x)$ the element of $\mathcal{P}_n$ containing x. Then*

$$-\lim_{n\to\infty} \frac{1}{n} \log \mu(\mathcal{P}_n(x)) = h(\mathcal{P}, T) \quad \mu\text{-a.e.}$$

Before we start with the proof, we recall the **information function**

$$I_{\mathcal{P}}(x) := -\log \mu(\mathcal{P}(x)) = -\sum_{P\in\mathcal{P}} 1_P(x) \log \mu(P),$$

with respect to which we have $H(\mathcal{P}) = \mathbb{E}(I_{\mathcal{P}})$. Inserting this in the definition of the entropy, we obtain

$$h(\mathcal{P}, T) = \lim_{n\to\infty} \frac{1}{n} H(\mathcal{P}_n) = \lim_{n\to\infty} \frac{1}{n} \int_X I_{\mathcal{P}_n}(x)\, d\mu. \tag{46}$$

The Shannon-McMillan-Breiman Theorem says that in fact the integrand converges to $h(\mathcal{P}, T)$ μ-a.e.

Similarly to conditional entropy, we define the **conditional information function**

$$I_{\mathcal{P}|\mathcal{Q}}(x) := -\sum_{P\in\mathcal{P}} \sum_{Q\in\mathcal{Q}} 1_{P\cap Q}(x) \log \frac{\mu(P\cap Q)}{\mu(Q)}.$$

Comparing this to the definition of conditional entropy, we get

$$\int_X I_{\mathcal{P}|\mathcal{Q}}\, d\mu = -\sum_{P\in\mathcal{P}} \sum_{Q\in\mathcal{Q}} \mu(P\cap Q) \log \frac{\mu(P\cap Q)}{\mu(Q)} = H_\mu(\mathcal{P}|\mathcal{Q}). \tag{47}$$

One can check (using Proposition 13 and the definition) that

$$I_{\mathcal{P}\vee\mathcal{Q}} = I_{\mathcal{P}} + I_{\mathcal{Q}|\mathcal{P}}. \tag{48}$$

Because of (45) and $1_P 1_Q = 1_{P\cap Q}$ we have

$$\begin{aligned}
-\log \mathbb{E}_\mu(1_{\mathcal{P}(x)}|\mathcal{Q}) &= -\log \mathbb{E}_\mu(\sum_{P\in\mathcal{P}} 1_P|\mathcal{Q}) = -\log \sum_{Q\in\mathcal{Q}} \frac{1}{\mu(Q)} \int_Q \sum_{P\in\mathcal{P}} 1_P\, d\mu \\
&= -\log \sum_{P\in\mathcal{P}} \sum_{Q\in\mathcal{Q}} 1_{P\cap Q} \frac{\mu(P\cap Q)}{\mu(Q)} \int_Q 1_P\, d\mu = I_{\mathcal{P}|\mathcal{Q}}(x).
\end{aligned}$$

Proof of Theorem 28. Write $g_k(x) = I_{\mathcal{P}|\vee_{j=1}^{k-1} T^{-j}\mathcal{P}}(x)$ for $k \geq 2$ and $g_1(x) = I_{\mathcal{P}}$. Then by (48)

$$\begin{aligned}
I_{\vee_{j=0}^{n-1} T^{-j}\mathcal{P}}(x) &= I_{\vee_{j=1}^{n-1} T^{-j}\mathcal{P}}(x) + I_{\mathcal{P}|\vee_{j=1}^{n-1} T^{-j}\mathcal{P}}(x) \\
&= I_{\vee_{j=0}^{n-2} T^{-j}\mathcal{P}}(Tx) + g_n(x) \\
&= I_{\vee_{j=1}^{n-2} T^{-j}\mathcal{P}}(Tx) + I_{\mathcal{P}|\vee_{j=1}^{n-2} T^{-j}\mathcal{P}}(Tx) + g_n(x) \\
&= I_{\vee_{j=0}^{n-3} T^{-j}\mathcal{P}}(T^2x) + g_{n-1}(Tx) + g_n(x) \\
&\quad \vdots \qquad \vdots \qquad \vdots \\
&= g_1(T^{n-1}(x)) + \cdots + g_{n-1}(T(x)) + g_n(x) = \sum_{j=0}^{n-1} g_{n-j}(T^j x).
\end{aligned}$$

Let $g = \lim_{n\to\infty} g_n$, which exists μ-a.e. and belongs to $L^1(\mu)$ because of the Martingale Convergence Theorem. We write the previous equality as

$$\frac{1}{n} I_{\vee_{j=0}^{n-1} T^{-j}\mathcal{P}}(x) = \frac{1}{n}\sum_{j=0}^{n-1} g(T^j x) + \frac{1}{n}\sum_{j=0}^{n-1} (g_{n-j} - g)(T^j x).$$

Since μ is ergodic, the first sum converges μ-a.e. to $\int_X g\, d\mu$, which is equal to $H_\mu(\mathcal{P} | \vee_{j=1}^{\infty} T^{-j}\mathcal{P})$ by (47), which in turn is equal to $h(\mathcal{P}, T)$ by Corollary 5.

For the second sum, we define

$$G_N = \sup_{k\geq n} |g_k - g| \quad \text{ and } \quad g^* = \sup_{n\geq 1} g_n.$$

Then $0 \leq G_N \leq g + g^*$ and $g + g^* \in L^1(\mu)$; this is because $\int g_n\, d\mu = H_\mu(\mathcal{P} | \bigvee_{j=1}^{n-1} \mathcal{P})$ is decreasing in n. Moreover, $G_N \to 0$ μ-a.e., so by the Dominated Convergence Theorem, $\lim_{N\to\infty} \int_X G_N\, d\mu = \int_X \lim_{N\to\infty} G_N\, d\mu = 0$. Now for any $N \geq 1$ and $n \geq N$ we split the second sum:

$$\begin{aligned} \frac{1}{n}\sum_{j=0}^{n-1}(g_{n-j} - g)(T^j x) &= \frac{1}{n}\sum_{j=0}^{n-N-1}(g_{n-j} - g)(T^j x) + \frac{1}{n}\sum_{j=n-N}^{n-1}(g_{n-j} - g)(T^j x) \\ &\leq \frac{1}{n}\sum_{k=0}^{n-N-1} G_N(T^k x) + \frac{1}{n}\sum_{j=n-N}^{n-1}(g_{n-j} - g)(T^j x). \end{aligned}$$

First take the limit $n \to \infty$. The the second sum tends to zero, and by the Ergodic Theorem, the first sum tends to $\int_X G_N\, d\mu$. Finally, taking $N \to \infty$ gives $\frac{1}{n}\sum_{j=0}^{n-1}(g_{n-j} - g)(T^j x) \to 0$, and hence $I_{\vee_{j=0}^{n-1} T^{-j}\mathcal{P}}(x) \to h(\mathcal{P}, T)$ μ-a.e., as required. □

20 Equilibrium states and Gibbs measures

20.1 Introductory example of the Ising model

This extended example is meant to give a feel for many of the ingredients in thermodynamic formalism. It is centred around a simplified Ising model, which can be computed completely.

We take the configuration space $\Omega = \{-1, +1\}^{\mathbb{Z}}$, that is the space of all bi-infinite sequences of $+1$'s and -1's. This give a rough model of ferro-magnetic atoms arranged on a line, having spin either upwards ($+1$) or downwards (-1). If all spins are upwards (or all downwards), then the material is fully magnetized, but usually the heat in the material means that atom rotate directing their spin in all directions over time, which we discretize to either up or down.

Of course, infinitely many atoms is unrealistic, and hence a configuration space $\{-1,+1\}^{[-n,n]}$ would be better (where $[-n,n]$ is our notation of the integer interval $\{-n,-n+1,\dots,n-1,n\}$), but for simplicity, let us look at the infinite line for the moment.

A **probability measure** μ indicates how likely it is to find a particular configuration, or rather a particular ensemble of configurations. For example, the fully magnetized states are expressed by the measures:

$$\delta_+(A) = \begin{cases} 1 & \text{if } A \ni (\dots,+1,+1,+1,+1,\dots) \\ 0 & \text{if } A \not\ni (\dots,+1,+1,+1,+1,\dots) \end{cases}$$

and δ_- with the analogous definition. For these two measures, only one configuration is likely to occur. Usually a single configuration occurs with probability zero, and we have to look at ensembles instead. Define **cylinder sets**

$$C_{m,n}(\omega) = \{\omega' \in \Omega : \omega'_i = \omega_i \text{ for } i \in [m,n]\}$$

as the set of all configurations that agree with configuration ω on sites i for $m \leqslant i \leqslant n$. Its **length** is $n-m+1$. Another notation would be $C_{m,n}(\omega) = [\omega_m\omega_{m+1}\dots\omega_n]$.

The Bernoulli measure (stationary product measure) μ_p is defined as[6]

$$\mu_p([\omega_m\omega_{m+1}\dots\omega_n]) = \prod_{i=m}^{n} p(\omega_i), \text{ where } p(+1) = p \text{ and } p(-1) = 1-p.$$

There is a Bernoulli measure μ_p for each $p \in [0,1]$ and $\mu_1 = \delta_+$, $\mu_0 = \delta_-$. However, for $p \in (0,1)$, every single configuration has measure 0. The Law of Large Numbers implies that the set of configurations in which the frequency of $+1$'s is anything else than p has zero measure.

Since physical problem is translation invariant. Define the **left-shift** as

$$\sigma(\omega)_i = \omega_{i+1}.$$

Translation invariance of a measure then means shift-invariance: $\mu(A) = \mu(\sigma(A))$ for each ensemble $A \subset \Omega$. Many probability measures on Ω are not translation invariant, but fortunately, the examples μ_p above are.

Another example of shift-invariant measures are the Gibbs measures, associated to some potential function $\psi : \Omega \to \mathbb{R}$; the integral $\int_\Omega \psi \, d\mu$ is called the (potential) **energy** of μ.

Definition 18. *A measure μ is a* **Gibbs measure** *w.r.t. potential function $\psi : \Omega \to \mathbb{R}$ if there are constants $C > 0$ and $P \in \mathbb{R}$ such that for all cylinder sets $C_{m,n}$ and all $\omega \in C_{m,n}$,*

$$\frac{1}{C} \leqslant \frac{\mu(C_{m,n})}{\exp \sum_{i=m}^{n}(\psi \circ \sigma^i(\omega) - P)} \leqslant C. \tag{49}$$

[6] This measure extends uniquely to all measurable sets by Kolmogorov's Extension Theorem.

The number P is called the **pressure***; in this setting it is a sort of normalizing constant, adjusting the exponential decrease of the denominator to the exponential decrease of the numerator*[7]

If we choose the potential to be

$$\psi(\omega) = \begin{cases} \log p & \text{if } \omega_0 = +1 \\ \log 1 - p & \text{if } \omega_0 = -1 \end{cases},$$

then the Bernoulli measure μ_p is actually a Gibbs measure, with pressure $P = 0$ and "distortion constant" $C = 1$. Indeed,

$$\mu(C_{m,n}(\omega)) = \prod_{i=m}^{n} p(\omega_i) = \prod_{i=m}^{n} e^{\psi(\sigma^i(\omega))} = \exp(\sum_{i=m}^{n} \psi(\sigma^i(\omega))),$$

and (49) follows.

The next ingredient is **entropy**. We postpone the precise definition, except for to say that there are different kinds. The system itself can have **topological entropy** $h_{top}(\sigma)$ which is independent of the measure, while each shift-invariant measure μ has its metric entropy or rather **measure theoretical entropy** $h_\mu(\sigma)$. For the Bernoulli measure μ_p, the measure theoretical entropy is

$$h_{\mu_p}(\sigma) = -(p \log p + (1-p)\log(1-p))$$

is the minus the expectation of ψ.

Exercise 7. *For $\varphi : [0,1] \to \mathbb{R}$ defined as $\varphi(x) = -(x \log x + (1-x)\log(1-x))$, we can write $h_{\mu_p}(\sigma) = \varphi(p)$. Compute the limits $\lim_{x\to 0}\varphi(x)$ and $\lim_{x\to 1}\varphi(x)$. Conclude that δ_+ and δ_- have zero entropy. (This agrees with the idea that entropy is suppose to measure disorder.) Where does φ assume its maximum? What does this suggest about the measure of maximal entropy?*

Exercise 8. *Compute its first and second derivative. Is φ (strictly) concave?*

Let us fix the potential

$$\psi(\omega) = \begin{cases} 0 & \text{if } \omega_0 = +1 \\ 1 & \text{if } \omega_0 = -1. \end{cases} \tag{50}$$

The potential energy $E(\mu) = \int_\Omega \psi \, d\mu$ becomes smaller for measures that favours configurations ω where many entries are $+1$. We can think of ψ as representing a fixed external magnetic field; the better the atoms align themselves to this field, the smaller the potential energy of their configuration. In extremo, $E(\delta_+) = 0$, but the entropy of δ_+ is zero, so we don't maximise entropy with this choice.

[7] This is the definition for one-dimensional lattices. For a d-dimensional lattice, we need to add an extra factor $(n - m + 1)^{d-1}$ in the lower and upper bounds in (49).

Pressure can also be defined by the Variational Principle. We introduce a weighing parameter $\beta \in \mathbb{R}$ between energy and entropy content of the measure. The physical interpretation of $\beta = 1/T$, where T stands for the absolute temperature (*i.e.,* degrees Kelvin normalised in some way), and thus it makes only physical sense to take $\beta \in (0, \infty)$, but we will frequently look at limit case $\beta \to 0$ and $\beta \to \infty$.

Now let the **(Variational) Pressure** be

$$P(\beta) = \sup\{h_\mu(\sigma) - \beta \int \psi \, d\mu \ : \ \mu \text{ is a shift-invariant probability measure}\} \tag{51}$$

A shift-invariant probability measure μ is called **equilibrium state** or **equilibrium measure**, if it assume the pressure in (51).

For the limit case $T \to \infty$, *i.e.,* $\beta \to 0$, the potential energy plays no role, and we are just maximising entropy. For the limit case $T \to 0$, *i.e.,* $\beta \to \infty$, the potential energy becomes all important, so in our example we expect δ_+ to be the limit equilibrium state. The physical interpretation of this statement is: as the temperature decreases to zero for some fixed external magnetic field (and also as the external magnetic field grows to infinity), the material becomes totally magnetized.

The question is now: do we find total magnetization (*i.e.,* the measure δ_+ as equilibrium state) also for some positive temperature (or finite external magnetic field)?

For each fixed measure, the function $\beta \mapsto h_\mu(\sigma) + \beta \int \psi \, d\mu$ is a straight line with slope $-\int \psi \, d\mu$ (non-positive because our potential ψ is non-negative) and abscissa $h_\mu(\sigma)$. If we look at (51) again, we can view the pressure function $\beta \mapsto P(\beta)$ as the envelope of all these straight lines. From this it follows immediately that $\beta \mapsto P(\beta)$ is continuous and convex (and non-increasing due to ψ being non-negative).

Once full magnetization is obtained, increasing β further will not change the equilibrium state anymore. Indeed, there is no measure that favours $\omega_i = +1$ more than δ_+. So if there is a finite β_0 such that δ_+ is equilibrium state, then $P(\beta) = 0$ for all $\beta \geqslant \beta_0$. We can call this a **freezing phase transition**, because at this parameter, the equilibrium state doesn't change anymore (as if the system is frozen in one configuration). The right-hand slope of the pressure function at β_0 is 0; how abrupt this **phase transition** is depends also on the left slope at β_0 which might be different from 0, but always $\geqslant 0$ because of convexity.

Let us now do the computation if there really is a phase transition at a finite β_0. For simplicity (and without justification at the moment) we will only compute the supremum in (51) over the Bernoulli measures μ_p. So then (51) simplifies to

$$P(\beta) = \sup_{p \in [0,1]} -(p \log p + (1-p)\log(1-p)) - \beta(1-p) =: \sup_{p \in [0,1]} F(\mu_p, \beta)$$

The quantity $F(\mu_p, \beta)$ is called the **free energy** of the measure μ_p. In our simplified case, it is a smooth curve in p, so to find the supremum (= maximum), we simply compute the derivative and put it equal to 0:

$$0 = \frac{\partial}{\partial p} F(\mu_p, \beta) = -(\log p - \log(1-p)) + \beta.$$

This is equivalent to $\log \frac{p}{1-p} = \beta$, *i.e.*,

$$p = \frac{e^\beta}{1+e^\beta}, \quad 1 - p = \frac{1}{1+e^\beta}$$

Substituting in $P(\beta)$, we find that the pressure is

$$\begin{aligned}
P(\beta) &= -\left(\frac{e^\beta}{1+e^\beta}\log\frac{e^\beta}{1+e^\beta} + \frac{1}{1+e^\beta}\log\frac{1}{1+e^\beta}\right) - \beta\frac{1}{1+e^\beta} \\
&= -\left(\underbrace{\frac{e^\beta+1}{1+e^\beta}\log\frac{e^\beta}{1+e^\beta}}_{\log\frac{e^\beta}{1+e^\beta}} + \underbrace{\frac{1}{1+e^\beta}\log\frac{1}{1+e^\beta} - \frac{1}{1+e^\beta}\log\frac{e^\beta}{1+e^\beta}}_{-\frac{\beta}{1+e^\beta}}\right) - \beta\frac{1}{1+e^\beta} \\
&= -\left(\log\frac{e^\beta}{1+e^\beta} - \frac{\beta}{1+e^\beta}\right) - \frac{\beta}{1+e^\beta} \\
&= \log(1+e^{-\beta}) \qquad \begin{cases} \to 0 & \text{as } \beta \to \infty \\ = \log 2 & \text{if } \beta = 0 \\ \sim -\beta & \text{as } \beta \to -\infty \end{cases}
\end{aligned}$$

So the pressure function is smooth (even real analytic) and never reaches the line $\beta \equiv 0$ for any finite β. Hence, there is no phase transition.

Exercise 9. *Verify that for potential* (50)*, μ_p is indeed a Gibbs measure. For which value of the pressure? Here it is important to incorporate the factor $-\beta$ in the potential, so $\psi_\beta(\omega) = 0$ if $\omega_0 = 1$ and $\psi_\beta(\omega) = -\beta$ if $\omega_0 = -1$.*

In the proper Ising model, the potential also contains also a local interaction term between nearest neighbors:

$$\psi(\omega) = \sum_i J\omega_i\omega_{i+1} + \psi_{ext}(\omega),$$

where $J < 0$, so neighboring atomic magnets with the same spin have lower joint energy than neighboring atoms with opposite spin. The term $\psi_{ext}(\omega)$ still stands for the external magnetic field, and can be taken as ψ in (50). This gives a problem for the infinite lattice, because here all configurations have a divergent sum $\sum_i J\omega_i\omega_{i+1}$. Ising's solution to this problem lies in first dealing with a large lattice $[-n, n]$, so the configuration space is $\{-1, +1\}^{[-n,n]}$, and considering the Gibbs measures and/or equilibrium states projected to fixed finite lattice $[-m, m]$ (these projections are called **marginal measures**), and then letting n tend to infinity. Such limits are called **thermodynamic**

limits. If there is no external magnetic field (*i.e.,* $\psi_{ext} \equiv 0$), then as $\beta \to \infty$, $n \to \infty$, there are two ergodic thermodynamic limits, namely δ_+ and δ_-. There is no preference from one over the other; this preference would arise if the is an external magnetic field of definite direction. However, no such magnetization takes place for a finite β. For this reason, Ising dismissed the model as a good explanation for magnetization of iron (and other substances). However, as was found much later, on higher dimensional lattices, the Ising model does produce phase transitions and magnetization at finite values of β (*i.e.,* positive temperature).

20.2 The Griffith-Ruelle Theorem

Let (Σ, σ) now be a one-sided or two-sided subshift of finite type. Throughout we will assume that the transition matrix is aperiodic and irreducible, so the Perron-Frobenius Theorem applies in its full force. Let $\psi : \Sigma \to \mathbb{R}$ be a potential function, which we will assume to be Hölder continuous, *i.e.,* there is $C > 0$ and $\vartheta \in (0, 1)$ such that if x_k and y_k agree for $|k| < n$, then $|\psi(x) - \psi(y)| \leqslant C\vartheta^n$. The Hölder property can be applied to ergodic sums on n-cylinders Z:

$$\begin{aligned} \sup\{S_n\psi(x) : x \in Z\} &\geqslant \inf\{S_n\psi(x) : x \in Z\} \\ &\geqslant \sup\{S_n\psi(x) : x \in Z\} - \underbrace{\sum_{k=0}^{n-1} C\vartheta^k}_{=C\frac{1-\vartheta^n}{1-\vartheta} < \frac{C}{1-\vartheta}} \end{aligned} \tag{52}$$

Definition 19. *We say that a shift-invariant probability measure μ satisfies the* **Gibbs property** *if there are constants $C_2 \geqslant C_1 > 0$ such that for all n, all n-cylinders Z and all $x \in Z$,*

$$C_1 \leqslant \frac{\mu(Z)}{e^{S_n\psi(x) - Pn}} \leqslant C_2. \tag{53}$$

Here P is some constant, which, as we will see later, coincides with the topological pressure of the system. It is the number by which we need to translate the potential such that the measure of an n-cylinder scales as $e^{S_n(\psi - P)}$.

The main theorem of this section is sometimes called, in physics the Griffith-Ruelle Theorem (which actually also include analyticity of the pressure function):

Theorem 29. *If ψ is Hölder continuous potential function on an aperiodic irreducible subshift of finite type, then there is a unique Gibbs measure μ; this measure is the unique equilibrium state for (Σ, σ, ψ).*

We will prove this theorem in various steps. We start by a trick to reduce the potentially two-sided shift space to a one-sided shift.

Definition 20. *Two potential functions ψ and χ on Σ are called* **cohomologous** *if there is a function u such that*

$$\psi = \chi + u - u \circ \sigma. \tag{54}$$

From this definition, the following consequence are immediate for σ-invariant measure:

$$\begin{aligned} S_n\psi(x) &= S_n\chi(x) + u(x) - u \circ \sigma^n(x), \\ \lim_{n\to\infty} \frac{1}{n} S_n\psi(x) &= \lim_{n\to\infty} \frac{1}{n} S_n\chi(x) \quad \mu\text{-a.e.}, \\ \int \psi \, d\mu &= \int \chi \, d\mu. \end{aligned}$$

From this it is easy to derive that cohomologous potentials have the same equilibrium states. This will be used, in the next proposition, to reduce our task from two-sided shifts spaces to one-sided shift spaces.

Proposition 15. *If (Σ, σ) is a two-sided subshift of finite type and ψ a Hölder potential, then there is a potential χ which is also Hölder continuous but depending only on forward coordinates $(x_k)_{k\geqslant 0}$ of $x \in \Sigma$, such that ψ and χ are cohomologous.*

Proof. For each symbol $b \in \{0, \dots, N-1\}$ pick a fix sequence $e^b \in \Sigma$ such that the zeroeth symbol $e^b_0 = n$. For $x \in \Sigma$, let x^* be the sequence with $x^*_k = x_k$ if $k \geqslant 0$ and $x^*_k = e^b_k$ if $k < 0$ and $x_0 = b$. Next choose

$$u(x) = \sum_{j=0}^{\infty} \psi \circ \sigma^j(x) - \psi \circ \sigma^j(x^*).$$

Note that $|\psi \circ \sigma^j(x) - \psi \circ \sigma^j(x^*)| < C\vartheta^j$, so the sum $u(x)$ converges and is continuous in x. Let n be arbitrary and set $m = \lfloor n/2 \rfloor$. If x_k and y_k coincide for all $|k| < n$, then

$$\begin{aligned} |u(x) - u(y)| &\leqslant \sum_{j=0}^{m} \big(|\psi \circ \sigma^j(x) - \psi \circ \sigma^j(y)| + |\psi \circ \sigma^j(x^*) - \psi \circ \sigma^j(y^*)| \big) \\ &\quad + \sum_{j>m} \big(|\psi \circ \sigma^j(x) - \psi \circ \sigma^j(x^*)| + |\psi \circ \sigma^j(y) - \psi \circ \sigma^j(y^*)| \big) \\ &\leqslant 2\sum_{j=0}^{m} C\vartheta^{n-j} + 2\sum_{j>m} C\vartheta^j \leqslant 4C\frac{\vartheta^m}{1-\vartheta} \leq \frac{C}{1-\vartheta}\sqrt{\vartheta}^n. \end{aligned}$$

Hence u is Hölder continuous with Hölder exponent $\sqrt{\vartheta}$ instead of ϑ.

Now for $\chi = \psi - u + u \circ \sigma$, which is also Hölder, we have

$$\begin{aligned}
\chi(x) &= \psi(x) - \sum_{j=0}^{\infty} \underbrace{\psi \circ \sigma^j(x) - \psi \circ \sigma^j(x^*)}_{\text{separate term } j=0} + \sum_{j=0}^{\infty} \psi \circ \sigma^j(\sigma x) - \psi \circ \sigma^j((\sigma x)^*) \\
&= \psi(x^*) - \sum_{j=1}^{\infty} \psi \circ \sigma^j(x) - \psi \circ \sigma^j(x^*) + \sum_{j=0}^{\infty} \psi \circ \sigma^j(\sigma x) - \psi \circ \sigma^j((\sigma x)^*) \\
&= \psi(x^*) + \sum_{j=0}^{\infty} \psi \circ \sigma^j((\sigma x)^*) - \psi \circ \sigma^{j+1}(x^*).
\end{aligned}$$

This depends only on the forward coordinates of x. □

Now that we can work on one-sided shift spaces, it is instructive to see why:

Proposition 16. *Gibbs measures of Hölder potentials are equilibrium states (*i.e., *measures that achieve the supremum in the Variational Principle).*

Proof. Let $\mathcal{P}$ is the partition into 1-cylinders, and recall that $\mathcal{P}_n = \bigvee_{k=0}^{n-1} \sigma^{-k}\mathcal{P}$ is the partition into n-cylinders. Write

$$\mathcal{Z}_n = \sum_{Z \in \mathcal{P}_n} e^{\sup\{S_n\psi(x) \,:\, x \in Z\}}.$$

be the n-th **partition function**. For Hölder continuous ψ, due to (52), whether we choose sup or inf, the result only changes by a multiplicative factor $e^{\frac{C}{1-\vartheta}}$, independently of n.

Now suppose that μ satisfies the Gibbs property (53). Summing over all n-cylinders gives

$$C_1 \frac{\mathcal{Z}_n}{e^{Pn}} \leqslant \sum_{Z \in \mathcal{P}_n} \mu(Z) = 1 \leqslant C_2 \frac{\mathcal{Z}_n}{e^{Pn}}.$$

therefore $P = \lim_n \frac{1}{n} \log \mathcal{Z}_n$. Combining this with (41) in the proof of the Variational Principle, with $\mathcal{P}_n$ instead of $\mathcal{Q}_n$, we can write

$$H_\mu(\mathcal{P}_n) + \int_\Sigma S_n\psi \; d\mu \leqslant \log \mathcal{Z}_n$$

Now we divide by n and take the limit $n \to \infty$ to obtain $h_\mu(\sigma) + \int \psi \; d\mu \leqslant P$.

For any x in an n-cylinder Z, we have

$$\begin{aligned}
-\mu(Z)\log\mu(Z) + \int_Z S_n\psi \; d\mu &\geqslant -\mu(Z)\left[\log\mu(Z) - S_n\psi(x) + \frac{C}{1-\vartheta}\right] \\
&\geqslant -\mu(Z)\left[\log C_2 e^{-Pn + S_n\psi(x)} - S_n\psi(x) + \frac{C}{1-\vartheta}\right] \\
&= \mu(Z)\left[Pn - \log C_2 - \frac{C}{1-\vartheta}\right].
\end{aligned}$$

Summing over all n-cylinders $Z \in \mathcal{P}_n$ gives

$$\begin{aligned} H_\mu(\mathcal{P}_n) + \int_\Sigma S_n\psi \, d\mu &\geqslant \sum_{Z \in \mathcal{P}_n} \mu(Z) \left[Pn - \log C_2 - \frac{C}{1-\vartheta} \right] \\ &= Pn - \log C_2 - \frac{C}{1-\vartheta}. \end{aligned}$$

Dividing by n and letting $n \to \infty$, we find $h_\mu(\sigma) + \int_\Sigma \psi \, d\mu \geqslant P$. Therefore we have equality $h_\mu(\sigma) + \int_\Sigma \psi \, d\mu = P$.

To show that $P = P_{top}(\sigma, \psi)$, take $\varepsilon > 0$ arbitrary and M such that $2^{-(M+1)} \leqslant \varepsilon < 2^{-M}$. Taking a point x in each $n+M$-cylinder then produces an (n, ε)-separated set $E_n(\varepsilon)$ of maximal cardinality. Therefore, as in (29), we find

$$\mathcal{Z}_{n+M} = \sup \left\{ \sum_{x \in E} e^{S_n\psi(x)} \ : \ E \text{ is } (n,\varepsilon)\text{-separated} \right\} =: P_n(\sigma, \psi, \varepsilon).$$

The ε-dependence on the left hand side is only in the choice of M. This dependence disappears when we take the limit $\lim_n \frac{1}{n} \log \mathcal{Z}_n = \lim_n \frac{1}{n} \log P_n(\sigma, \psi, \varepsilon)$, and therefore taking the limit $\varepsilon \to 0$ gives

$$P = \lim_{n\to\infty} \frac{1}{n} \log \mathcal{Z}_n = \lim_{\varepsilon \to 0} \lim_{n \to \infty} \frac{1}{n} \log P_n(\sigma, \psi, \varepsilon) = P_{top}(\sigma, \psi).$$

This completes the proof. □

Next we give somewhat abstract results from functional analysis to find a candidate Gibbs measure as the combination of the eigenfunction and eigenmeasure of a particular operator and its dual.

Definition 21. *The Ruelle-Perron-Frobenius operator acting on functions $f : \Sigma \to \mathbb{R}$ is defined as*

$$\mathcal{L}_\psi f(x) = \sum_{\sigma y = x} e^{\psi(y)} f(y). \tag{55}$$

The dual operator $\mathcal{L}_\psi^*$ acts on measures: $\int f \, d(\mathcal{L}_\psi^* \nu) = \int \mathcal{L}_\psi f \, d\nu$ for all $f \in L^1(\nu)$.

This operator describes how densities are transformed by the dynamics. For instance, if instead of σ we had a differentiable transformation $T : [0,1] \to [0,1]$ and $\psi = -\log|T'|$, then $\mathcal{L}_\psi f(x) = \sum_{Ty=x} \frac{1}{|T'(y)|} f(y)$ which, when integrated over $[0,1]$, we can recognise as the integral formula for a change of coordinates $x = T(y)$.

The following theorem can be seen as the operator-version of the Perron-Frobenius Theorem for matrices:

Theorem 30. *If (Σ, σ) is a one-sided subshift of finite type, with aperiodic irreducible transition matrix, then there is a unique $\lambda > 0$ and continuous positive (or more precisely: bounded away from zero) function h and a probability measure ν such that*

$$\mathcal{L}_\psi h = \lambda\ h \qquad\qquad \mathcal{L}_\psi^* \nu = \lambda\ \nu.$$

The Ruelle-Perron-Frobenius operator has the properties:

1. $\mathcal{L}_\psi$ is positive: $f \geqslant 0$ implies $\mathcal{L}_\psi f \geqslant 0$.
2. $\mathcal{L}_\psi^n f(x) = \sum_{\sigma^n y = x} e^{S_n\psi(y)} f(y)$.
3. ν is in general **not** σ-invariant. Instead it satisfies
$$\nu(\sigma A) = \lambda \int_A e^{-\psi}\ d\nu \tag{56}$$
whenever $\sigma : A \to \sigma(A)$ is one-to-one and A is measurable. Measures with this property are called $\lambda e^{-\psi}$**-conformal**.
4. Instead, the measure $d\mu = h\ d\nu$ is σ-invariant. We can always scale h such that μ is a probability measure too.
5. We will see later that $\lambda = e^P$ where P is the topological pressure.

Proof. Property 1. is obvious, since $e^{\psi(y)}$ is always positive. Property 2. follows by direct computation. For Property 3., we have

$$\begin{aligned}
\lambda \int_A e^{-\psi}\ d\nu &= \lambda \int_\Sigma e^{-\psi} 1\!\!1_A\ d\nu = \int_\Sigma e^{-\psi} 1\!\!1_A\ d(\lambda\nu) \\
&= \int_\Sigma e^{-\psi(x)} 1\!\!1_A(x)\ d(\mathcal{L}_\psi^*\nu) = \int_\Sigma \mathcal{L}_\psi(e^{-\psi(x)} 1\!\!1_A(x))\ d\nu \\
&= \int_\Sigma \sum_{\sigma y = x} e^{\psi(y)} e^{-\psi(y)} 1\!\!1_A(y)\ d\nu = \int_\Sigma \sum_{\sigma y = x} 1\!\!1_A(y)\ d\nu
\end{aligned}$$

Since $\sigma : A \to \sigma(A)$ is one-to-one, $\sum_{\sigma y = x} 1\!\!1_A(y) = 1$ if $x \in \sigma(A)$ and $= 0$ otherwise. Hence the integral $\int \sum_{\sigma y = x} 1\!\!1_A(y)\ d\nu = \nu(\sigma A)$ as required.

For Property 4., first check that

$$\mathcal{L}_\psi g(x) \cdot f(x) = \sum_{\sigma y = x} e^{\psi(y)} g(y) f(x) = \sum_{\sigma y = x} e^{\psi(y)} g(y) f(\sigma y) = \mathcal{L}_\psi(g \cdot f \circ \sigma)(x).$$

This gives

$$\begin{aligned}\int_\Sigma f\ d\mu &= \int_\Sigma f\cdot h\ d\nu = \frac{1}{\lambda}\int_\Sigma f\cdot \mathcal{L}_\psi h\ d\nu\\ &= \frac{1}{\lambda}\int_\Sigma \mathcal{L}_\psi(h\cdot f\circ\sigma)\ d\nu = \frac{1}{\lambda}\int_\Sigma h\cdot f\circ\sigma\ d(\mathcal{L}_\psi^*\nu)\\ &= \int_\Sigma f\circ\sigma\cdot h\ d\nu = \int_\Sigma f\circ\sigma\ d\mu\end{aligned}$$

Property 5. will follow from the next proposition. □

Proposition 17. *For Hölder potential ψ, the measure $d\mu = h\,d\nu$ satisfies the Gibbs property with $P = \log\lambda$.*

Proof. For each $z \in \Sigma$ and n-cylinder Z, there is at most one $y \in Z$ with $\sigma^n y = z$. Take $x \in Z$ arbitrary. Then

$$\mathcal{L}_\psi^n(h\cdot 1\!\!1_Z) = \sum_{\sigma^n y = z} e^{S_n\psi(y)}h(y)1\!\!1_Z(y) \leqslant \underbrace{e^{\frac{C}{1-\vartheta}}\|h\|_\infty}_{C_2} e^{S_n\psi(x)}.$$

Hence

$$\begin{aligned}\mu(Z) &= \int_Z h\ d\nu = \int_\Sigma h\cdot 1\!\!1_Z\ d\nu = \lambda^{-n}\int_\Sigma h\cdot 1\!\!1_Z\ d(\mathcal{L}_\psi^{*n}\nu)\\ &= \lambda^{-n}\int_\Sigma \mathcal{L}_\psi^n(h\cdot 1\!\!1_Z)\ d\nu \leqslant C_2\lambda^{-n}e^{S_n\psi(x)}.\end{aligned} \tag{57}$$

On the other hand, since the subshift of finite type is irreducible, there is some uniform integer M and $y \in Z$ such that $\sigma^{n+M}(y) = z$. Therefore

$$\mathcal{L}_\psi^n(h\cdot 1\!\!1_Z) \geqslant e^{S_{n+M}\psi(y)}h(y) \geqslant \underbrace{e^{-M\|\psi\|_\infty}e^{-\frac{C}{1-\vartheta}}\cdot\inf h}_{C_1} e^{S_n\psi(x)}.$$

Integrating over Z gives us $\mu(Z) \geqslant C_1\lambda^{-n}e^{S_n\psi(x)}$ by the same reasoning as in (57). Therefore

$$C_1 \leqslant \frac{\mu(Z)}{\lambda^{-n}e^{S_n\psi(x)}} \leqslant C_2,$$

for all n-cylinders and thus if we choose $e^P = \lambda$, we obtain the Gibbs property. □

Lemma 13. *The Gibbs measure is unique.*

Proof. If both μ and μ' satisfy (53) for some constants C_1, C_2, P and C_1', C_2', P' then we can first take (53) for μ' and sum over all n-cylinders. This gives

$$C_1'e^{-P'n}\sum_{Z\in\mathcal{P}_n} e^{S_n\psi(x)} \leqslant 1 \leqslant C_2'e^{-P'n}\sum_{Z\in\mathcal{P}_n} e^{S_n\psi(x)},$$

so that $P' = \lim_n \frac{1}{n} \log \sum_{Z \in \mathcal{P}_n} e^{S_n\psi(x)}$, independently of μ'. Therefore $P' = P$.

Now divide (53) for μ' by the same expression for μ. This gives

$$\frac{C_1'}{C_2} \leqslant \frac{\mu'(Z)}{\mu(Z)} \leqslant \frac{C_2'}{C_1}$$

independently of Z. Therefore μ' and μ are equivalent: they have the same null-sets. In particular, for each continuous f, the set of points $x \in \Sigma$ for which the Birkhoff Ergodic Theorem holds for μ' and μ differs by at most a nullset. For any point which is typical for both, we find $\int f \, d\mu' = \lim_n \frac{1}{n} S_n f(x) = \int f \, d\mu$. Therefore $\mu = \mu'$. $\square$

20.3 Upper semicontinuity of entropy

For a continuous potential $\psi : X \to \mathbb{R}$, and a sequence of measures $(\mu_n)_{n\in\mathbb{N}}$ such that $\mu_n \to \mu$ in the weak* topology, we always have $\int \psi \, d\mu_n \to \int \psi \, d\mu$, simply because that is the definition of weak* convergence. However, entropy isn't continuous in this sense. For example, if (Σ, σ) is the full shift on two symbols, then the $\frac{1}{2}$-$\frac{1}{2}$ Bernoulli measure μ is the measure of maximal entropy $\log 2$. If $x \in \Sigma$ is a typical point (in the sense of the Birkhoff Ergodic Theorem), then we can create a sequence of measure μ_n by

$$\mu_n = \frac{1}{n} \sum_{j=0}^{n-1} \delta_{\sigma^j y}$$

where $y = \overline{x_0 x_1 \dots x_{n-1}}$ is the n-periodic point in the same n-cylinder as x. For these measure $\mu_n \to \mu$ in the weak* topology, but since μ_n is supported on a single periodic orbit, the entropy $h_{\mu_n}(\sigma) = 0$ for every n. Therefore

$$\lim_{n\to\infty} h_{\mu_n}(\sigma) = 0 < \log 2 = h_\mu(\sigma).$$

Lacking continuity, the best we can hope for is **upper semicontinuity** (USC) of the entropy function, *i.e.,*

$$\mu_n \to \mu \text{ implies } \limsup_{n\to\infty} h_{\mu_n}(\sigma) \leqslant h_\mu(\sigma).$$

In other words, the value of h can make a jump **upwards** at the limit measure, but not downwards. Fortunately, the entropy function $\mu \mapsto h_\mu(\sigma)$ is indeed USC for subshifts on a finite alphabet, and USC is enough to guarantee the existence of equilibrium states.

Proposition 18. *Let (X, T) be a continuous dynamical system on a compact metric space X. Assume that potential $\psi : X \to \mathbb{R}$ is continuous. If the entropy function is USC, then there is an equilibrium state,* *Proof.* We use the Variation Principle

$$P(\psi) = \sup\{h_\nu(T) + \int \psi \, d\nu \ : \ \nu \text{ is } T\text{-invariant probability measure}\}. \tag{58}$$

Hence there exists a sequence $(\mu_n)_{n\in\mathbb{N}}$ such that $P(\psi) = \lim_n h_{\mu_n}(T) + \int \psi \, d\mu_n$. Passing to a subsequence (n_k) if necessary, we can assume that $\mu_{n_k} \to \mu$ as $k \to \infty$ in the weak* topology, and therefore $\int \psi \, d\mu_{n_k} \to \int \psi \, d\mu$ as $k \to \infty$. Due to upper semicontinuity,

$$P(\psi) = \limsup_{k\to\infty} h_{\mu_{n_k}}(T) + \int \psi \, d\mu_{n_k} \leqslant h_\mu(T) + \int \psi \, d\mu,$$

but also $h_\mu(T) + \int \psi \, d\mu \leqslant P(\psi)$ by (58). Hence μ is an equilibrium state. □

The following corollary follows in the same way.

Corollary 6. *Let (X,T) be a continuous dynamical system on a compact metric space X, and suppose that the entropy function is USC. Let ψ_β be a family (continuous in β) of continuous potentials and $\beta \to \beta^*$. If μ_β are equilibrium states for ψ_β and $\mu_\beta \to \mu_{\beta^*}$ in the weak* topology as $\beta \to \beta^*$, then μ_{β^*} is an equilibrium state for ψ_{β^*}.*

In particular, the important case $\beta \mapsto P(\beta \cdot \psi)$ is a continuous function. Upper semi-continuity of entropy also gives us another way of characterizing entropy:

Lemma 14 (Dual Variational Principle)**.** *Let (X,T) be a continuous dynamical system on a compact metric space. Assume that the entropy function is upper semi-continuous and that $P(0) < \infty$. Then*

$$h_\mu(T) = \inf\{P(\psi) - \int \psi \, d\mu \ : \ \psi : X \to \mathbb{R} \text{ continuous}\}.$$

Proof. See [14, Theorem 4.2.9] or [24, Theorem 9.12]. □

20.4 Smoothness of the pressure function

In Section 20.2 we have given conditions under which a Gibbs measure is unique. Gibbs measures are equilibrium states, but that doesn't prove uniqueness of equilibrium states. There could in principle be equilibrium states that are not Gibbs measures. In this section we will connect uniqueness of equilibrium states of a parametrised family ψ_β of potentials to smoothness of the pressure function $\beta \mapsto P(\psi_\beta)$. In fact, the remaining part of the Griffith-Ruelle Theorem is about smoothness, more precisely analyticity, of pressure function when $\psi_\beta = \beta \cdot \psi$, for inverse temperature $\beta \in \mathbb{R}$.

Theorem 31 (Griffith-Ruelle Theorem (continued)). *If ψ is Hölder continuous potential function on an aperiodic irreducible subshift of finite type, then the pressure function*

$$\beta \mapsto P(\beta \cdot \psi)$$

is real analytic.

We will not prove analyticity here (which depends on perturbation theory of operators), but rather focus on how differentiability of $\beta \mapsto P(\psi_\beta)$ is related to equilibrium states. In the simplest case when $\psi_\beta = \beta \cdot \psi$, then the graph of

$$\beta \mapsto P(\beta \cdot \psi) := \sup\{h_\nu(T) + \beta \int \psi \, d\nu : \nu \text{ is } T\text{-invariant probability measure}\}$$

is the envelope of straight lines $\beta \mapsto h_\nu(T) + \beta \int \psi \, d\nu$, and therefore continuous. We think of ψ (or at least $\int \psi \, d\nu$) as non-positive, so that maximising $P(\beta)$ corresponds to maximising entropy and minimising energy in agreement with the Laws of Thermodynamics. Hence the graph $\beta \mapsto P(\beta)$, as the envelop of non-increasing lines, is non-increasing and convex.

Furthermore, if μ_0 is an equilibrium state for β_0, and $\beta \mapsto P(\beta)$ is differentiable at $\beta = \beta_0$, then $P'(\beta_0) = \int \psi \, d\mu_0$. Hence if μ_0 and μ_0' are two different equilibrium states for β_0 with $\int \psi \, d\mu_0 \neq \int \psi \, d\mu_0'$, then $\beta \mapsto P(\beta)$ cannot be differentiable at $\beta = \beta_0$.

Definition 22. *Given a continuous potential $\psi : X \to \mathbb{R}$, we say that:*

- *a measure μ on X is a* **tangent measure** *if*

$$P(\psi + \phi) \geqslant P(\psi) + \int \phi \, d\mu \text{ for all continuous } \phi : X \to \mathbb{R}. \tag{59}$$

- *P is* **differentiable** *at ψ if $P(\psi) < \infty$ and there is a unique tangent measure.*

It would be more correct to speak of **tangent functional** *since a priori, we just have $\nu \in C^*(X)$, but in all cases ν turns out to be indeed an "unsigned" probability measure.*

So, compared to differentiability of $\beta \mapsto P(\beta \cdot \psi)$, differentiability in the above sense requires (59) not just for $\phi = (\beta - 1) \cdot \psi$ (which follows from convexity of $\beta \mapsto P(\beta \cdot \psi)$), but for all continuous $\phi : X \to \mathbb{R}$.

Theorem 32. *Let (X, T) be a continuous dynamical system on a compact metric space X, and suppose that the entropy function is USC. Let $\psi : X \to \mathbb{R}$ be a continuous potential. Then P is differentiable at ψ with derivative μ if and only if*

$$\lim_{\varepsilon \to 0} \frac{P(\psi + \varepsilon\phi) - P(\psi)}{\varepsilon} = \int \phi \, d\mu \tag{60}$$

for all continuous $\phi : X \to \mathbb{R}$. In this case, μ is the unique equilibrium state for ψ.

Proof. We start by proving that the tangent measures are exactly the equilibrium states. Assume that μ is an equilibrium state for ψ. Then

$$\begin{aligned} P(\psi+\phi) &= \sup_\nu\{h_\nu(T) + \int \psi\ d\nu + \int \phi\ d\nu\} \\ &\geqslant h_\mu(T) + \int \psi\ d\mu + \int \phi\ d\mu = P(\psi) + \int \phi\ d\mu \end{aligned}$$

for all continuous $\phi : X \to \mathbb{R}$, so μ is a tangent measure.

For the converse, assume that μ satisfies (59). Since μ satisfies (59), we have

$$\begin{cases} P(\psi) + 1 = P(\psi+1) \geqslant P(\psi) + \int d\mu, \\ P(\psi) - 1 = P(\psi-1) \geqslant P(\psi) - \int d\mu, \end{cases}$$

so $\int d\mu = 1$ follows. Furthermore, if $\phi \geqslant 0$, we have

$$P(\psi) \geqslant P(\psi - \phi) \geqslant P(\psi) - \int \phi\ d\mu,$$

so $\int \phi\ d\mu \geqslant 0$. This shows that μ is an "unsigned" probability measure. To prove T-invariance, recall about cohomologous functions that

$$P(\psi) = P(\psi + \eta \cdot (\phi \circ T - \phi)) \geqslant P(\psi) + \eta \int \phi \circ T - \phi\ d\mu,$$

hence $0 \geqslant \eta \int \phi \circ T - \phi\ d\mu$. Since η can be both positive or negative, there is only one possibility: $0 = \int \phi \circ T - \phi\ d\mu$, and so μ is indeed T-invariant.

Since $\psi : X \to \mathbb{R}$ is continuous on a compact space, so $\inf \psi > -\infty$. We have $P(0) + \inf \psi \leqslant P(\psi) < \infty$ by assumption, so also $P(0) \leqslant P(\psi) - \inf \psi < \infty$. Therefore we can apply Lemma 14, and obtain

$$\begin{aligned} h_\mu(T) &= \inf\{P(\psi+\phi) - \int \psi + \phi\ d\mu\ :\ \phi : X \to \mathbb{R} \text{ continuous}\} \\ &\geqslant \inf\{P(\psi) + \int \phi\ d\mu - \int \psi + \phi\ d\mu\ :\ \phi : X \to \mathbb{R} \text{ continuous}\} \\ &= P(\psi) - \int \phi\ d\mu \geqslant h_\mu(T). \end{aligned}$$

Therefore μ is indeed an equilibrium state.

Now for the second half of the proof, assume that P is differentiable at ψ with derivative μ, so μ is the only tangent measure, and therefore only equilibrium state. We need to establish (60). For $\varepsilon \neq 0$ and given ϕ, let μ_ε be an equilibrium state for $\psi + \varepsilon\phi$. Then $\mu_\varepsilon \to \mu$ as $\varepsilon \to 0$ by Corollary 6. Since μ is a tangent measure

$$P(\psi + \varepsilon\phi) - P(\psi) \geqslant \varepsilon \int \phi\ d\mu,$$

and since μ_ε are also tangent measures,

$$-\left(P(\psi + \varepsilon\phi) - P(\psi)\right) = P(\psi + \varepsilon\phi - \varepsilon\phi) - P(\psi + \varepsilon\phi) \geqslant -\varepsilon \int \phi \, d\mu_\varepsilon.$$

Combining the two, we find

$$\int \phi \, d\mu \leqslant \frac{P(\psi + \varepsilon\phi) - P(\psi)}{\varepsilon} \leqslant \int \phi \, d\mu_\varepsilon$$

if $\varepsilon > 0$ or with reversed inequalities if $\varepsilon < 0$. Now $\int \phi \, d\mu_\varepsilon \to \int \psi \, d\mu$ as $\varepsilon \to 0$, so (60) follows.

Conversely, assume that (60) holds for μ and all continuous $\phi : X \to \mathbb{R}$. If ν is an arbitrary tangent measure (which we know exists, because an equilibrium measure exists by Proposition 18 and we just proved that and equilibrium measures are tangent measures), then

$$\int \phi \, d\mu = \lim_{\varepsilon \searrow 0} \frac{P(\psi + \varepsilon\phi) - P(\psi)}{\varepsilon} \geqslant \int \phi \, d\nu,$$

and also

$$\int \phi \, d\mu = \lim_{\varepsilon \nearrow 0} \frac{P(\psi + \varepsilon\phi) - P(\psi)}{\varepsilon} \leqslant \int \phi \, d\nu.$$

Hence $\int \phi \, d\mu = \int \phi \, d\nu$ for all continuous $\phi : X \to \mathbb{R}$, whence $\mu = \nu$, and P is indeed differentiable with single tangent measure as derivative. $\square$

In view of the Griffith-Ruelle Theorem, this motivates the definition:

Definition 23. *The system (X, T) with potential $\psi : X \to \mathbb{R}$ undergoes a* **phase transition** *at parameter β_0 if $\beta \mapsto P(\beta \cdot \psi)$ fails to be analytic at β_0.*

It is where pressure fails to be analytic, that equilibrium states may be non-existent (possible, if the potential is non-continuous), non-unique (possible, if the potential is non-Hölder) and/or discontinuous under change of parameters.

20.5 Phase transitions for non-Hölder potentials

The analyticity of the pressure function may fail if ψ is not Hölder continous. Hofbauer [12] gave the first such example for the full shift and a piecewise constant potential, and we will discuss this example later on. First we present a very much related example on the interval, namely the **Pomeau-Manneville map** $f = f_\alpha : [0,1] \to [0,1]$ for parameter $\alpha \in (0, \infty)$, defined as

$$f(x) = \begin{cases} x(1 + (2x)^\alpha) & x \in [0, \frac{1}{2}], \qquad \alpha \in (0, \infty) \\ 2x - 1 & x \in (\frac{1}{2}, 1]. \end{cases}$$

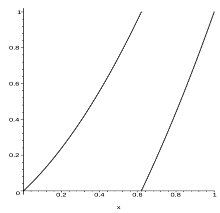

Figure 4: The Pomeau-Manneville map T_α with $\alpha = 1$, and the corresponding first return map F.

This map is non-uniformly expanding: the neutral fixed point at 0 prevents it from being uniformly expanding, and this changes the behaviour entirely. Here the potential of interest is $-\log|T'(x)| = -\log(1 + (1+\alpha)(2x)^\alpha) \approx (1+\alpha)(2x)^\alpha$ for x close to 0.

Theorem 33. *The Pomeau-Manneville map has an invariant probability measure $\mu \ll$ Leb if and only if $\alpha \in (0,1)$. This measure is also an equilibrium measure, and the Dirac measure δ_0 is an equilibrium measure as well. If $\alpha \geq 1$, and $\mu \ll$ Leb is f-invariant, then μ is σ-finite and infinite, and δ_0 is the unique equilibrium state.*

Sketch of Proof. Since f is not uniformly expanding, we will use the first return map $F = f^\tau : Y \to Y$ to the set $Y = (\frac{1}{2}, 1]$. Here $\tau(x) = \min\{n \geq 1 : f^n(x) \in Y$ is the first return time to Y. Define recursively

$$y_0 = 1, y_1 = \frac{1}{2}, y_n = f^{-1}(y_{n-1}) \cap [0, \frac{1}{2}).$$

It can be computed (see [9] and [23, Appendix B]) that $y_n = \frac{1}{2(\alpha n)^{1/\alpha}}$. Let $Z_n = (y_n, y_{n-1})$ and $\tilde{Z}_n = f^{-1}(Z_n) \cap (\frac{1}{2}, 1]$. It is not hard to check that $Z_n = \{y \in Y : \tau(y) = n\}$ and $F : \tilde{Z}_n \to (\frac{1}{2}, 1]$ is a diffeomorphism. Without proof, we say that $F : \tilde{Z}_n \to (\frac{1}{2}, 1]$ has uniform distortion bounds, and that therefore a version of the Folklore Theorem 9 applies. Therefore we have an F-invariant measure ν equivalent to Lebesgue, and $h(x) = \frac{d\nu(x)}{dx}$ is bounded and bounded away from zero.

Similar to Proposition 6, we obtain an f-invariant measure by the formula

$$\mu(A) = \sum_{k \geq 0} \nu(f^{-k} \cap \{y \in Y : \tau(y) > k\}); \tag{61}$$

it is σ-finite and can be normalised to a probability measure by dividing by $\Lambda = \mathbb{E}_\nu(\tau) = \sum_{n \geq 1} n\nu(\{y \in Y : \tau(y) = n\})$ provided $\Lambda < \infty$. Since $\{y \in Y : \tau(y) = n\} = \tilde{Z}_n$ is an interval of measure

$$\nu(\tilde{Z}_n) \sim \frac{h(\frac{1}{2})}{4\alpha^{1/\alpha}} \left(\frac{1}{n^{1/\alpha}} - \frac{1}{(n+1)^{1/\alpha}} \right) \sim \frac{h(\frac{1}{2})}{4\alpha^{1/\alpha}} \frac{1}{n^{1+1/\alpha}},$$

we find $\Lambda \sim \sum_{n\geq 1} \frac{h(\frac{1}{2})}{2\alpha^{1/\alpha}} \frac{1}{n^{1/\alpha}} < \infty$ precisely if $\alpha < 1$. Hence μ is finite if and only if $\alpha \in (0,1)$.

Let us look at the induced map F once more, for $\alpha < 1$, so $\Lambda < \infty$. It has countably many branches $F : \tilde{Z}_n \to [\frac{1}{2}, 1]$ on which F is C^2 expanding and onto. One can show that

$$\frac{F'(x)}{F'(y)} \leq K \qquad \text{for all } n \text{ and all } x, y \in \tilde{Z}_n.$$

Therefore, any k-cylinder J w.r.t. the partition $\{J\tilde{Z}_n\}_{n\in\mathbb{N}}$ satisfies

$$\nu(J) = \int_J h(x)\, dx \approx 2|J| = \frac{2|F^k(J)|}{DF^k(\xi)} \approx \frac{1}{DF^k(x)}$$

for all $x \in J$. Here we used the Mean Value Theorem in the second step, and we used $\approx$ in the sense of: $a \approx b$ if the is a uniform constant $C > 0$ such that $\frac{1}{C} \leq \frac{a}{b} \leq C$. If we compute the induced potential

$$\Psi(x) := \sum_{k=1}^{\tau(x)-1} -\log T'(T^k(x)),$$

then we get $\Psi(x) = -\log F'(x)$ by the Chain Rule. Combining with the above gives a constant $C > 0$ such that

$$\frac{1}{C} \leq \frac{\nu(J)}{e^{S_k\Psi(x)}} \leq C \qquad \text{for all } x, y \in J \text{ and all } k\text{-cylinders } J.$$

In other words, ν is a Gibbs measure with pressure $P = P(\Psi) = 0$. Since the Gibbs measures coincide with the equilibrium states, we conclude that ν is the equilibrium state of (Y, F, Ψ), and it is unique because the Gibbs measure is unique.

One can show that for $\psi = -\log T'$, and μ as in (61), but normalized by dividing by $\Lambda < \infty$

$$\begin{cases} h_\mu(T) = \frac{1}{\Lambda} h_\nu(F) & \text{(this is \textbf{Abramov's formula})}, \\ \int \psi\, d\mu = \frac{1}{\Lambda} \int \Psi\, d\nu, \end{cases} \tag{62}$$

and therefore $h_\mu(T) + \int \psi\, d\mu = 0$. If there was another measure $\tilde{\mu} \neq \delta_0$ such that $h_{\tilde{\mu}}(T) + \int \psi\, d\tilde{\mu} \geq 0$, then there would be a $\tilde{\nu}$ (related to $\tilde{\mu}$ by (61) and (62), then therefore $h_{\tilde{\nu}}(F) + \int \psi\, d\tilde{\nu} \geq 0$, contradicting that ν is the unique equilibrium state for (Y, F, Ψ).

It follows that μ is an equilibrium state, but not the unique one because the Dirac measure δ_0 is another one, but δ_0 doesn't "lift" to an F-invariant measure. $\square$

Symbolically, this can be interpreted as the full shift (Σ, σ) with the **Hofbauer potential**

$$\psi(e) = \begin{cases} -\frac{\gamma}{n} & e \in [0^{n-1}1] \\ 0 & e = 0^\infty \end{cases} \qquad\qquad \gamma = \frac{1+\alpha}{2\alpha} h(\frac{1}{2})$$

corresponds to the Pomeau-Manneville map above. Therefore, for $e \in [0^{n-1}1]$,

$$|\psi(e) - \psi(0^\infty)| = \frac{\gamma}{n} \not< C\vartheta^n$$

for large n, so ψ fails to be Hölder continuous. For (Σ, σ, ψ), the Dirac measure $\delta_0 = \delta_{0^\infty}$ has $h_{\delta_0}(\sigma) + \int \psi d\delta_0 = 0$, and in fact $h_{\delta_0}(\sigma) + \beta \int \psi d\delta_0 = 0$ for all $\beta = 0$. It follows that the pressure function must be non-negative. It was shown (cf. [18]) that

$$P(\beta\psi) \begin{cases} > 0 & \beta < 1 \text{ and is a unique fully supported equilibrium state,} \\ = 0 & \beta \geq 1 \text{ and } \delta_0 \text{ is an equilibrium state (unique if } \beta > 1), \end{cases}$$

and for $\beta = 1$, δ_0 is the unique equilibrium state if and only if $\alpha \geq 1$. Furthermore, $\beta \to P(\beta\psi)$ is only C^0 at $\beta = 1$ if $\alpha < 1$ and C^1 if $\alpha \geq 1$, but in either case $\beta \to P(\beta\psi)$ is not real anlytic, so there is a phase transtion at $\beta = 1$.

21 Hausdorff Dimension of Repellors

Let $f : D \subset [0,1] \to [0,1]$ be defined on a domain $D = \cup_{k=0}^{N-1} D_k$, where each D_k is a closed interval and $f : D_k \to [0,1]$ is surjective, C^2-smooth and expanding, *i.e.,* $\inf\{|f'(x)| : x \in D\} > 1$. Recall that $f^n = f \circ \cdots \circ f$ is the n-fold composition of a map and define

$$X = \{x \in [0,1] : f^n(x) \in D \text{ for all } n \geqslant 0\}.$$

This set X is sometimes called the **repellor** of f, and is usually a Cantor set, *i.e.,* compact, totally disconnected and without isolated points.

Example 7. *If*

$$f(x) = \begin{cases} 3x & \text{if } x \in [0, \frac{1}{3}] = D_0, \\ 3x - 2 & \text{if } x \in [\frac{2}{3}, 1] = D_1, \end{cases}$$

then X becomes the middle third Cantor set.

Example 8. *The full tent-map is defined as*

$$T(x) = \begin{cases} 2x & \text{if } x \in [0, \frac{1}{2}] = D_0, \\ 2(1-x) & \text{if } x \in [\frac{1}{2}, 1] = D_1. \end{cases}$$

Here $X = [0,1]$, so not a Cantor set. (In this case, D_0 and D_1 overlap at one point, and that explains the difference.)

Definition 24. *Given some set A, an* **(open)** ε**-cover** $\mathcal{U} = \{U_j\}_{j \in \mathbb{N}}$ *of A is a collection of open sets such that $A \subset \cup_j U_j$ and the diameters* $\mathrm{diam}(U_j) < \varepsilon$ *for all j.*[8]

[8] We can include $U_j = \emptyset$ for some j, so finite covers $\{U_j\}_{j=1}^R$ can always be extended to countable covers $\{U_j\}_{j \in \mathbb{N}}$ if necessary.

The δ**-dimensional Hausdorff measure** *is defined as*

$$\mu_\delta(A) = \lim_{\varepsilon\to 0} \inf\{\sum_j (\mathrm{diam}(U_j))^\delta \ : \ \mathcal{U} \textit{ is an open } \varepsilon\textit{-cover of } A\}.$$

It turns out that there is a unique δ_0 *such that*

$$\mu_\delta(A) = \begin{cases} \infty & \textit{if } \delta < \delta_0, \\ 0 & \textit{if } \delta > \delta_0. \end{cases}$$

This δ_0 *is called the* **Hausdorff dimension** *of* A*, and it is denoted as* $\dim_H(A)$*.*

Lebesgue measure on the unit cube $[0,1]^n$ coincides, up to a multiplicative constant, with n-dimensional Hausdorff measure. However, for "fractal" sets such as the middle third Cantor sets, the "correct" value of δ_0 can be non-integer, as we will argue in the next example.

Example 9. *Let* X *be the middle third Cantor set. For each* n*, we can cover* X *with* 2^n *closed intervals of length* 3^{-n}*, namely*

$$[0,3^{-n}] \cup [2\cdot 3^{-n}, 3\cdot 3^{-n}] \cup [6\cdot 3^{-n}, 7\cdot 3^{-n}] \cup [8\cdot 3^{-n}, 9\cdot 3^{-n}] \cup \dots \cup [(3^n-1)\cdot 3^{-n}, 1].$$

We can make this into an open cover $\mathcal{U}_\varepsilon$ *(with* $\varepsilon = 3^{-n}(1+2\cdot 3^{-n})$*) by thickening these intervals a little bit,* i.e., *replacing* $[m\cdot 3^{-n}, (m+1)\cdot 3^{-n}]$ *by* $(m\cdot 3^{-n} - 3^{-2n}, (m+1)\cdot 3^{-n} + 3^{-2n})$*. Then*

$$\mu_\delta(X) \leqslant 2^n \cdot (3^{-n} + 2\cdot 3^{-2n})^\delta = 2^n \cdot 3^{-\delta n}\cdot (1+2\cdot 3^{-n})^\delta =: E_n.$$

Then

$$\lim_{n\to\infty} E_n = \begin{cases} \infty & \textit{if } \delta < \frac{\log 2}{\log 3}, \\ 1 & \textit{if } \delta = \frac{\log 2}{\log 3}, \\ 0 & \textit{if } \delta > \frac{\log 2}{\log 3}. \end{cases}$$

This shows that $\dim_H(X) \leqslant \frac{\log 2}{\log 3}$*. In fact,* $\dim_H(X) = \frac{\log 2}{\log 3}$*, but showing that covers* $\mathcal{U}_\varepsilon$ *are "optimal" is a bit messy, and we will skip this part.*

Coming back to our expanding interval map f, we choose the potential

$$\psi_\beta(x) = -\beta \log |f'(x)|,$$

which is C^1-smooth on each D_k, and negative for $\beta > 0$. The ergodic sum

$$\begin{aligned} S_n\psi_\beta(x) &= -\beta \sum_{k=0}^{n-1} \log|f'\circ f^k(x)| \\ &= -\beta \log \prod_{k=0}^{n-1} |f'\circ f^k(x)| = \log|(f^n)'(x)|^{-\beta} \qquad (63) \end{aligned}$$

by the Chain Rule.

Theorem 34. *Let* $([0,1], f)$ *with repellor*

$$X = \{x \in [0,1] \ : \ f^n(x) \in D = \cup_k D_k \text{ for all } n \geqslant 0\}$$

and potential $\psi_\beta = -\beta \log |f'|$ *be as above. Then there is a unique* β_0 *at which the pressure* $P(\psi_\beta)$ *vanishes, and* $\dim_H(X) = \beta_0$.

Sketch of Proof. We use symbolic dynamics on X by setting

$$e(x) = y_0 y_1 y_2 \dots \text{ with } y_n = k \in \{0, \dots, N-1\} \text{ if } f^n(x) \in D_k.$$

This uniquely associates a code $y \in \Sigma := \{0, \dots, N-1\}^{\mathbb{N}_0}$ to x provided the D_k's don't overlap, as in Example 8. If some D_k's overlap at one point, this affects only countably many points, and therefore we can neglect them. Conversely, since f is expanding, each code $y \in \Sigma$ is associated to no more than one $x \in X$.

To each n-cylinder set $[y_0 y_1 \dots y_{n-1}] = Z \subset \Sigma$, we can associate a closed interval J such that $f^k(J) \subset D_{y_k}$ for $0 \leqslant k < n$, and in fact $f^{n-1}(J) = D_{y_{n-1}}$ and $f^n(J) = [0,1]$.

The C^2-smoothness of f guarantees that ψ_β transfers to a Hölder potential $\tilde\psi_\beta(y) := \psi_\beta \circ e^{-1}(y)$ on Σ, and therefore, for each β, we can apply the Griffith-Ruelle Theorem and obtain a unique equilibrium state which is also a Gibbs measure. Use the coding map $e : X \to \Sigma$ to transfer this to (X, f, ψ_β): For each $\beta \in \mathbb{R}$, there is a unique equilibrium state μ_β which is also a Gibbs measure, for ψ_β.

Therefore, there are $C_1, C_2 > 0$ depending only on f and β, such that for all n, all interval J associated to n-cylinders and all $x \in J \cap X$,

$$C_1 \leqslant \frac{\mu_\beta(J \cap X)}{e^{S_n(\psi_\beta(x) - P)}} \leqslant C_2, \tag{64}$$

where $P = P(\psi_\beta)$ is the pressure.

Recall from (63) that $e^{S_n(\psi_\beta(x)-P)} = e^{-nP}|(f^n)'(x)|^{-\beta}$ for $x \in J \cap X$; in fact the same holds for all $x \in J$. By the Mean Value Theorem, and since $f^n(J) = [0,1]$, there is $x_J \in J$ such that $|(f^n)'(x_J)| = 1/\mathrm{diam}(J)$. Now we don't know if $x_J \in X$, but we use a distortion argument[9] to rewrite (64) to

$$\frac{\mu_\beta(J)}{C_2} \leqslant e^{-Pn} \mathrm{diam}(J)^\beta \leqslant \frac{\mu_\beta(J)}{C_1}$$

and summing over all cylinder sets, we arrive at

$$\frac{1}{C_2} \leqslant e^{-Pn} \sum_J \mathrm{diam}(J)^\beta \leqslant \frac{1}{C_1}. \tag{65}$$

[9] which we will sweep under the carpet

Now for $\beta = 0$, this gives $\frac{1}{C_2} \leqslant e^{-Pn}\#\{$ intervals $J\} \leqslant \frac{1}{C_1}$, and since there are N^n intervals, we get $P(\psi_0) = \lim_n \frac{1}{n} \log N^n = \log N > 0$, which is indeed the topological entropy of the map f.

We have $\sum_J \operatorname{diam}(J) \leqslant 1$, and therefore, for $\beta > 1$, $\sum_J \operatorname{diam}(J)^\beta \to 0$ exponentially in n. Hence (65) implies that $P(\psi_\beta) < 0$ for all $\beta > 1$. Now since $\beta \mapsto P(\psi_\beta)$ is non-increasing and convex, this means that there is a unique β_0 such that $P(\psi_\beta) = 0$ for $\beta = \beta_0$.

For this β_0, we find

$$\frac{1}{C_2} \leqslant \sum_J \operatorname{diam}(J)^{\beta_0} \leqslant \frac{1}{C_1}.$$

The sets J can be thickened a bit to produce an open ε-cover $\mathcal{U}_\varepsilon$ (with $\varepsilon < 2(\inf |f'|)^{-n}) \to 0$ as $n \to \infty$). This gives $\dim_H(X) \leqslant \beta_0$. To show that also $\dim_H(X) \geqslant \beta_0$, we need a similar argument that covers $\mathcal{U}_\varepsilon$ are "optimal" that we skipped in Example 9, and which we will omit here as well. □

Exercise 10. *Assume that $\cup_k D_k = [0,1]$ as in Example 8. Show that $\beta_0 = 1$ and that the unique equilibrium state μ_1 is equivalent to Lebesgue measure.*

22 Gibbs distributions and large deviations - an example

The following is an adaptation of Example 1.2.1. from Keller's book [14]. Assume first that the entire system consists of a single particle that can assume states in alphabet $\mathcal{A} = \{0, \dots, N-1\}$, with energies $-\beta\psi_0(a)$ (where parameter $\beta \in \mathbb{R}$ denotes the inverse temperature). We call

$$\mathbb{P}(x = a) = q_\beta(a) := \frac{e^{-\beta\psi_0(a)}}{\sum_{a' \in \mathcal{A}} e^{-\beta\psi_0(a')}} \tag{66}$$

a **Gibbs distribution**. (The Gibbs distribution in this section should not be confused with a Gibbs measure that satisfies the Gibbs property (49).) Note that a Gibbs distribution isn't a fixed state the particle is in, it is a probability distribution indicating (presumably) what proportion of time the particle assumes state $a \in \mathcal{A}$.

In this simple case, the configuration space $\mathcal{A}$ and as there is no dynamics, entropy is just

$$H(q_\beta) = -\sum_{p \in \mathcal{P}} q_\beta(p) \log q_\beta(p)$$

with respect to the only sensible partition, namely into single symbols: $\mathcal{P} = \{\omega = a\}_{a \in \mathcal{A}}$.

We know from Corollary 4 that

$$H(q_\beta) - \beta \int \psi_0 \, dq_\beta \geqslant H(\pi) - \beta \int \psi_0 \, d\pi$$

for every probability measure π on $\mathcal{A}$ with equality if and only if $\pi = q_\beta$. Hence the Gibbs measure is the equilibrium state for ψ_0. We take this as inspiration to measure how far π is from the "optimal" measure q_β by defining

$$d_\beta(\pi) = \left(H(q_\beta) - \beta \int \psi_0 \, dq_\beta \right) - \left(H(\pi) - \beta \int \psi_0 \, d\pi \right). \tag{67}$$

Let us now replace the single site by a finite lattice or any finite collection G of sites, say $n = \#G$, with particles at every site assuming states in $\mathcal{A}$. Thus now the configuration space is $\Sigma = \mathcal{A}^G$ of cardinality $\#\Sigma = N^n$, where we think of n as huge (number of Avogadro or like).

Assume that the energy $\psi(\omega)$ of configuration $\omega \in \Sigma$ is just the sum of the energies of the separate particles: $\psi(\omega) = \sum_{g \in G} \psi_0(\omega_g)$. So there is no interaction between particles whatsoever; no coherence in the set G.

We can still define the Gibbs measure (and hence equilibrium state for ψ) as before; it becomes the product measure of the Gibbs measures at each site:

$$\mu_\beta(\omega) = \frac{e^{-\beta\psi(\omega)}}{\sum_{\omega' \in \Sigma} e^{-\beta\psi(\omega')}} = \prod_{g \in G} \frac{e^{-\beta\psi_0(\omega_g)}}{\sum_{a' \in \mathcal{A}} e^{-\beta\psi_0(a')}}.$$

It is convenient to denote the denominator, *i.e.,* **partition function**, as $\mathcal{Z}(\beta) = \sum_{\omega' \in \Sigma} e^{-\beta\psi(\omega')}$.

The measures $\mu_\beta(\omega)$ for each singular configuration are minute, even if ω minimises energy. Note however, that for small temperature (large β), configurations with minimal energies are extremely more likely to occur than those with large energies. For high temperature (small β), this relative difference is much smaller. As argued by Boltzmann, see the Ehrenfest paper [11], the vast majority of configurations (measure by μ_β) has the property that if you count proportions at which states $a \in \mathcal{A}$ occur, *i.e.,*

$$\pi_\omega(a) = \frac{1}{n} \#\{g \in G \ : \ \omega_g = a\}$$

you find that π_ω is extremely close to q_β. So without interactions, the effect of many particles averages out to q_β.

We can quantify "large majority" using distance d_β of (67). Write

$$U_{\beta,r} = \{\omega \in \Sigma \ : \ d_\beta(\pi_\omega) < r\}$$

as the collection of configurations whose emperical distributions π_ω (*i.e.,* frequencies of particles taking the respective states in $\mathcal{A}$) are r-close to q_β.

Theorem 35. *For $0 < r < H(q_\beta) - \beta \int \psi_0 \ dq_\beta$, we have*

$$\lim_{n\to\infty} \frac{1}{n} \log \mu_\beta(\Sigma \setminus U_{\beta,r}) = -r,$$

so $\mu_\beta(\Sigma \setminus U_{\beta,r}) \sim e^{-nr}$ as $n = \#G$ grows large.

Proof. It is an exercise to check that $H(\mu_\beta) = nH(q_\beta)$. Next, for some configuration $\omega \in \Sigma$, we have

$$\begin{aligned}
\log \mu_\beta(\omega) &= -\beta \sum_{g\in G} \psi_0(\omega_g) - \log \mathcal{Z}(\beta) \qquad \text{rewrite } \mathcal{Z}(\beta) \text{ by Corollary 4} \\
&= -\beta n \int \psi_0 \ d\pi_\omega - \left(H(\mu_\beta) - \beta \int \psi \ d\mu_\beta \right) \\
&= -\beta n \int \psi_0 \ d\pi_\omega - n \left(H(q_\beta) - \beta \int \psi_0 \ dq_\beta \right) \\
&= -n \left((H(q_\beta) - \beta \int \psi_0 \ dq_\beta) - (H(\pi_\omega) - \beta \int \psi_0 \ d\pi_\omega) \right) - nH(\pi_\omega) \\
&= n d_\beta(\pi_\omega) - nH(\pi_\omega).
\end{aligned}$$

Every π_ω represents a way to choose $n = \#G$ times from $N = \#\mathcal{A}$ boxes. The order of choosing is not important, only how many are drawn from each box. This can be indicated by a non-negative integer vector $v = (v_a)_{a\in\mathcal{A}}$ where $\sum_{a\in\mathcal{A}} v_a = n$. In fact, $\frac{v}{n}$ indicates the same probability distribution on $\mathcal{A}$ as π_ω. We can compute

$$M(v) := \#\{\omega \in \Sigma \ : \ \pi_\omega \text{ leads to } v\} = \frac{n!}{\prod_{a\in\mathcal{A}} v_a!}.$$

Stirling's formula gives $n! \sim \sqrt{2\pi n}\ n^n e^{-n}$, neglecting an error factor that tends to 1 as $n \to \infty$. Thus

$$M(v) \sim \frac{\sqrt{2\pi n}\ n^n e^{-n}}{\prod_{a\in\mathcal{A}} \sqrt{2\pi v_a}\ v_a^{v_a} e^{-v_a}} \sim \sqrt{\frac{2\pi n}{\prod_{a\in\mathcal{A}} 2\pi v_a}} \prod_{a\in\mathcal{A}} \left(\frac{v_a}{n}\right)^{-v_a},$$

and

$$\log M(v) \sim \frac{1}{2} \log \frac{2\pi n}{\prod_{a\in\mathcal{A}} 2\pi v_a} + n \sum_{a\in\mathcal{A}} -\frac{v_a}{n} \log \frac{v_a}{n}.$$

Note that $\frac{v_a}{n} = \pi_\omega(x = a)$, so the dominating term in $\log M(v)$ is just $nH(\pi_\omega)$! The remaining terms, including the one we neglected in our version of Stirling's formula, are $O(\log n)$.

Therefore

$$\begin{aligned}
\frac{1}{n}\log\mu_\beta(\Sigma\setminus U_{\beta,r}) &= \frac{1}{n}\log\sum_{\omega\in\Sigma\setminus U_{\beta,r}}\mu_\beta(\omega) = \frac{1}{n}\log\sum_{\omega\in\Sigma\setminus U_{\beta,r}} e^{-nd_\beta(\pi_\omega)-nH(\pi_\omega)}\\
&\leqslant \frac{1}{n}\log\sum_{v=(v_a)_{a\in\mathcal{A}}} M(v)\cdot e^{-nr-nH(\frac{v}{n})}\\
&= \frac{1}{n}\log\sum_{v=(v_a)_{a\in\mathcal{A}}} e^{nH(\frac{v}{n})+O(\log n)}\cdot e^{-nr-nH(\frac{v}{n})}\\
&= \frac{1}{n}\log\sum_{v=(v_a)_{a\in\mathcal{A}}} e^{-nr+O(\log n)} \leqslant \frac{1}{n}\log n^N e^{-nr+O(\log n)} \to -r
\end{aligned}$$

as $n\to\infty$, where we used in the last line that there are no more than n^N ways of choosing non-negative integer vectors $v=(v_a)_{a\in\mathcal{A}}$ with $\sum_{a\in\mathcal{A}} v_a = n$.

Now for the lower bound, take $r'>r$. For sufficiently large n, we can find some vector $v=(v_a)_{a\in\mathcal{A}}$ such that $r<d_\beta(\frac{v}{n})<r'$. Therefore

$$\begin{aligned}
\frac{1}{n}\log\mu_\beta(\Sigma\setminus U_{\beta,r}) &\geqslant \frac{1}{n}\log\sum_{\omega,\pi_\omega=\frac{v}{n}}\mu_\beta(\omega) = \frac{1}{n}\log\sum_{\omega,\pi_\omega=\frac{v}{n}} e^{-nd_\beta(\pi_\omega)-nH(\pi_\omega)}\\
&\geqslant \frac{1}{n}\log\Big(M(v)\cdot e^{-nr'-nH(\frac{v}{n})}\Big)\\
&\geqslant \frac{1}{n}\Big(nH(\frac{v}{n})+O(\log n)-nr'-nH(\frac{v}{n})\Big)\to -r'
\end{aligned}$$

as $n\to\infty$. Since $r'>r$ is arbitrary, $\lim_n \frac{1}{n}\log\mu_\beta(\Sigma\setminus U_{\beta,r}) = -r$ as claimed. □

References

[1] R. L. Adler, A. G. Konheim, M. H. McAndrew, *Topological entropy,* Trans. Amer. Math. Soc. **114** (1965), 309-319.

[2] V. I. Arnol'd, A. Avez, *Ergodic Problems in Classical Mechanics,* New York: Benjamin (1968).

[3] V. I. Arnold, *Mathematical Methods of Classical Mechanics,* Springer-Verlag (1989).

[4] R. Bowen, *Equilibrium states and the ergodic theory of Anosov diffeomorphisms,* Lect. Math. **470** Spring 1974 and second revised edition 2008.

[5] A. Boyarski, P. Góra, *Laws of chaos,* Birkhäuser, Basel New York (1997)

[6] K. M. Brucks, H. Bruin, *Topics from one-dimensional dynamics,* London Mathematical Society, Student Texts **62** Cambridge University Press 2004.

[7] H Bruin, J. Hawkins, *Rigidity of smooth one-sided Bernoulli endomorphisms,* New York Journ. Maths. **15** (2009), 451–483.

[8] H. Bruin, M. Todd, *Transience and thermodynamic formalism for infinitely branched interval maps,* J. London Math. Soc. **86** 2012, 171–194.

[9] N. G. de Bruyn, *Asymptotic methods in analysis,* Dover Publ. New York 1981.

[10] E. I. Dinaburg, *The relation between topological entropy and metric enropy,* Dokl. Akad. Nauk SSSR., **190** (1970), 19–22, Soviet Math. Dok. **11** (1970), 13-16.

[11] P. and T. Ehrenfest, *Begriffliche Grundlagen der statistischen Auffassung in der Mechanik,* Enzy. d. Mathem. Wiss. IV, 2, II (1912), 3-90. English translation: *The conceptual foundations of the statistical approach in mechanics,* Cornell Univ. Press (1959).

[12] F. Hofbauer, *Examples for the nonuniqueness of the equilibrium state,* Trans. Amer. Math. Soc. **228** (1977) 223–241

[13] M. Keane, M. Smorodinsky, *Bernoulli schemes of the same entropy are finitarily isomorphic,* Ann. of Math. (**109 (1979)**, 397–406.

[14] G. Keller, *Equilibrium states in Ergodic Theory,* London Math. Society, Student Texts **42** Cambridge Univ. Press, Cambridge 1998.

[15] W. Krieger, *On entropy and generators of measure-preserving transformations,* Trans. Amer. Math. Soc. **149** (1970), 453–464 and Erraum **168** (1972), 519.

[16] A. Lasota, J. Yorke,, *On the existence of invariant measures for piecewise monotonic transformations,* Trans. A.M.S. **186** 481-488 (1973)

[17] F. Ledrappier, *Some properties of absolutely continuous invariant measures on an interval,* Erg. Th. and Dyn. Sys. **1** (1981), 77–94.

[18] A. Lopes, *The zeta function, non differentiability of pressure, and the critical exponent of transition, Adv. Math.* **101** (1993) 133–165.

[19] M. Misiurewicz and W. Szlenk, *Entropy of piecewise monotone mappings,* Studia Math. **67** (1980), no. 1, 45-63.

[20] K. Petersen, *Ergodic theory,* Cambridge University Press, 1983.

[21] D. Ornstein, *Ergodic theory, randomness and dynamical systems,* Yale Math. Monographs, **5** Yale Univ. New Haven, 1974.

[22] W. Parry, *Intrinsic Markov chains*, Trans. AMS. **112** (1965), 55-65.

[23] D. Terhesiu, *Improved mixing rates for infinite measure preserving transformations,* Ergodic Th. and Dyn. Sys. **35** (2015) 585–614.

[24] P. Walters, *An introduction to ergodic theory,* Springer Verlag (1982).