Information and entropy in dynamical systems. An introduction

François Ledrappier

Będlewo, June/July 2011

1. Information and conditional information
2. Mean entropy of a transformation.
3. Shannon-McMillan-Breiman Theorem. Local entropy.
4. Entropy of linear automorphisms of tori
5. Entropy, exponents and dimension for linear hyperbolic automorphisms.
6. Ergodic linear automorphisms
7. Linear endomorphisms of tori
8. General $C^{1+\alpha}$ diffeomorphism of a compact manifold. Pesin theory
9. Entropy and absolutely continuous conditional measures
10. Notes and comments

1.a. Information. Countable partitions

$(X, \mathcal{A}, m)$ is a Lebesgue probability space

$(X, \mathcal{A}, m)$ is measurably isomorphic to $([0, 1]$, Borel, Leb.)

$P = \{P_1, P_2, \cdots\}$ is a mod.0 partition of $(X, \mathcal{A}, m)$

$P_i \in \mathcal{A}, m(P_i \cap P_j) = 0$ for $i \neq j$ and $m(\cup_i P_i) = 1$.

For a.e. $x \in X$, $P(x) := P_{i(x)}$ s.t. $x \in P_{i(x)}$

Definition *Information Function* $I(P)$

$$I(P)(x) \ = \ -\ln m(P(x)).$$

$Q = \{Q_1, Q_2, \cdots\}$ another partition. The *conditional measure with respect to* Q m_x^Q is given, for $A \in \mathcal{A}$, by

$$m_x^Q(A) = \frac{m(Q(x) \cap A)}{m(Q(x))}.$$

Definition *Conditional Information* $I(P|Q)$

$$I(P|Q)(x) = -\ln m_x^Q(P(x)).$$

$P \vee Q$ partition s.t. $P \vee Q(x) = P(x) \cap Q(x)$:

$$I(P \vee Q)(x) = I(Q)(x) + I(P|Q)(x).$$

$I(P)(x)$ measures the information brought by learning in which element of the partition P the point x is; $I(P|Q)(x)$ the same information if one knows already in which element of Q the point x lies. The addition formula is consistent with this image. Moreover (*exercise*):

$$I(P|Q)(x) = 0 \text{ a.e. } \iff P \prec Q$$

($P \prec Q$ means that Q refines P: no new information)

$$I(P|Q) = I(P) \text{ a.e. } \iff P \perp Q$$

($P \perp Q$ means that P and Q are independent: previous information is irrelevant).

1.b Conditional measures. General case

$\mathcal{B}$ a subσ-algebra of $\mathcal{A}$. There exist conditional probability measures $m_x^{\mathcal{B}}$ on $\mathcal{A}$ satisfying:

- $A \mapsto m_x^{\mathcal{B}}(A)$ is a probability measure for all x,
- $x \mapsto m_x^{\mathcal{B}}(A)$ is $\mathcal{B}$-measurable, and
- $\forall A \in \mathcal{A}, B \in \mathcal{B},\ \int_B m_x^{\mathcal{B}}(A) dm = m(A \cap B)$.

Moreover, this family is essentially unique.

Proof (exercise) $\mathcal{A}_0$ countable algebra generating $\mathcal{A}$, $\mathbb{E}^{\mathcal{B}}1_A(x)$ the conditional expectations with respect to $\mathcal{B}$. For a.e. x, $x \mapsto \mathbb{E}^{\mathcal{B}}1_A(x)$ satisfies Kolmogorov conditions and extends from $\mathcal{A}_0$ into in a probability measure on $\mathcal{A}$.

Remark: $\xi_{\mathcal{B}}$ s.t. $x \overset{\xi_{\mathcal{B}}}{\sim} y \iff m_x^{\mathcal{B}} = m_y^{\mathcal{B}}$ defines a partition of X which is *countably separated by $\mathcal{A}$ measurable sets.*

We call such a partition a *measurable partition.*

Conversely, let ξ be a measurable partition, define $\mathcal{B}_\xi$ as the σ-algebra of $\mathcal{A}$-measurable sets which are unions of elements of ξ. Then, on a set of full measure,

$$\xi_{\mathcal{B}_\xi} = \xi,$$

$m_x^\xi := m_x^{\mathcal{B}_\xi}$ has support $\xi(x)$ and

for all $B \in \mathcal{B}_\xi$, $m_x^\xi(B) = 0$ or 1.

Examples

1. Let ε be the partition of X into points. $\mathcal{B}_\varepsilon = \mathcal{A}, m_x^{\mathcal{A}} = \delta_x$ and $\xi_{\mathcal{A}} = \varepsilon$.

2. $X = [0,1] \times [0,1]$, any probability measure m. η partition into vertical lines. $\mathcal{B}_\eta$ is the σ-algebra of vertical sets. $m_x^{\mathcal{B}_\eta}$ is the conditional measure carried by $\eta(x)$, and $\xi_{\mathcal{B}_\eta} = \eta$.

3. An example of a non-measurable partition. Let η be the partition into orbits of a measure preserving transformation. Then $\mathcal{B}_\eta$ is the σ-algebra $\mathcal{I}$ of invariant measurable sets. Then, the conditional measures $m_x^{\mathcal{I}}$ are invariant ergodic (*See Omri Sarig's lectures:* One uses the Pointwise Ergodic Theorem). In general $\xi_{\mathcal{I}} \neq \eta$. For example, $\xi_{\mathcal{I}}$ is trivial if the transformation is ergodic.

Martingale Theorems[see Chapter 10 b.]

I. Assume $\xi_n \prec \xi_{n+1}$ (ξ_{n+1} refines ξ_n), then $\mathcal{B}_{\xi_n} \subset \mathcal{B}_{\xi_{n+1}}$. Let $\mathcal{B}_\infty := \sigma(\cup_n \mathcal{B}_n)$.

Then, as $n \to \infty$, for a.e. x,

$$m_x^{\xi_n} \to m_x^{\mathcal{B}_\infty}$$

and

$$\xi_{\mathcal{B}_\infty}(x) = \cap_n \xi_n(x).$$

II. Assume $\xi_n \succ \xi_{n+1}$ (ξ_n refines ξ_{n+1}), then $\mathcal{B}_{\xi_{n+1}} \subset \mathcal{B}_{\xi_n}$. Let $\mathcal{B}_\infty := \cap_n \mathcal{B}_n$.

Then, as $n \to \infty$, for a.e. x,

$$m_x^{\xi_n} \to m_x^{\mathcal{B}_\infty}$$

but, in general,

$$\xi_{\mathcal{B}_\infty}(x) \neq \cup_n \xi_n(x).$$

We denote $\wedge_n \xi_n := \xi_{\mathcal{B}_\infty}$.

Example of II: Kolmogorov 0-1 law:

$X = [0,1]^{\mathbb{N}}, m = \otimes_{\mathbb{N}}\{1/2, 1/2\}$,

$\xi_n(x) = \{y : y_p = x_p \text{ for } p \geq n\}$.

$m_x^{\xi_n} = \otimes_{0,1,\cdots,n-1}\{1/2, 1/2\} \otimes \delta_{x_n} \otimes \cdots \otimes \delta_{x_{n+p}} \otimes \cdots$

As $n \to \infty$, $m_x^{\xi_n} \to m$, $\wedge_n \xi_n = \xi_{\cap_n \mathcal{B}_{\xi_n}}$ is trivial but

$$\cup_n \xi_n(x) = \{y : \exists n, \text{ s.t. } y_p = x_p \text{ for } p \geq n\}.$$

1.c Conditional information. General case

P a countable partition, ξ a measurable partition.

Definition *Conditional Information* $I(P|\xi)$

$$I(P|\xi)(x) \;=\; -\ln m_x^{\xi}(P(x)).$$

We have again (*exercise*):

$$I(P|\xi)(x) = 0 \text{ a.e. } \iff P \prec \xi,$$

$$I(P|\xi) = I(P) \text{ a.e. } \iff P \perp \xi \quad \text{and}$$

$$I(P \vee Q|\xi)(x) = I(Q|\xi)(x) + I(P|Q \vee \xi)(x),$$

$$I(P \vee Q|\xi)(x) = I(Q|\xi)(x) + I(P|Q \vee \xi)(x),$$

where $Q \vee \xi$ is the measurable partition obtained by cutting each element of ξ by the Q_j such that $m_x^\xi(Q_j) > 0$. Verify then that

$$m_x^{Q\vee\xi}(A) = \frac{m_x^\xi(A \cap Q(x))}{m_x^\xi(Q(x))}. \qquad (1)$$

In particular, $I(P|\xi)(x) = I(Q|\xi)(x)$ a.e. iff

$$P \vee \xi = Q \vee \xi.$$

If $\xi_n \prec \xi_{n+1}$ and $\xi_\infty(x) := \cap_n \xi_n(x)$,

or if $\xi_n \succ \xi_{n+1}$ and $\xi_\infty := \wedge_n \xi_n$,

$$I(P|\xi_n) \to I(P|\xi_\infty) \text{ a.e..}$$

Theorem *Assume* $\xi_n \prec \xi_{n+1}$ *and* $\int I(P|\xi_1) < +\infty$. *Then,*

$$\int \sup_n I(P|\xi_n) < +\infty.$$

Proof By definition, for all $t > 0$:

$$m_x^{\xi_1}(\{y : \sup_n I(P|\xi_n)(y) > t\}) = \sum_i m_x^{\xi_1}(P_i \cap \{y : \inf_n m_x^{\xi_n}(P_i) < e^{-t}\})$$

Claim: $m_x^{\xi_1}(P_i \cap \{y : \inf_n m_x^{\xi_n}(P_i) < e^{-t}\}) \leq \min\{m_x^{\xi_1}(P_i), e^{-t}\}$.

Using the claim and $\int_0^{+\infty} \min\{a, e^{-t}\} dt = a - a \ln a$, we get:

$$\begin{aligned} \int \sup_n I(P|\xi_n) dm_x^{\xi_1} &\leq \sum_i m_x^{\xi_1}(P_i) - \sum_i m_x^{\xi_1}(P_i) \ln m_x^{\xi_1}(P_i) \\ &\leq 1 + \int I(P|\xi_1) dm_x^{\xi_1} \end{aligned}$$

and the conclusion follows by integrating with respect to m.

Proof of the claim:

$m_x^{\xi_1}(P_i \cap \{y : \inf_n m_x^{\xi_n}(P_i) < e^{-t}\}) \leq m_x^{\xi_1}(P_i)$ is clear.

Moreover, let $\nu(y)$ be the smallest integer n such that $m_y^{\xi_n}(P_i) < e^{-t}$.

Since $\{y : \nu(y) = n\} \in \mathcal{B}_{\xi_n}$, we may write:

$$\begin{aligned} m_x^{\xi_1}(P_i \cap \{y : \nu(y) = n\}) &= \int_{\{y:\nu(y)=n\}} m_y^{\xi_n}(P_i) dm_x^{\xi_1}(y) \\ &\leq e^{-t} m_x^{\xi_1}(\{y : \nu(y) = n\}). \end{aligned}$$

Therefore,

$$\begin{aligned} m_x^{\xi_1}(P_i \cap \{y : \inf_n m_x^{\xi_n}(P_i) < e^{-t}\}) &= \sum_n m_x^{\xi_1}(P_i \cap \{y : \nu(y) = n\}) \\ &\leq e^{-t} \sum_n m_x^{\xi_1}(\{y : \nu(y) = n\}) \leq e^{-t}. \end{aligned}$$

2.a Entropy

Definition *The entropy* $H(P)$ *of a countable partition* P *is the integral of the information function:*

$$H(P) = \int I(P)(x)dm(x) = -\sum_i m(P_i) \ln m(P_i).$$

$0 \leq H(P) \leq +\infty$. $H(P) = 0 \iff P = \{X\}$.

Definition ξ *a measurable partition. The conditional entropy of the partition* P *with respect to* ξ *is given by:*

$$H(P|\xi) = \int I(P|\xi)dm(x) = \int H_{m_x^\xi}(P)dm(x).$$

$0 \leq H(P|\xi) \leq +\infty$. $H(P|\xi) = 0 \iff P \prec \xi$.

$$H(P \vee Q|\xi) = H(Q|\xi) + H(P|Q \vee \xi).$$

Proposition 1) *If* $\xi_1 \prec \xi_2$, $H(P|\xi_2) \leq H(P|\xi_1)$.

2) *If Card* $(P) = K$, *then* $H(P) \leq \ln K$, *with equality iff* $m(P_i) = 1/K$ *for all* i.

Proof: Both statements follow from the strict convexity of the function $t \mapsto \phi(t) = t \ln t$.

For 2), write that

$$-\frac{1}{K}\ln K = \phi(\frac{1}{K}) = \phi\left(\frac{1}{K}\sum_{i=1}^{K} m(P_i)\right) \leq \frac{1}{K}\sum_{i=1}^{K}\phi\left(m(P_i)\right) = -\frac{1}{K}H(P)$$

and we have equality only if all $m(P_i)$ coincide.

For 1), since $\xi_1 \prec \xi_2$, $m_x^{\xi_1}(P_i) = \int m_y^{\xi_2}(P_i) dm_x^{\xi_1}(dy)$ and therefore $\phi\left(m_x^{\xi_1}(P_i)\right) \leq \int \phi\left(m_y^{\xi_2}(P_i)\right) dm_x^{\xi_1}(dy)$. Summing on i yields:

$$H_{m_x^{\xi_1}}(P) \geq \int H_{m_y^{\xi_2}}(P) dm_x^{\xi_1}(dy).$$

1) follows by integrating on x. $\square$

We have equality in 1) iff $m_y^{\xi_2}(P_i)$ is constant $m_x^{\xi_1}$-a.e. (P is said to be *conditionally independent* of $\mathcal{B}_{\xi_2}$ relatively to $\mathcal{B}_{\xi_1}$).

In particular, $H(P|\xi) \leq H(P)$ with equality iff P is independent of $\mathcal{B}_\xi$.

Proposition 1) *If* $\xi_n \prec \xi_{n+1}$, $\xi_\infty(x) := \cap_n \xi_n(x)$ **and** $H(P|\xi_1) < +\infty$, *then*

$$H(P|\xi_1) \searrow H(P|\xi_\infty).$$

2) *If* $\xi_n \succ \xi_{n+1}$ *and* $\xi_\infty := \wedge_n \xi_n$, *then*

$$H(P|\xi_1) \nearrow H(P|\xi_\infty).$$

Proof 1) We know that $I(P|\xi_n)$ converges towards $I(P|\xi_\infty)$ and by Theorem page 15, that $\int \sup_n I(P|\xi_n) < +\infty$. The convergence follows by the Dominated Convergence Theorem. The limit is non-increasing by the previous Proposition.

2) As just seen, $H_{m_x^{\xi_n}}(P) \le \int H_{m_y^{\xi_{n+1}}}(P) dm_x^{\xi_n}(dy)$. This means that $x \mapsto H_{m_x^{\xi_n}}(P)$ is a supermartingale with respect to the decreasing family of σ-algebras $\mathcal{B}_{\xi_n}$. By Doob's Theorem (see Chapter 10 b), it converges m-a.e. and in L^1 if it makes sense. In particular, $\int H_{m_x^{\xi_n}}(P) dm(x)$ converges towards $\int H_{m_x^{\xi_\infty}}(P) dm(x)$, even if the limit is infinite. $\square$

Complements 1) If $H(Q|\xi) < +\infty$, there exists $P, H(P) < +\infty$ such that $Q \vee \xi = P \vee \xi$. Moreover, $H(P) = O(H(Q|\xi))$.

Sketch of a proof: α) One can choose P_i such that $P \vee \xi = Q \vee \xi$ and $m_x^\xi(P_i) \geq m_x^\xi(P_{i+1})$. Then, $m_x^\xi(P_i) \geq 1/i$ and $\int_X \left(\sum_i m_x^\xi(P_i) \ln i\right) dm(x) \leq H(P|\xi) = H(Q|\xi)$. Then, $\sum_i m(P_i) \ln i \leq H(Q|\xi)$.

β) Set, for $a > 1$, $q_i = \frac{i^{-a}}{\zeta(a)}$. $\sum_i q_i = 1$ and: $\sum_i m(P_i) \ln \frac{m(P_i)}{q_i} = \sum_i q_i \phi\left(\frac{m(P_i)}{q_i}\right) \geq \phi\left(\sum_i q_i \frac{m(P_i)}{q_i}\right) = 0$, i.e. $H(P) \leq -\sum_i m(P_i) \ln q_i \leq \ln \zeta(a) + a \sum_i m(P_i) \ln i \leq \ln \zeta(a) + aH(Q|\xi) < +\infty$.

γ) Use $\ln \zeta(a) \leq \frac{1}{a-1}$ and optimize at $a = 1 + \frac{1}{\sqrt{H(Q|\xi)}}$.

2) Define, for ξ, η measurable partitions

$$H(\eta|\xi) := \sup_{Q : Q \text{countable} Q \prec \eta} H(Q|\xi).$$

Then, $\Delta(\xi, \eta) := H(\eta|\xi) + H(\xi|\eta)$ is a "distance" (it might be infinite) for which the space of measurable partitions is complete.

2.b Mean entropy

Let $T : X \to X$ be a measurable, measure preserving transformation, P a partition; set $T^{-1}P$ for the partition into the sets

$$T^{-1}P_i = \{y : Ty \in P_i\}.$$

Then, $H(T^{-1}P) = H(P)$. Denote

$$P_m^n := \vee_{m \le j < n} T^{-j} P, \;\; P_m^\infty := \vee_{m \le j} T^{-j} P.$$

Then, $H(P_m^n) \leq H(P_0^m)+H(P_0^n)$. By Fekete's Lemma:

$$\frac{1}{n}H(P_0^n) \to \inf_n \frac{1}{n}H(P_0^n) =: h(P,T).$$

$h(P,T)$ is called the *mean entropy* of P. We have:

$$h(P,T) = H(P|P_1^\infty).$$

Proof: Write $H(P_0^n) = H(T^{-n}P) + H(T^{-(n-1)}P|T^{-n}P) + \cdots + H(P|P_1^n)$. The general term in the sum is $H(P|P_1^j)$ which converges towards $H(P|P_1^\infty)$. The average has the same limit. $\square$

Properties of the mean entropy

$$h(P_0^k, T) = h(P, T); \quad h(P_0^k, T^k) = kh(P, T).$$

$$|h(P, T) - h(Q, T)| \leq \Delta(P, Q).$$

Proof: Write $H(P_0^n \vee Q_0^n) = H(P_0^n) + H(Q_0^n | P_0^n) = H(Q_0^n) + H(P_0^n | Q_0^n)$. It follows that

$$\begin{aligned} |H(P_0^n) - H(Q_0^n)| &\leq H(P_0^n | Q_0^n) + H(Q_0^n | P_0^n) \\ &\leq \sum_{i=1}^{n} H(T^{-i}P | Q_0^n) + H(T^{-i}Q | P_0^n) \leq n\Delta(P, Q). \end{aligned} \qquad \square$$

In particular, $Q \prec P_0^\infty \Rightarrow h(Q, T) \leq h(P, T)$.

Because there are partitions $Q_k \prec P_0^k$ with $\Delta(Q, Q_k) \to 0$.

Entropy of a transformation

Definition $h_m(T) := \sup_{P; P \text{finite}} h(P,T)$.

Example 1: Rotations ($X = \mathbb{R}/\mathbb{Z}$, Borel, Leb. ,$Tx = x + \alpha$ mod1). Then $h(T) = 0$. *Proof:* Let $P_k = \{\cdots, [\frac{j}{k}, \frac{j+1}{k}], \cdots\}$. Since $(P_k)_0^n$ has less than kn elements, $h(P_k, T) = 0$. On the other hand, for any Q finite, $H(Q|P_k) \to 0$. $\square$

Example 2: Bernoulli shifts (A, p) a finite or countable probability space, $H(p) < +\infty$.

$X = A^{\mathbb{N}}, \mathcal{A}, m = \otimes_n p_n, Tx = y$ with $y_n = x_{n+1}$. Then,

$$h_m(T) = H(p)$$

Proof: P defined by x_0. $H(P_0^n) = nH(p)$ and $P_0^\infty = \mathcal{A}$. $\square$

Remark the formula also holds if $H(p) = +\infty$.

2.c Relative entropy. Pinsker formula

We assume that the transformation T is invertible. For $m, n \in \mathbb{Z}, m < n$, define P_m^n, P_m^∞, $P_{-\infty}^n, P_{-\infty}^\infty$. We have:

$$h(P,T) = H(P|P_1^\infty) = H(P|P_{-\infty}^0) = h(P,T^{-1}),$$

since all these numbers are $\lim_n \frac{1}{n} H(P_0^n)$.

In particular, $h_m(T) = h_m(T^{-1})$.

Exercise: For $k \in \mathbb{Z}$, $h_m(T^k) = |k| h_m(T)$.

The measurable partition ξ is called invariant if the σ-algebra $\mathcal{B}_\xi$ is invariant. Then:

$$\xi(Tx) = T\xi(x); T_*(m_x^\xi) = m_{Tx}^\xi$$

$$I(T^{-1}P|\xi)(x) = I(P|\xi)(Tx); H(T^{-1}P|\xi) = H(P|\xi).$$

The *relative mean entropy* is given by:

$$\begin{aligned} h(P,T|\xi) &= \lim_n \frac{1}{n} H(P_0^n|\xi) = \inf_n \frac{1}{n} H(P_0^n|\xi) \\ &= H(P|P_1^\infty \vee \xi) = H(P|P_{-\infty}^0 \vee \xi). \end{aligned}$$

Proofs and properties are the same as the ones for the absolute mean entropy.

Proposition [Pinsker Formula]

$$h(P \vee Q, T|\xi) = h(Q, T|\xi) + h(P, T|\xi \vee Q_{-\infty}^{\infty}).$$

Proof : Write

$$\begin{aligned} H((P \vee Q)_0^n|\xi) &= H(Q_0^n|\xi) + H(P_0^n|\xi \vee Q_0^n) \\ &= H(Q_0^n|\xi) + \sum_j H(P|P_{-j}^0 \vee \xi \vee Q_{-j}^{n-j}). \end{aligned}$$

Most of the terms in the sum are closer and closer to $H(P|P_{-\infty}^0 \vee \xi \vee Q_{-\infty}^{\infty}) = h(P, T|\xi)$. $\square$

Corollary The family of sets $A \in \mathcal{A}$ s.t. $h(\{A, A^c\}, T) = 0$ form an invariant σ-algebra, the Pinsker σ-algebra.

Let π be the associated measurable partition, $h(P, T|\pi) = h(P, T)$ for all $P, H(P) < \infty$.

Proof: Exercise.

Proposition $P, H(P) < +\infty, P_{-\infty}^{\infty} = \varepsilon$. *Then,*

$$\pi = \wedge_{n\in\mathbb{Z}} P_{-\infty}^{n} = \wedge_{n\in\mathbb{Z}} P_{n}^{\infty}.$$

Proof: 1) Assume $A \in \mathcal{B}_{\wedge_n \xi_n}$ and denote $P_A := \{A, A^c\}$. Then,

$h((P \vee P_A), T) = h(P_A, T) + h(P, T|(P_A)_{-\infty}^{\infty}) = h(P, T) + h(P_A, T|P_{-\infty}^{\infty})$.

$h(P_A, T|P_{-\infty}^{\infty}) = 0$ since $P_A \prec P_{-\infty}^{\infty}$; $h(P, T|(P_A)_{-\infty}^{\infty}) = h(P, T)$ since $(P_A)_{-\infty}^{\infty} \prec P_{-\infty}^{0}$. Therefore, $h(P_A, T) = 0$.

2) Conversely, assume that $h(P_A, T) = 0$. Then, as above,

$H(P_0^n|P_{-\infty}^0 \vee (P_A)_{-\infty}^{\infty}) = H(P_0^n|P_{-\infty}^0)$ for all $n > 0$. Also

$H(P_0^n|P_{-\infty}^0 \vee P_A) = H(P_0^n|P_{-\infty}^0)$, since $P_A \prec (P_A)_{-\infty}^{\infty}$.

This implies: $H(P_A|P_{-\infty}^0) = H(P_A|P_{-\infty}^n) = H(P_A|P_{-\infty}^{\infty}) = 0$.

So, $P_A \prec P_{-\infty}^0$. Thus, $P_A \prec P_{-\infty}^n \forall n \in \mathbb{Z}$, i.e. $P_A \prec \wedge_n P_{-\infty}^n$. $\square$

ξ is called increasing if $\xi \prec T^{-1}\xi$, i.e. $\xi(Tx) \subset T(\xi(x))$. Then,

$$h(\xi, T) := H(T^{-1}\xi|\xi) \leq h(T).$$

Proof: $Q \prec T^{-1}\xi$, Q finite. Then, $H(Q|\xi) \leq H(Q|Q^0_{-\infty}) = h(Q,T) \leq h(T)$. $\square$

Exercise: ξ increasing. Then,

$$h(\xi, T) = \sup_{Q \text{ finite}} \lim_n \frac{1}{n} H(Q_0^n|\xi).$$

Remark. One may have ξ increasing, $T^{-n}\xi \nearrow \varepsilon$ and $h(\xi, T) < h(T)$. See an example on page 64.

3.a Shannon-McMillan-Breiman Theorem

Theorem $(X, \mathcal{A}, m, T)$ *ergodic,* P *countable partition with* $H(P) < +\infty$. *Then,*

$$\frac{1}{n} I(P_0^n)(x) \overset{a.s., L^1}{\longrightarrow} h(P, T).$$

Proof: Write

$$I(P_0^n)(x) = \sum_{j=0}^{n-1} I(T^{-j}P | P_{j+1}^{n+1})(x) = \sum_{j=0}^{n-1} I(P | P_1^{n-j+1})(T^j x).$$

The Theorem follows since $I(P|P_1^{n-j+1})(x) \to I(P|P_1^{\infty})(x)$ a.s. and in L^1, $\sup_n I(P|P_1^n) \in L^1$ by Theorem page 15, and:

Claim (exercise) *If* $f_k \in L^1, f_k \overset{a.s., L^1}{\longrightarrow} f_\infty$ *and* $\sup_k |f_k| \in L^1$*, then*

$$\frac{1}{n} \sum_{k=0}^{n-1} f_k \circ T^k \overset{a.s., L^1}{\longrightarrow} \int f_\infty \ \ as \ \ k \to \infty. \qquad \square$$

SMB Theorem, variant 1 *$(X, \mathcal{A}, m, T)$ ergodic, P countable partition with $H(P) < +\infty$. Then,*

$$\frac{1}{2n} I(P_{-n}^{n})(x) \overset{a.s., L^1}{\longrightarrow} h(P, T).$$

SMB Theorem, variant 2 *$(X, \mathcal{A}, m, T)$ ergodic, ξ increasing measurable partition. Then,*

$$\frac{1}{n} I(T^{-n}\xi | \xi)(x) \overset{a.s., L^1}{\longrightarrow} h(\xi, T).$$

SMB Theorem, variant 3 $(X, \mathcal{A}, m, T)$ *ergodic,* ξ *invariant measurable partition,* P *countable partition with* $H(P|\xi) < +\infty$. *Then,*

$$\frac{1}{n} I(P_0^n|\xi)(x) \stackrel{a.s., L^1}{\longrightarrow} h(P, T|\xi).$$

SMB Theorem, variant 4 $(X, \mathcal{A}, m, T)$ *non necessarily ergodic,* P *countable partition with* $H(P) < +\infty$. *Then,*

$$\frac{1}{n} I(P_0^n)(x) \stackrel{a.s., L^1}{\longrightarrow} h_{m_x}(P, T),$$

where m_x *is the ergodic decomposition of* m.

Proof: Variant 1, 2 and 3 are proven by the same argument as SMB Theorem.

For variant 4, recall from page 21 that

$$H_{m_x^{P_{-\infty}^0}}(P) \geq \int H_{m_y^{P_{-\infty}^0} \vee \eta}(P) dm_x^{P_{-\infty}^0}(dy),$$

where η is the measurable partition associated to the σ-algebra of invariant sets. In other words:

$$I_m(P|P_{-\infty}^0)(y) \geq \int I_{m_y^\eta}(P|P_{-\infty}^0)(z) dm_y^{P_{-\infty}^0}(dz).$$

The proof of SMB yields, in the nonergodic case,

$$\frac{1}{n} I(P_0^n)(x) \xrightarrow{a.s., L^1} \int I_m(P|P_{-\infty}^0)(y) dm_x^\eta(dy)$$

(recall that m_x^η is the decomposition of m into ergodic components, cf. page 9). Integrating the above inequality in y, we see that

$$\liminf_n \frac{1}{n} I(P_0^n)(x) \geq \int I_{m_y^\eta}(P|P_{-\infty}^0)(y) dm_x^\eta(dy) = h_{m_x^\eta}(P, T).$$

$$\liminf_n \frac{1}{n} I(P_0^n)(x) \geq \int I_{m_y^\eta}(P|P_{-\infty}^0)(y) dm_x^\eta(dy) = h_{m_x^\eta}(P,T).$$

The integral of the first term goes to $h(P,T)$. The integral of the last term is $h(P,T|\eta)$. The Theorem follows from:

Claim η the measurable partition associated to the invariant sets. Then, $h(P,T) = h(P,T|\eta)$.

Clear for $Q \prec \eta$, Q finite, and

$$h(P,T|\eta) = H(P|P_{-\infty}^0 \vee \eta) = \inf_{Q;Q\prec\eta} H(P|P_{-\infty}^0 \vee Q). \quad \square$$

3.b Local entropy

X is a metric compact space, $\mathcal{A}$ the Borel σ-algebra, T a homeomorphism of X and m a T-invariant probability measure (which exists by a fixed point Theorem). For $\varepsilon > 0, n \in \mathbb{N}$, define the *Bowen dynamical ball* $B(x, n, \varepsilon)$ as:

$$B(x, n, \varepsilon) := \{y : d(T^k x, T^k y) < \varepsilon, \text{ for } 0 \leq k < n\}.$$

Theorem [Brin-Katok] (X, T, m) *as above. Assume m ergodic. Then,*

$$\begin{aligned} h_m(T) &= \lim_{\varepsilon \to 0} \liminf_n -\frac{1}{n} \ln m(B(x, n, \varepsilon)) \\ &= \lim_{\varepsilon \to 0} \limsup_n -\frac{1}{n} \ln m(B(x, n, \varepsilon)). \end{aligned}$$

Proof: 1) If the elements of P have diameter smaller than ε, then $P_0^n(x) \subset B(x, n, \varepsilon)$ and therefore:

$$\limsup_n -\frac{1}{n} \ln m(B(x, n, \varepsilon)) \leq \limsup_n -\frac{1}{n} I(P_0^n)(x) = h_m(P, T) \leq h(T).$$

2) Let $\delta, \rho > 0$, ρ to be chosen later. Choose Q a finite partition in closed sets and $\varepsilon > 0$ such that $m(\mathcal{U}_\varepsilon(\partial Q)) < \rho$ (∂Q is the union of the boundaries of the elements of Q, $\mathcal{U}_\varepsilon$ its ε-neighborhood) and $h(Q, T) \geq h(T) - \delta$ if $h(T)$ is finite, $> 1/\delta$ otherwise.

Consider the following sets E_n and F_n:

$$E_n := \{x : \forall p \geq n, \frac{1}{p}\sum_{j=0}^{n-1} \chi_{\mathcal{U}_e(\partial Q)}(T^j x) < 2\rho\}$$

$$F_n := \{x : \forall p \geq n, m(Q_0^p(x)) \leq e^{-p(h(Q,T)-\delta)}\}$$

By the Ergodic Theorem (for E_n) and the SMB Theorem (for F_n), for n large enough, $m(E_n \cap F_n) \geq 1-\delta$. For $x \in E_n \cap F_n, p \geq n$, $B(x,n,\varepsilon)$ is contained in less than $\binom{p}{2\rho p}$ elements of Q_0^p.

Either they all have measure $\leq e^{-p(h(Q,T)-4\delta)}$ and then $m(B(x,n,\varepsilon)) \leq e^{-p(h(Q,T)-4\delta)}\binom{p}{2\rho p} \leq e^{-p(h(Q,T)-5\delta)}$ by choosing ρ small enough, n large enough that $\binom{p}{2\rho p} \leq e^{p\delta}$ for all $p \geq n$,
or at least one of those elements of Q_0^p has measure $> e^{-p(h(Q,T)-4\delta)}$.

The total measure of points $x \in E_n \cap F_n$ belonging to those bad $B(y,n,\varepsilon)$ is at most

$\#\{\text{bad at. of } Q_0^n\} \times \#\{\text{at. of } Q_0^n \in F_n \text{ touching each bad at.}\} \times e^{-p(h(Q,T)-\delta)}$,

We have $\#\{\text{bad at. of } Q_0^n\} \leq e^{p(h(Q,T)-4\delta)}$ and

$$\#\{\text{at. of } Q_0^n \in F_n \text{ touching each bad at.}\} \leq \binom{p}{2\rho p} \leq e^{p\delta}$$

Altogether, the measure of points $x \in E_n \cap F_n$ belonging to bad $B(y, n, \varepsilon)$ is at most $e^{p(h(Q,T)-4\delta)} \times e^{p\delta} \times e^{-p(h(Q,T)-\delta)} \leq e^{-2p\delta}$.

By Borel Cantelli, a.e. point in $E_n \cap F_n$ eventually does not belong to the bad set, so that

$$\liminf_n -\frac{1}{n} \ln m(B(x,n,\varepsilon)) \geq h(Q,T) - 5\delta \quad \text{on} \quad E_n \cap F_n.$$

By our choice of Q, if $h(T) < \infty$, $\liminf_n -\frac{1}{n} \ln m(B(x,n,\varepsilon)) \geq h(T) - 6\delta$ and, if $h(T) = \infty$, $\liminf_n -\frac{1}{n} \ln m(B(x,n,\varepsilon)) \geq 1/\delta - 5\delta$.

The conclusion follows. $\square$

Exercise Write variants 1, 2 and 3 of Brin-Katok Theorem.

BK variant 4 (X, T, m) *as above. Let* $m_x := m_x^{\mathcal{I}}$ *be the ergodic decomposition of* m. *Then,*

$$\begin{aligned} h_{m_x}(T) &= \lim_{\varepsilon \to 0} \liminf_n -\frac{1}{n} \ln m(B(x, n, \varepsilon)) \\ &= \lim_{\varepsilon \to 0} \limsup_n -\frac{1}{n} \ln m(B(x, n, \varepsilon)). \end{aligned}$$

Proof: 1)By SMB variant 4, we get, using the same partition P,

$$\limsup_n -\frac{1}{n} \ln m(B(x, n, \varepsilon)) \le \limsup_n -\frac{1}{n} I(P_0^n)(x) = h_{m_x}(P, T) \le h_{m_x}(T).$$

2) Let A be a finite partition in invariant sets such that $A_0 = \{x : h_{m_x}(T) \geq 1/\delta\}$ and $x \mapsto h_{m_x}(T)$ varies by less than δ on each other element of A.

Set $h_i = \inf_{\{x \in A_i\}} h_{m_x}(T)$ if $i > 0$, $h_0 = 1/\delta$.

One can find a finite partition Q in closed sets and $\varepsilon > 0$ such that $m(\mathcal{U}_\varepsilon(\partial Q)) < \rho$ and $H(A|Q)$ is so small that outside a set G of measure δ, each element Q_j of Q is contained in one element $A_{i(j)} \cap G^c$. One can also impose that for $x \in A_i$, $h_{m_x}(Q,T) \geq h_i - \delta$. With such a partition Q, the proof of BK Theorem 2) gives, for $x \in A_i \cap G^c$:

$$\liminf_n -\frac{1}{n} \ln m(B(x,n,\varepsilon)) \geq h_i - 6\delta.$$

The conclusion follows. $\square$

Linear automorphisms of tori

In the next three sections, X is the d-dimensional torus $\mathbb{T}^d = \mathbb{R}^d/\mathbb{Z}^d$, $\mathcal{A}$ the Borel σ-algebra, T the transformation of X defined by the quotient of a linear action on $\mathbb{R}^d$ of a matrix $A \in SL(d, \mathbb{R})$, m an invariant probability measure. One particular invariant probability measure is the Lebesgue measure λ.

4 a. Ergodicity, mixing, entropy

Theorem $(X, \mathcal{A}, \lambda, T)$ *is ergodic and mixing if, and only if, no eigenvalue of* A *is a root of unity.*

Proof: $L^2(\mathbb{T}^d, \lambda)$ admits an orthonormal basis $e_m, m \in \mathbb{Z}^d$, where $e_m(x) = e^{2i\pi\langle m,x\rangle}$. Clearly $e_m(Tx) = e_{A^t m}(x)$. Therefore, spectral properties of the Koopman operator T on $L^2(\mathbb{T}^d, \lambda)$ can be read on the properties of the action of A^t on $\mathbb{Z}^d$.

$A^t e_0 = e_0$ corresponds to the space of constant functions.

If all other orbits of the action of A^t on $\mathbb{Z}^d$ are infinite, $(X, \mathcal{A}, \lambda, T)$ is mixing (exercise).

If $(X, \mathcal{A}, \lambda, T)$ is non mixing, there is a finite orbit, i.e. there are $m \in \mathbb{Z}^d, k \in \mathbb{N}$ such that $(A^t)^k m = m$. A^t (and therefore A) has an eigenvalue root of unity.

Finally, if A^t has an eigenvalue root of unity, there are $v \in \mathbb{R}^d, k \in \mathbb{N}$ such that $(A^t)^k v = v$. Clearly v is rational and can be chosen with integer coordinates. The orbit of v in $\mathbb{Z}^d$ is then finite and the function $f(x) = \sum_{j=0}^{k} e^{2i\pi\langle (A^t)^j v, x\rangle}$ is invariant: $(X, \mathcal{A}, \lambda, T)$ is not ergodic. $\square$

Let A be a linear automorphism of $\mathbb{T}^d$. We denote $\lambda_1, \lambda_2 \cdots, \lambda_K$ the different moduli of the eigenvalues of A. Each λ_i appears with multiplicity m_i, $\sum_{i=1}^{K} m_i = d$.

Theorem $h_\lambda(T) = \sum_{i=1}^{K} m_i \max\{0, \ln \lambda_i\}$.

Proof: Clearly, for all n, ε, $\lambda(B(x,n,\varepsilon)) = \lambda(B(0,n,\varepsilon))$. For $i = 1, \cdots K$, let V_i be the sum of eigenspaces of eigenvalues of modulus λ_i, and write $x = \sum x_i, x_i \in V_i$. Fix ε. Then, for all large n:

$$\{y = \sum_i y_i : \max\{1, \lambda_i^n\}|y_i| \le \varepsilon e^{-n\varepsilon}\} \subset B(0,n,\varepsilon)$$
$$\subset \{y = \sum_i y_i : \max\{1, \lambda_i^n\}|y_i| \le \varepsilon e^{n\varepsilon}\}.$$

(Use the real Jordan form of A.) Then, for all x,

$$|\limsup_n -\frac{1}{n} \ln \lambda(B(x,n,\varepsilon)) - \sum_{i=1}^{K} m_i \max\{0, \ln \lambda_i\}| \le \varepsilon.$$

The proposition follows by Brin-Katok Theorem (use variant 4 if $(X, \mathcal{A}, \lambda, T)$ is not ergodic). $\Box$

Proposition m *any T-invariant mesure.*

$$h_m(T) \leq \sum_{i=1}^{K} m_i \max\{0, \ln \lambda_i\}.$$

Proof: May assume that m is ergodic. Set

$$E_n := \{y : e^{-n(h_m(T)+\delta)} \leq m(B(y,n,\varepsilon)) \leq e^{-n(h_m(T)-\delta)}\}.$$

Then, for n large enough, ε small, $m(E_n) \geq 1/2$. Choose a set $A_n \subset E_n$ maximal with the property that that for $y_1, y_2 \in A_k$, $B(y_1, n, \varepsilon/2) \cap B(y_2, n, \varepsilon/2) = \emptyset$. Clearly,

$$\#A_n \leq (\min_x \lambda(B(x,n,\varepsilon/2)))^{-1} \leq (\frac{2}{\varepsilon})^K e^{K\varepsilon} \prod_i (\max\{1, \lambda_i^n\})^{m_i}.$$

But, by maximality, $\cup_{\{x \in A_n\}} B(x,n,\varepsilon)$ covers E_n and therefore $\#A_n \geq 1/2 e^{n(h_m(T)-\delta)}$. Comparing the two estimates gives the result. $\square$

4 b. Dimension on $\mathbb{T}^2$

In this subsection, we assume T is an ergodic linear automorphism of $\mathbb{T}^2$. Then (why?), the eigenvalues of the matrix A are $\lambda > 1 > \lambda^{-1}$. We have for any ergodic probability measure m:

Theorem $\displaystyle \lim_{\varepsilon \to 0} \frac{\ln m(B(x,\varepsilon))}{\ln \varepsilon} = \frac{2h_m(T)}{\ln \lambda}.$

Proof: Write $BB(x,n,\varepsilon) := \{y : d(T^k x, T^k y) < \varepsilon,$ for $|k| < n\}$ Observe that for all n

$$BB(x,n,C^{-1}\varepsilon) \subset B(x,\varepsilon\lambda^{-n}) \subset BB(x,n,C\varepsilon),$$

for some constant C depending on the angle between the invariant directions V_1, V_2. Conclude using BK variant 1. $\square$

Hausdorff dimension

Recall the definition of the Hausdorff dimension of a subset E of a metric space (X, d).

For a cover $\mathcal{U} = \{U_i\}$ of E and $s \geq 0$, write
$H_{s,\mathcal{U}} := \sum_i (\mathrm{Diam} U_i)^s$ and $H_s(E) := \inf_{\mathcal{U} \text{ covers } E} H_{s,\mathcal{U}}$. Then,

$$\text{H-dim}(E) := \sup\{s : H_s(E) = \infty\} = \inf\{s : H_s(E) = 0\}.$$

A measure m on a metric space X is said to be *exact dimensional* if $\lim_{\varepsilon \to 0} \dfrac{\ln m(B(x,\varepsilon))}{\ln \varepsilon}$ exists and is constant m-a.e. and the limit constant is called the dimension of m.

Verify that, if m is exact dimensional of dimension δ, then:

$$\delta = \inf\{\text{H-dim}(E), E \text{ Borel and } m(E) = 1\}.$$

Let $X=[0,1]^p\times[0,1]^q$ with the max metric, π the projection on the first coordinate and $m=\int\left(m_s^\eta(dt)\right)[\pi_* m](ds)$ as in Example 2 page 9.

Lemma *Assume $\pi_* m$ is exact dimensional of dimension δ and that for $[\pi_* m]$-a.e.s, m_s^η is exact dimensional of dimension γ. Then,*

$$\liminf_{\varepsilon\to 0}\frac{\ln m(B(x,\varepsilon))}{\ln\varepsilon}\geq\delta+\gamma.$$

Typically, the inequality is strict. For example, let $\{(s,B_s)\}\subset[0,\infty)\times\mathbb{R}]$ be the graph of the Brownian process B and m the measure which projects on Lebesgue on the s coordinate. it is known that m has dimension 3/2, but the projection has dimension 1 and the conditional measures δ_{B_s} have dimension 0. In that case, it is true that by considering the other order of projections, one gets indeed $3/2=1+1/2$.

Proof: Fix $\varepsilon > 0$. Choose A_1 and N_1 s.t. $m(A_1) > 1-\varepsilon$ and, for $(s,t) \in A_1, n \geq N_1$,

$$m_s(B^q(t, 2e^{-n})) \leq e^{-n\gamma} e^{n\varepsilon}.$$

Then, by Lebesgue Density Theorem, choose A_2 and $N_2 \geq N_1$ s.t. $m(A_2) > 1-\varepsilon$ and, for $(s,t) \in A_2, n \geq N_2$,

$$m(A_1 \cap B((s,t), e^{-n})) \geq \frac{1}{2} m(B((s,t), e^{-n})).$$

We get for $(s,t) \in A_2$ and $n \geq N_2$,

$$\begin{aligned} m(B((s,t), e^{-n})) &\leq 2 \int_{B^p(s,e^{-n})} m_s(A_1 \cap B^q(t, e^{-n}))[\pi_* m](dt) \\ &\leq 2 e^{-n\gamma} e^{n\varepsilon} [\pi_* m](B^p(s, e^{-n})). \end{aligned}$$

□

Lemma (variant) *Let $X = \Omega \times [0,1]^q$, π the projection on the first coordinate and $m = \int \Big(m_\omega(dt)\Big)[\pi_* m](d\omega)$. Assume that for $\pi_* m$-a.e. ω, m_ω is exact dimensional of dimension γ. Then, at m-a.e. (ω, t):*

$$\liminf_{\varepsilon \to 0} \frac{\ln m(B^q((\omega, t), \varepsilon))}{\ln \varepsilon} \geq \gamma.$$

5. Entropy, exponents and dimension for linear hyperbolic automorphisms

A linear automorphism of a torus is called *hyperbolic* if the matrix A has no eigenvalue of modulus 1. We have:

Theorem *Let T be a hyperbolic linear automophism of the torus, m an invariant ergodic probability measure.*
Then, m is exact dimensional: there is δ s.t.

$$\lim_{\varepsilon \to 0} \frac{\ln m(B(x,\varepsilon))}{\ln \varepsilon} = \delta \text{ for } m \text{ a.e.} x.$$

Proof If there are different moduli, then dynamical Bowen ball are no more directly comparable to usual metric ball. The proof will go by analyzing one exponent at a time.

We write $\lambda_1, \cdots, \lambda^u$ for the moduli of eigenvalues which are > 1, $\lambda^s, \cdots, \lambda_K$ for the others. We regroup $W_i = \oplus_{j \leq i} V_j$ for $i \leq u$, $W_i = \oplus_{j \geq i} V_j$ for $i \geq s$.

In general, the partition $\mathcal{W}_i$ of $\mathbb{T}^d$ obtained by projecting the affine planes parallel to W_i in $\mathbb{R}^d$ is not a measurable partition. A measurable partition ξ_i is said to be *subordinated* to $\mathcal{W}_i$ if for m-a.e. x, $\xi(x)$ is an open neighborhood of x inside $\mathcal{W}_i(x)$. We will show that, if the measurable partition ξ_i is subordinated to $\mathcal{W}_i$, then the conditional measures $m_x^{\xi_i}$ are exact dimensional.

More precisely there are numbers $\delta_i, h_i, i = 1, \cdots, K$, such that, if ξ_i is a measurable partition subordinated to $\mathcal{W}_i$, then:

More precisely there are numbers $\delta_i, h_i, i = 1, \cdots, K$, such that, if ξ_i is a measurable partition subordinated to $\mathcal{W}_i$, then:

$$\lim_{\varepsilon\to 0} \frac{\ln m_x^{\xi_i}(B(x,\varepsilon))}{\ln \varepsilon} = \delta_i, \text{and} \lim_n -\frac{1}{n} \ln m_x^{\xi_i}(BB(x,n,\varepsilon_0)) = h_i \quad (2)$$

for some small ε_0 and that moreover,

$$h_u = h_s = h. \quad (3)$$

We will show that (2) for u and s and (3) imply that the measure m is exact dimensional, with dimension $\delta_u + \delta_s$. We will then prove (2) by induction on $i \leq u$. The proof is the same for $i \geq s$.

Let us first show (3): we have $h_u \leq h$ (page 33). To see that $h_u = h$, choose P be a partition in small cubes, small enough that $P_{-\infty}^{0}(x) \subset B^{W_u(x)}(x,\varepsilon_0)$, $P_0^{+\infty}(x) \subset B^{W_s(x)}(x,\varepsilon_0)$ so that $P_{-\infty}^{+\infty}(x) = \{x\}$ and thus $h(P,T) = h(T)$. If ξ is a decreasing partition adapted to $\mathcal{W}_u$ with $\xi \prec P_0^{\infty}$ (see below LS Lemma page 59), then

$$h_u \geq \lim_n \frac{1}{n} H(P_0^n | T^{-1}\xi) = H(P|P_1^{\infty} \vee T^{-1}\xi) = H(TP|P_1^{\infty}) = h.$$

(2) and (3) imply m is exact dimensional and $\delta = \delta_u + \delta_s$.

1) Locally, $\mathcal{W}_u$ and $\mathcal{W}_s$ form a system of coordinates, and the max distance is equivalent to the original distance on $\mathbb{T}^d$. Let π be the projection on some W_s space along $\mathcal{W}_u$. Then $m = \int m^u d[\pi_* m]$ and $\pi_* m$ is an average of measures m^s. By Lemma page 51, $\liminf_{\varepsilon\to 0} \frac{\ln[\pi_* m](B^{W_s}(t,\varepsilon))}{\ln \varepsilon} \geq \delta_s$ $\pi_* m$-a.e. and by Lemma page 50, $\liminf_{\varepsilon\to 0} \frac{\ln m(B(x,\varepsilon))}{\ln \varepsilon} \geq \delta_u + \delta_s$ m-a.e..

2) Conversely, to estimate $\limsup_{\varepsilon\to 0} \frac{\ln m(B(x,\varepsilon))}{\ln \varepsilon}$ from above, we want to estimate from below $m(B(x, 4e^{-n}))$. Let P be a partition as above, small enough that $P_{-\infty}^{0}(x) \subset B^{W_u(x)}(x, \varepsilon_0)$, $P_0^{+\infty}(x) \subset B^{W_s(x)}(x, \varepsilon_0)$ so that $P_{-\infty}^{+\infty}(x) = \{x\}$ and $h(P,T) = h(T)$. Choose a large enough that the diameter of the atoms of P_{-an}^{an} is smaller than e^{-n}, say $a = 2\max\{(\ln \lambda_u)^{-1}, |\ln \lambda_s|^{-1}\}$. Using (2) and (3), we have, on a set A_1 of large measure, for $n \geq N_1$ large enough:

$$m_x^s(P_{-na}^{0}(x)), m_x^u(P_0^{na}(x)) \leq e^{-na(h-2\varepsilon)},\ m(P_{-na}^{na}(x)) \geq e^{-2na(h+2\varepsilon)}.$$

Using (2) and Lebesgue Density Theorem, choose $A_2 \subset A_1$ of large measure and N_2 such that for $x \in A_2$, $n \geq N_2$:

$$m_x^u(A_1 \cap B^{W_u}(x, e^{-n})) \geq e^{-n(\delta_u+\varepsilon)}.$$

Analogously, choose $A_3 \subset A_2$ of large measure and N_3 such that for $x \in A_3$, $n \geq N_3$:

$$m_x^s(A_2 \cap B^{W_s}(x, e^{-n})) \geq e^{-n(\delta_s+\varepsilon)}.$$

All the above construction ensures that if $x \in A_3$, $B(x, 4e^{-n})$ contains at least $e^{-n(\delta_u+\delta_s+2\varepsilon-2ah+2a\varepsilon)}$ atoms of P_{-na}^{na} with measure at least $e^{-2na(h+2\varepsilon)}$. This shows the claim.

Proof of (2) We prove by induction on $i, 1 \leq i \leq u$ (and similarly on $j = K - i, 0 \leq j \leq K - s$) that there are numbers $\gamma_i, i = 1, \cdots, K, 0 \leq \gamma_i \leq m_i$ such that

for $i \leq u$, $\delta_i = \sum_{j \leq i} \gamma_i, h_i = \sum_{j \leq i} \gamma_i \ln \lambda_i$,

for $i \geq s$, $\delta_i = \sum_{j \geq i} \gamma_i, h_i = \sum_{j \geq i} \gamma_i \ln \lambda_i$.

Lemma [L.-Strelcyn] *There exist, for* $1 \leq i \leq u$, *increasing measurable partitions* ξ_i *such that* ξ_i *is subordinated to* $\mathcal{W}_i$ *and* $\xi_i \prec \xi_{i-1}$ *(Set* $\xi_0 = \varepsilon$*).*

We admit LS Lemma, which will be proven later.

Step 1: $m_x^{\xi_1}$ *are exact dimensional, with dimension* $\delta_1 = \gamma_1$ *and the entropy* $h_1 = h(\xi_1, T) = \gamma_1 \ln \lambda_1$.
We have, as above,

$$\{y_1 : \lambda_1^n |y_1| \leq \varepsilon e^{-n\varepsilon}\} \subset B(0, n, \varepsilon) \cap W_1(0) \subset \{y_1 : \lambda_1^n |y_1| \leq \varepsilon e^{n\varepsilon}\}.$$

By BK variant 2, we know that $h(\xi_1, T) = \lim_n -\frac{1}{n} \ln m_x^{\xi_1}(B(x, n, \varepsilon_0)) =: h_1$. It follows that, for a.e. x, all $\varepsilon > 0$:

$$\limsup_{r \to 0} \frac{\ln m_x^{\xi_i}(B(x,r))}{\ln r} - \varepsilon \leq h_1 / \ln \lambda_1 \leq \liminf_{r \to 0} \frac{\ln m_x^{\xi_i}(B(x,r))}{\ln r} + \varepsilon.$$

Step 1 follows.

We assume that $m_x^{\xi_{i-1}}$ are exact dimensional, with dimension δ_{i-1}. We know that $h_{i-1} = h(\xi_{i-1}, T) = \lim_n -\frac{1}{n} \ln m_x^{\xi_{i-1}}(B(x,n,\varepsilon_0))$, $h_i = h(\xi_i, T) = \lim_n -\frac{1}{n} \ln m_x^{\xi_i}(B(x,n,\varepsilon_0))$. We need to show that

$$\lim_{r\to 0} \frac{\ln m_x^{\xi_i}(B(x,r))}{\ln r} = \delta_{i-1} + \frac{h_i - h_{i-1}}{\ln \lambda_i}.$$

Step 2: $\displaystyle\limsup_{r\to 0} \frac{\ln m_x^{\xi_i}(B(x,r))}{\ln r} \le \delta_{i-1} + \frac{h_i - h_{i-1}}{\ln \lambda_i}.$

We want to estimate from below $m_x^{\xi_i}(B(x, \lambda_i^{-n}\varepsilon))$. The set $B(x, \lambda_i^{-n}\varepsilon)$ is made of $B(.,n,\varepsilon)$ balls of measure at least $e^{-n(h_i+\varepsilon)}$. In order to count them, intersect with a typical $\mathcal{W}_{i-1}$ leaf. The intersections have $m_x^{\xi_{i-1}}$-measure less than $e^{-n(h_{i-1}-\varepsilon)}$ and should fill up at least $e^{-n \ln \lambda_i(\delta_{i-1}+\varepsilon)}$. Their number is at least $e^{-n(\ln \lambda_i \delta_{i-1} - h_{i-1} + (\ln \lambda_i + 1)\varepsilon)}$. The estimate from below of the measure of balls follow. The precise argument uses Lebesgue Density Theorem in a way similar to the argument on pages 57/58.

Step 3. Approximating $I(T^{-1}\xi_{i-1}|\xi_{i-1})$

The estimate in Step 2 is a priori way off since there might be different good $B(n,\varepsilon)$ balls to count for different typical $\mathcal{W}_{i-1}$ leaves. Each atom of ξ_i is an open set in a plane parallel to W_i and ξ_{i-1} partitions it into open sets of planes parallel to W_{i-1}. We call π_i the projection on the V_i direction parallel to W_{i-1} and

$$B^{T_i}(x,\delta) = \{y; y \in \xi_i, d_{V_i}(\pi_i(x), \pi_i(y)) < \delta\}.$$

Set:

$$g_\delta(x) = \frac{1}{m_x^{\xi_i}(B^{T_i}(x,\delta))} \int_{B^{T_i}(x,\delta)} m_z^{\xi_i}([T^{-1}\xi_{i-1}](z)) dm_x^{\xi_i}(z)$$

$$\text{and } g^* = \inf_{\delta>0} g_\delta.$$

Lemma *As* $\delta \to 0$, $\gamma_\delta(x) \to I(T^{-1}\xi_{i-1}|\xi_{i-1})(x)$ *at* m*-a.e.* x *and* $\int -\ln g^* < \infty$.

The convergence follows from Lebesgue Density Theorem applied to the measures $(\pi_i)_* m_x^{\xi_i}$. See Theorem page 95.

Step 4: Transversal dimension

We show that the measures $\mu_x^i := (\pi_i)_* m_x^{\xi_i}$ are exact dimensional with dimension $\gamma_i = \dfrac{h_i - h_{i-1}}{\ln \lambda_i}$. Assume first that there is no Jordan block in V_i and write $a(k,x) := B^{T_i}(x, \lambda_i^{-k})$. We have:

$$m_x^{\xi_i}(a(n,x)) = \prod_{k=0}^{n-1} \frac{m_{f^k x}^{\xi_i}(a(n-k, f^k x))}{m_{f^{k+1}x}^{\xi_i}(a(n-k-1, f^{k+1}x))}.$$

By composition of conditional measures and invariance, we can write:

$$m_{f^{k+1}x}^{\xi_i}(a(n-k-1, f^{k+1}x)) = \frac{m_{f^k x}^{\xi_i}(f^{-1}[a(n-k-1, f^{k+1}x)])}{m_{f^k x}^{\xi_i}([f^{-1}\xi_i](f^k x))}$$

and, by the contraction along V_i:

$$\xi_i(f^k x) \cap f^{-1}[a(n-k-1, f^{k+1}x)] = [f^{-1}\xi_i](f^k x) \cap a(n-k, f^k x)]$$

It follows that

$$m_x^{\xi_i}(a(n,x)) = \prod_{k=0}^{n-1} \frac{m_{f^k x}^{\xi_i}(a(n-k, f^k x)) m_{f^k x}^{\xi_i}([f^{-1}\xi_i](f^k x))}{m_{f^k x}^{\xi_i}([f^{-1}\xi_i](f^k x) \cap a(n-k, f^k x))} = \prod_{k=0}^{n-1} \frac{m_{f^k x}^{\xi_i}([f^{-1}\xi_i](f^k x)}{g_{\lambda_i^{-(n-k)}}(f^k x)}$$

so that

$$-\frac{1}{n\ln\lambda_i}\ln(m_x^{\xi_i}(a(n,x))) = \frac{1}{n\ln\lambda_i}\sum_{k=0}^{n-1}\Big(\ln g_{\lambda_i^{-(n-k)}}(f^k x) + I(T^{-1}\xi_i|\xi_i)(f^k x)\Big).$$

The first average converges to $-h_{i-1}/\ln\lambda_i$ by step 3 and Claim page 34, the second to $h_i/\ln\lambda_i$ by the Birkhoff Ergodic Theorem.

In the case of Jordan blocks, define $a_\pm(k,x) := B^{T_i}(x, c_{k,\pm})$ in such a way that $\lambda_i^{-k(1+\varepsilon)} \leq c_{k,-} \leq c_{k,+} \leq \lambda_i^{-k(1-\varepsilon)}$,

$$\begin{aligned} \xi_i(f^k x)\cap f^{-1}[a_-(n-k-1, f^{k+1}x)] \quad &\subset \quad [f^{-1}\xi_i](f^k x)\cap a_-(n-k, f^k x) \\ \text{and } [f^{-1}\xi_i](f^k x)\cap a_+(n-k, f^k x) \quad &\subset \quad \xi_i(f^k x)\cap f^{-1}[a_+(n-k-1, f^{k+1}x)]. \end{aligned}$$

The conclusion follows in the same way, up to ε, for all ε.

Step 5 $\displaystyle\liminf_{r\to 0}\frac{\ln m_x^{\xi_i}(B(x,r))}{\ln r} \geq \delta_{i-1} + \frac{h_i - h_{i-1}}{\ln\lambda_i}.$

Follows now from Step 4 and the dimension Lemma page 50.

This finishes the proof of exact dimension, provided we construct the partitions ξ_i with the properties of LS Lemma.

LS Lemma *There exist, for $1 \leq i \leq u$, increasing measurable partitions ξ_i such that ξ_i is subordinated to $\mathcal{W}_i$ and $\xi_i \prec \xi_{i-1}$.*

We choose on each axis of coordinates in $\mathbb{T}^d$ a ε_0 dense set of points $a_{j,1}, \cdots, a_{j,L}$ such that $|a_{j,k+1} - a_{j,k}| < \varepsilon_0/2$ and satisfying another condition seen later. We consider the partition P of $\mathbb{T}^d$ defined by the hyperplanes $H_{j,k} := \{x : x_j = a_{j,k}\}$ and, for $i = 1, \cdots, u$, η_i the partition obtained by partitioning each element of P by planes parallel to V_i. Set $\eta_0 = \varepsilon$. Then, η_i is a measurable partition for all i and $\eta_i \prec \eta_{i-1}$. Set $\xi_i = \vee_{n \geq 0} T^n \eta_i$. We have to show that we can choose the $a_{j,k}$ in such a way that $\xi_i(x)$ is a neighborhood of x in $W_i(x)$ for m-a.e. x. The other properties are straightforward. Let $0 < \lambda < \lambda_u$ and assume that $\sum_{j,k,n} m(\mathcal{U}(H_{j,k}, \lambda^{-n})) < \infty$. Then, for m-a.e. x,

$$C(x) := \inf_{n \geq 0} \lambda^n d^{W_i}(T^{-n}x, \cup_{j,k} H_{j,k}) > 0.$$

Clearly, $B^{W_i}(x, C(x)) \subset \xi_i(x)$. Therefore, we only have to choose the $a_{j,k}$ such that $\sum_n m_j([a_{j,k} - \lambda^{-n}, a_{j,k} + \lambda^{-n}]) < \infty$, where m_j is the projection of m on the j coordinate. Lebesgue almost every point on the interval $[0, 1]$ has this property (exercise; use the last statement of Lebesgue Density Theorem page 94).

6. Ergodic linear automorphisms

In general, an ergodic linear automorphisms is not hyperbolic: there are eigenvalues of modulus one (but not roots of unity).

Let V_{u+1} be the corresponding eigenspace, $\mathcal{W}_{u+1}$ the foliation of $\mathbb{T}^d$ into planes parallel to V_{u+1}. Some results (and proofs) of Section 4a and 5 apply:

for any invariant measure, the entropy is at most $\sum_{i=1}^{K} m_i \max\{0, \ln \lambda_i\}$;

the Lebesgue measure has maximal entropy; the conditional measures on $\mathcal{W}_i, i \neq u+1$ of an ergodic measure are exact dimensional and, if ξ_i is a measurable partition subordinated to $\mathcal{W}_i$, then: $h_i = \lim_n -\frac{1}{n} \ln m_x^{\xi_i}(BB(x,n,\varepsilon_0))$.

Lemma $h_u = h_s = h_m(T)$.

Proof: Consider the partition P of the proof of LS Lemma, and set $\xi_{u+1} = P^0_{-\infty}$. We have $h_m(T) = H(P|P^0_{-\infty}) = H(T^{-1}\xi_{u+1}|\xi_{u+1})$ The arguments of step 4 above, applied to the projection on V_{u+1} with the sets $a(k,x) := B^{T_{u+1}}(x, e^{-k\varepsilon})$ yield:

$$\frac{h-h_u}{\varepsilon} \leq \mathrm{esssup}_{m-\text{a.e. } x} \limsup_n \frac{\ln \mu_x^{u+1}(a(n,x))}{n\varepsilon} \leq m_{u+1}.$$

The last inequality holds since μ_x^{u+1} lives on a m_{u+1} dimensional plane. $\square$

Corollary *If m is ergodic, the Pinsker partition π is trivial.*

7. Linear endomorphisms of tori

Consider in this section a $d \times d$ matrix A with integer coefficients and T the associated transformation on $\mathbb{T}^d$.
For $x \in \mathbb{T}^d$, $\#\{T^{-1}x\} = |\text{ Det } A|$.

The Lebesgue measure λ is T invariant (exercise) and, with the same notations,

$$h_\lambda(T) = \sum_{i=1}^{K} m_i \max\{0, \ln \lambda_i\}$$

(same proof, cf. page 48).

For any invariant measure m,

$$h_m(T) \leq \sum_{i=1}^{K} m_i \max\{0, \ln \lambda_i\}.$$

(same proof, cf. page 49).

In particular, if $d = 1$, A is given by a 1×1 matrix (p) and $Tx = px(\text{mod}.1)$. $h_\lambda(T_p) = \ln p$ and for any invariant measure m, $h_m(T_p) \leq \ln p$. An ergodic invariant measure m is exact dimensional and

$$\lim_{\varepsilon \to 0} \frac{\ln m(B(x, \varepsilon))}{\ln \varepsilon} = \frac{h_m(T_p)}{\ln p}.$$

Proof: $B(x, n, \varepsilon) = [x - p^n \varepsilon, x + p^n \varepsilon]$. $\square$

Folding entropy Let m be an invariant measure. The quantity $F_m(T) = H(\varepsilon|T^{-1}\varepsilon)$ is called the folding entropy of the transformation. We have:

$$F_m(T) \leq \ln|\text{ Det } A| \text{ and } F_\lambda(T) = \ln|\text{ Det } A|.$$

Observe that $h_\lambda(T) = F_\lambda(T) - \sum_{i=1}^K m_i \min\{0, \ln\lambda_i\}$. One also has

Theorem [Shu] *Let T be a hyperbolic endomorphism of $\mathbb{T}^d$, m an ergodic invariant measure. Then m is exact dimensional. More precisely, there are numbers $\gamma_i, i = 1 \cdots, K$ such that, for m-a.e. x,*

$$\lim_{\varepsilon\to 0} \frac{\ln m(B(x,\varepsilon))}{\ln \varepsilon} = \sum_i \gamma_i,\ F_m(T) = \sum_{i=1}^K \gamma_i \ln\lambda_i$$

and $h_m(T) = \sum_{i=1}^K \gamma_i \max\{0, \ln\lambda_i\} = F_m(T) - \sum_{i=1}^K \gamma_i \min\{0, \ln\lambda_i\}$.

8. General $C^{1+\alpha}$ diffeomorphism of a compact manifold. Pesin theory

From now on, X is a compact Riemannian manifold, without boundary, $\mathcal{A}$ the Borel σ-algebra, T a C^1-diffeomorphism of X, m an invariant probability measure.
Oseledets Multiplicative Ergodic Theorem applies to the differential DT. It is remarkable that, as soon as the differential is Hölder continuous, much of the linear theory of sections 4 to 7 carries over to the non-linear case.

Theorem [Oseledets MET] *Let* (X, α, m, T) *be a smooth dynamical system with* m *ergodic. Then, there exist numbers*

$$\lambda_1 > \cdots > \lambda_K$$

and, at m*-a.e. every* x*, a splitting* $T_xM = \oplus_{j=1}^{K} V_i(x)$ *such that:*

1. $v \in T_xM$ *belongs to* $V_i(x)$ *if, and only if,*

$$\lim_{n\to\pm\infty} \frac{1}{n} \ln \|D_xT^n v\|_{T^n x} = \ln \lambda_i,$$

2. $\lim_{n\to\pm\infty} \frac{1}{n} \ln |sin\angle(V_i(T^n x), V_j(T^n x))| = 0$ *for* $i \neq j$.

To deduce our MET from the general MET for cocycles, choose an arbitrary measurable trivialization of the tangent bundle $\Phi : X \times \mathbb{R}^d \to TX$ and apply MET to the cocycle based on $A_1(x) = \Phi_{Tx}^{-1} \circ D_x T \circ \Phi_x$. Observe also that the subbundles V_i are automatically measurable and invariant: $D_x T V_i(x) = V_i(Tx)$. In particular, the dimension of V_i is an invariant function and therefore an a.e. constant m_i.

Theorem [Margulis-Ruelle inequality] *Assume m ergodic, then*

$$h_m(T) \leq \sum_{i=1}^{K} m_i \max\{0, \ln \lambda_i\}.$$

Exercise Write the non-ergodic variant of MET and of Ruelle-Margulis inequality.

Sketch of the proof of MR inequality:

1.For a linear mapping A of $\mathbb{R}^d$, denote $t_1 \geq t_2 \geq \cdots \geq t_d$ the eigenvalues of $(A^tA)^{1/2}$ and observe that $\mathcal{U}\Big(A(B(0,r),r)\Big)$ can be covered by less than $C\prod_{i=1}^{d}\max(1,t_i)$ balls of radius $r/2$.

2. Fix $n > 0$ and choose r small enough that on each ball $B(x,r)$, T^n is well approximated by $\exp_{T^nx} \circ D_xT^n \circ \exp_x^{-1}$.

3. Choose a finite partition P_r such that elements of P_r have inner diameter $r/2$ and outer diameter r.

4. Using 1,2, and 3, show that

$$I(T^{-n}P_r|P_r)(x) \leq \sum_{i=1}^{d} \ln\max(1, t_i(D_xT^n)) + C.$$

5. Conclude.

We now assume that the diffeomorphism T is $C^{1+\alpha}$.

(i.e. There are local charts Φ such that, on each domain where it is defined, $x \mapsto \Phi^{-1} \circ D_x T \circ \Phi$ and $x \mapsto \Phi^{-1} \circ D_x(T^{-1}) \circ \Phi$ are α-Hölder continuous functions with values in $GL(d, \mathbb{R})$.

Recall that a function f on a metric space is α-Hölder if $\mathrm{Lip}_\alpha := \sup_{x \neq y} \frac{|f(x)-f(y)|}{(d(x,y))^\alpha}$ is finite.)

Let, as in the linear case, u, s be such that

$$\lambda_i > 0 \Leftrightarrow i \leq u; \ \lambda_i < 0 \Leftrightarrow i \geq s.$$

Define the following equivalence relations:

$$\text{for } i \geq s, \quad y \overset{\mathcal{W}_i}{\sim} x \Leftrightarrow \limsup_{n\to\infty} \frac{1}{n} \ln d(T^n x, T^n y) \leq \ln \lambda_i$$

$$\text{for } i \leq u, \quad y \overset{\mathcal{W}_i}{\sim} x \Leftrightarrow \liminf_{n\to-\infty} \frac{1}{n} \ln d(T^n x, T^n y) \geq \ln \lambda_i$$

In the linear case, equivalence classes are planes parallel to the corresponding W_i. Set for $i \leq u$, $W_i(x) = \oplus_{j\leq i} V_j(x)$and for $i \geq s$, $W_i(x) = \oplus_{j\geq i} V_j(x)$.

Theorem [Pesin] *Let* $(X, \mathcal{A}, T, m)$ *be a smooth dynamical system with* T *of class* $C^{1+\alpha}$ *and* m *ergodic. Then, for each* $i \leq u$ *or* $i \geq s$, *the equivalence classes of* $\mathcal{W}_i$ *form a measurable lamination of* m*-almost all* X. *Individual classes* $\mathcal{W}_i(x)$ *are homeomorphic to* $\mathbb{R}^{d_i}$ *and depend measurably on* x. *There is a measurable invariant set* $\Lambda, m(\Lambda) = 1$, *such that for* $x \in \Lambda$, $T_x\mathcal{W}_i(x) = W_i(x)$.

Moreover, for fixed small ε, *there is a measurable function* $\ell : \Lambda \to (0, +\infty)$, *satisfying* $\ell(T^{\pm 1}x) \leq e^{\varepsilon}\ell(x)$ *with the following properties:*

– *The connected component* $\mathcal{W}_{i,loc}(x)$ *of* x *inside* $\mathcal{W}_i(x) \cap B(x, (\ell(x))^{-1})$ *is the exponential of a graph of a function* $\varphi_{i,x} : W_i(x) \to \mathbb{R}^d \ominus W_i(x)$,

$-\varphi_{i,x}$ *is* $C^{1+\alpha}$, $D_0\varphi_{i,x} = 0$ *and* $Lip_\alpha(D\varphi_{i,x}) \leq \ell(x)$,

– *there is a metric* $\delta_{i,x}$ *on* $\mathcal{W}_{i,loc}(x)$ *and a constant* C *such that, for* $y, z \in W_{i,loc}(x)$:

$$Cd(y,z) \leq \delta_{i,x}(y,z) \leq \ell(x)d(y,z),$$

– *for* $i \geq s$, $T\mathcal{W}_{i,loc}(x) \subset \mathcal{W}_{i,loc}(Tx)$ *and, for* $y, z \in W_{i,loc}(x)$:

$$\delta_{i,Tx}(Ty, Tz) \leq \lambda_i e^{\varepsilon} \delta_{i,x}(y, z),$$

– *for* $i \leq u$, $T^{-1}\mathcal{W}_{i,loc}(x) \subset \mathcal{W}_{i,loc}(T^{-1}x)$ *and, for* $y, z \in W_{i,loc}(x)$:

$$\delta_{i,T^{-1}x}(T^{-1}y, T^{-1}z) \leq \lambda_i^{-1} e^{\varepsilon} \delta_{i,x}(y, z).$$

In short, the nonlinear picture is the same as the linear picture, at least on a small ball around almost every point. How small depends on the point, but is slowly varying along the orbit. In particular, for $i \le u$,

$$\mathcal{W}_0(x) := \{x\} \subset \mathcal{W}_1(x) \subset \mathcal{W}_2(x) \subset \cdots \subset \mathcal{W}_u(x)$$

and each $\mathcal{W}_{i-1}$ family defines a foliation of $\mathcal{W}_i(x)$. Let $\pi_{i,x} : \mathcal{W}_{i,loc} \to \exp_x \varphi_{i,x} V_i(x)$ be the projection along $\mathcal{W}_{i-1,loc}$ leaves (compare with page 61). Then:

Lemma[Barreira-Pesin-Schmeling] *The mapping $\pi_{i,x}$ is Lipschitz and the Lipschitz constant depends only on $\ell(x)$.*

Given Pesin Theorem and BPS Lemma, we can do all the geometrical arguments of Section 5. For the dynamical arguments, we can compare the orbits of x and y only as long as $d(T^i x, T^j y) \leq \ell^{-1}(T^j x)$.

Lemma [Mañé] *Let* $(X, \mathcal{A}, m, T)$ *be a smooth ergodic dynamical system,* $f : x \to (0, +\infty)$ *a function satisfying* $f(Tx) \leq Af(x)$ *for some* A*. Then, there exist a partition* P *with finite entropy such that for* m*-a.e.* x*, for* n *large enough,*

$$y \in P_0^{+\infty}(x) \Longrightarrow d(T^n y, T^n x) \leq f(T^n x).$$

Proof Choose $\delta_0 > 0$ such that the set $B := \{x : f(x) \geq \delta_0\}$ has positive measure and consider for $n > 0$:

$$B_n := \{x, x \in B, T^j x \notin B \text{ for } 1 \leq j < n, T^n x \in B\}.$$

Then $\sum_n nm(B_n) = 1$ (Kac Lemma). Let L be the Lipschitz constant of T. Since X is a Riemannian compact manifold, there is a number D such that for each N, one can find a partition P_N of X into less than D^N sets of diameter smaller than $(AL)^{-N}\delta_0$. Let Q be the partition $Q = (B_0 = B^c, B_1, \cdots)$ and P the partition obtained by cutting each B_n by P_n. We have:

$H(Q) < \infty$ since, by the argument page 24 with $q_n = \frac{e^{-n}}{1-e^{-1}}$, $H(Q) \leq -\ln(1 - e^{-1}) + \sum_n nm(B_n) < \infty$.

$H(P) \leq H(Q) + H(P|Q) \leq H(Q) + \sum_n m(B_n) \ln D^n < \infty$, and

P has the desired property.

Let indeed k be the first time the orbit of x enters B. As soon as $n \geq k$, there is a last integer $k_n \leq n$ with the property that $T^{k_n}x \in B$. Then, $T^{k_n}x \in B_p$, for some $p, 0 \leq n - k_n < p$. If $y \in P_0^\infty(x)$, $T^{k_n}y \in P(T^{k_n}x)$ so that $T^{k_n}y \in B_p$ and $d(T^{k_n}x, T^{k_n}y) \leq (AL)^{-p}\delta_0$ By the Lipschitz property, $d(T^n x, T^n y) \leq L^{n-k_n}(AL)^{-p}\delta_0 \leq A^{-p}\delta_0$, whereas $f(T^n x) \geq A^{k_n - n} f(T^{k_n}x) \geq A^{-p}\delta_0$. $\square$

Exercise Let $(X, \mathcal{A}, m, T)$ be a smooth ergodic dynamical system, $f : x \to (0, +\infty)$ a function satisfying $f(Tx) \leq Af(x)$ for some A. Write and prove the variant of Brin-Katok Theorem page 40 involving

$$B(x, n, \varepsilon f) := \{y : d(T^j x, T^j y) < \varepsilon f(T^j x) \text{ for } 0 \leq j \leq n\}$$

Theorem [Barreira-Pesin-Schmeling] *Let $(X, \mathcal{A}, m, T)$ be a $C^{1+\alpha}$ ergodic dynamical system. Assume that* 0 *is not an exponent of the system ($s = u + 1$). Then, m is exact dimensional.*

The proof follows the arguments of the linear case. Observe though that one needs a new argument to show that, for m-a.e.x,

$$\liminf_{\varepsilon \to 0} \frac{\ln m(B(x,\varepsilon))}{\ln \varepsilon} \geq \delta_u + \delta_s.$$

Contrarily to 1) page 57, it is not true any more that $\mathcal{W}_u$ and $\mathcal{W}_s$ form a system of coordinates where the max distance is equivalent to the original distance on X.

Observe that the arguments for 2) page 57/58 are valid in the nonlinear case. With a little care, they can also englobe the W_{u+1} direction in the presence of 0 exponents and yield:

Theorem [L-Young] *Let* $(X, \mathcal{A}, m, T)$ *be a* $C^{1+\alpha}$ *ergodic dynamical system. Assume that the exponent* 0 *has multiplicity* m_{u+1}*. Then:*

$$\limsup_{\varepsilon \to 0} \frac{\ln m(B(x,\varepsilon))}{\ln \varepsilon} \leq \delta_u + \delta_s + m_{u+1}.$$

Actually, there are numbers $\gamma_i, 0 \leq m_i$, which can be seen as transverse dimensions within $\mathcal{W}_i$ leaves, such that

for $i \leq u$, $\delta_i = \sum_{j \leq i} \gamma_i, h_i = \sum_{j \leq i} \gamma_i \ln \lambda_i$,

for $i \geq s$, $\delta_i = \sum_{j \geq i} \gamma_i, h_i = \sum_{j \geq i} \gamma_i \ln \lambda_i$.

In particular, if m is absolutely continuous, then

$$\delta_u + \delta_s + m_{u+1} = \sum_{j \neq u+1} \gamma_j + m_{u+1} = d = \sum_j m_j.$$

It follows that if m is absolutely continuous, $\gamma_j = m_j$ for all $j \neq u+1$ and therefore:

Theorem [Pesin Formula] *Let* $(X, \mathcal{A}, m, T)$ *be a* $C^{1+\alpha}$ *ergodic dynamical system, with* m *absolutely continuous with respect to the volume on* X. *Then:*

$$h_m(T) = \sum_{i=1}^{K} m_i \max\{0, \ln \lambda_i\}.$$

9. Entropy and absolutely continuous conditional measures

$(X, \mathcal{A}, T)$ an ergodic linear automorphism of the torus, m an invariant ergodic measure, λ_i, m_i, h_i as above for $1 \leq i \leq u$.

Theorem $h_i \leq \sum_{j=1}^{i} m_j \ln \lambda_j$ *with equality iff* m *is the Lebesgue measure.*

Proof: Consider the increasing LS partition ξ_i such that $h_i = H(T^{-1}\xi_i|\xi_i)$. We first show that the conditional measures $m_x^{\xi_i}$ are proportional to the Lebesgue measure on $\mathcal{W}_i(x)$. Let $L_i(x)$ be the DimW_i-Lebesgue measure of $\xi_i(x)$. By definition, the DimW_i-Lebesgue measure of $[T^{-1}\xi_i](x)$ is given by $\frac{L_i(Tx)}{\text{Det}A|_{W_i}} = \frac{L_i(Tx)}{\prod_{j=1}^{i} \lambda_i^{m_i}}$. We have, at m-a.e. x:

Lemma $H_{m_x^{\xi_i}}(T^{-1}\xi_i) \leq \sum_{j=1}^{i} m_j \ln \lambda_j + \int (\ln L_i(y) - \ln L_i(Ty))[m_x^{\xi_i}](dy)$, *with equality iff* $m_x^{\xi_I}([T^{-1}\xi_i](x)) = \frac{\text{Leb.}[T^{-1}\xi_i](x)}{\text{Leb.}[\xi_i](x)}$.

Follows from convexity of $\phi(t) = t \ln t$ (see page 24: $\sum_i q_i \phi\left(\frac{m(P_i)}{q_i}\right) \geq \phi\left(\sum_i q_i \frac{m(P_i)}{q_i}\right) = 0$ with $m = m_x^{\xi_i}$ and $q_i = \frac{L_i(Tx)}{L_i(x) \prod_{j=1}^{i} \lambda_i^{m_i}}$.)

Observe that in the previous Lemma, $\int(\ln L_i(y) - \ln L_i(Ty))[m_x^{\xi_i}](dy)$ makes sense (and might be $+\infty$) because $\ln L_i(y)$ is constant on $\xi_i(x)$ and $-\ln L_i(Ty)$ is bounded from below. Me may integrate in m and get:

$$H(T^{-1}\xi|\xi) \leq \sum_{j=1}^{i} m_j \ln \lambda_j + \int (L_i(y) - L_i(Ty))m(dy).$$

In particular $\inf\{0, L_i - L_i \circ T\}$ is integrable. The following Lemma is an exercise of application of Birkhoff PET:

Lemma: *Let L be a measurable function such that* $\inf\{0, L - L \circ T\}$ *is integrable. Then,* $\int L - L \circ T = 0$.

It follows from the two Lemmas that $h_i \leq \sum_{j=1}^{i} m_j \ln \lambda_j$ and that in the case of equality, $m_x^{\xi_I}([T^{-1}\xi_i](x)) = \frac{\text{Leb.}[T^{-1}\xi_i](x)}{\text{Leb.}[\xi_i](x)}$. By the same argument, we also have

$$m_x^{\xi_I}([T^{-n}\xi_i](x)) = \frac{\text{Leb.}[T^{-n}\xi_i](x)}{\text{Leb.}[\xi_i](x)}.$$

The claim follows since $\cup_n \mathcal{B}_{T^{-n}\xi}$ generate the σ-algebra $\mathcal{A}$.

Since the automorphism is ergodic, the leaf $\mathcal{W}_i(0) = W_i$ is dense. Let τ be an element of W_i such that τ^n is dense. Since the conditional measures of m along $\mathcal{W}_i$ are proportional to Lebesgue, the measure m is invariant by $R_\tau x := x + \tau$(mod.1). By examining for instance its Fourier coefficients, one sees that the only R_τ invariant measure is the Lebesgue measure. $\square$

For a $C^{1+\alpha}$ diffeomorphism, part of the claim is still true, but this does not in general suffice to characterize the invariant measure:

Theorem *Let T be a $C^{1+\alpha}$ diffeomorphism of a compact manifold X, m an ergodic invariant measure. Then,*

$$h_i \leq \sum_{j=1}^{i} m_j \ln \lambda_j$$

with equality iff the conditional measures of m along $\mathcal{W}_i$ are absolutely continuous with respect to the DimW_i-dimensional Lebesgue measure.

In particular,

Corollary *Let T be a C^2 diffeomorphism of a compact manifold X, m an ergodic invariant measure. Then,*

$$h \leq \sum_{j=1}^{K} m_j \max\{0, \ln \lambda_j\}$$

with equality iff the conditional measures of m along $\mathcal{W}_u$ are absolutely continuous with respect to the DimW_u-dimensional Lebesgue measure.

Corollary follows from Theorem in the case $i = u$ once we prove $h_u = h$. In order to use in the nonlinear case the same arguments as the ones page 66, one needs to show that the analog of the projection over V_{u+1} is Lipschitz, so that the 'transversal dimension' makes sense and is finite. This is done in [LY] with the hypothesis that the diffeomorphism is C^2. It is very likely that the statement holds in the $C^{1+\alpha}$ case as well (see in particular [BW]), but this statement has not been formally written to this day.

10. Notes and comments

10 a. Density theorems on $\mathbb{R}^n$ We used several forms of Lebesgue Density Theorem in these notes. Here we recall the logic of these results. They all follow from

Theorem [Besicovich Covering Lemma] *Let* $E \subset \mathbb{R}^n$, $r : E \to (0, \infty)$ *a bounded function on* E. *Then, there is a* $c(n)$, *depending only on* n, *such that the cover* $\{B(x, r(x)), x \in E\}$ *admits a subcover* $\mathcal{C}$ *such that no* x *in* $\mathbb{R}^n$ *is covered by more than* $c(n)$ *balls from* $\mathcal{C}$.

Let now μ be a probability measure on $\mathbb{R}^n$, and $g \in L^1(\mu), g \geq 0$. Define:

$$g_\delta(x) := \frac{1}{\mu(B(x,\delta))} \int_{B(x,\delta)} g d\mu, \ g^* := \sup_\delta g_\delta \ \text{ and } g_* := \inf_\delta g_\delta.$$

Lemma [Maximal Lemma] *a) For* $\lambda > 0, \mu(g^* > \lambda) \leq \frac{c(n)}{\lambda} \int g d\mu$, *and b)* $\int_{g_* < \lambda} g d\mu \leq c(n)\lambda$.

Proof Use Besicovich Covering Theorem.

The Maximal Lemma is used in proving:

Theorem[Lebesgue Density Theorem] *Let* $g \in L^1(\mu)$. *Then,*

$$g_\delta \to g \ \mu\text{-}a.e. \text{ as } \delta \to 0.$$

In particular, for $A \in \mathcal{B}, \mu(A) > 0$, $\dfrac{\mu(A \cap B(x,\delta))}{\mu(B(x,\delta))} \to 1$ μ*-a.e. on* A *as* $\delta \to 0$ *and* $\limsup_\delta \dfrac{\mu(B(x,\delta))}{\lambda(B(x,\delta))} < \infty$ λ*-a.e.*.

We also used BCL on page 61: let (X, μ) be a Lebesgue space, $\pi : X \to \mathbb{R}^n$ measurable and $\{\mu_t, t \in \mathbb{R}^n\}$ an associated family of conditional probabilities. Let P be a countable partition of X with $H(P) < \infty$. Define $g^i(x) = \mu_{\pi(x)}(P_i)$, g^i_δ and g^i_* defined on $\mathbb{R}^n$ as above. Set:

$$g(x) := \sum_i \xi_{P_i}(x) g^i(x), \quad g_\delta(x) := \sum_i \xi_{P_i}(x) g^i_\delta(x)$$

$$\text{and} \quad g_*(x) := \sum_i \xi_{P_i}(x) g^i_*(x).$$

Theorem [cf. Lemma page 61] $\lim_{\delta\to 0} g_d = g$ *μ-a.e. and* $\int -\ln g_* < \infty$.

Proof Convergence follows from LDT applied to each g^i_δ. For the estimate, we write (compare with pages 16/17):

$$\int -\ln g_* d\mu = \int_0^\infty \mu(g_* < e^{-s})ds = \sum_i \int_0^\infty \mu(P_i \cap \{g^i_* < e^{-s}\})ds.$$

We clearly have $\mu(P_i \cap \{g^i_* < e^{-s}\}) \le \mu(P_i)$ and, by the Maximal Lemma b),

$$\mu(P_i \cap \{g^i_* < e^{-s}\}) = \int_{\{g^i_* < e^{-s}\}} g^i d(\pi_* \mu) \le c(n)e^{-s}.$$

It follows that

$$\begin{aligned}\int -\ln g_* d\mu &\le \sum_i \int_0^\infty \min(\mu(P_i), c(n)e^{-s})ds \\ &\le H(P) + \ln c(n) + 1 < \infty. \qquad \square\end{aligned}$$

10 b. Martingale theorems Let $\{\mathcal{F}_n\}_{n\in\mathbb{Z}}$ be an increasing family of σ-algebras in a probability space $(\Omega, \mathcal{A})$, $\mathcal{F}_{-\infty} := \cap_n \mathcal{F}_n, \mathcal{F}_\infty$ the σ-algebra generated by $\cup_n \mathcal{F}_n$. We assume for convenience that there are families of conditional measures μ_ω^n associated with $\mathcal{F}_n$. A process $\{x_n\}_n$, where x_n is a $\mathcal{F}_n$-measurable integrable function, is called a supermartingale if it satisfies, for all $m < n$ and for a.e. ω:

$$x_m(\omega) \geq \int x_n(\omega')\mu_\omega^m(d\omega').$$

A martingale is a process $\{x_n\}_n$ such that $\{x_n\}_n$ and $\{-x_n\}_n$ are supermartingales. L^1-bounded supermartingales converge a.e. as $n \to +\infty$. Supermartingales always converge as $n \to -\infty$ (this is sometimes called reversed martingale convergence). Namely:

Theorem [[D], Theorem II.3.13] *Let* $\{x_n, \mathcal{F}_n\}_n, n \in \mathbb{Z}$ *be a supermartingale such that* $\sup_{n\geq 0}|x_n|$ *is integrable. Then* $\{x_n\}_n$ *converges a.e. as* $n \to +\infty$.

Theorem [[D], Theorem II.3.17] *Let* $\{x_n, \mathcal{F}_n\}_n, n \in \mathbb{Z}$ *be a supermartingale. Then*

1. $\int x_0 \leq \int x_{-1} \leq \cdots \leq \int x_{-n} \leq \cdots$. *Let* $L := \lim_{n\to\infty} \int x_{-n}$,

2. $\lim_{n\to\infty} x_{-n}(\omega) =: x_{-\infty}(\omega)$ *exists in* $(-\infty, +\infty]$ *for a.e.* ω *and*

3. *if* $L < +\infty$, *then* $x_{-\infty} < \infty$ *a.e.,* $\sup_{n\leq 0}|x_n|$ *is integrable, and the convergence in 2) holds in* L^1.

10 c. Other comments

A more detailed account of most of the same material is to come in the Volume 2 of the book by Einsiedler, Lindenstrauss and Ward ([ELW]).

Here, information and entropy theory (Chapters 1,2 and 3) are taken from [Ro2] and [Pa]. See also [KH] pages 161 to 179 and [W]. The entropy was introduced by Shannon [Sha] as a measure of transmission of information. Kolmogorov [K] defined the mean entropy as a new invariant for dynamical systems. The properties of page 27 (Sinai [Si]) allow to compute the entropy of many transformations. BK Theorem, with all variants, comes from [BK]. For Lebesgue spaces and conditional measures, see [Ro1].

Relations between dimension exponents and entropy hold in many different contexts. Even for diffeomorphisms, they are known for much larger families than linear automorphisms, but the presentation is simpler in this context. We state (without proofs) some of the relevant general results in section 8. The proofs are extensions of the proofs we present in the linear case.

The theorem page 50 is due to Young in the general case; the one page 54 to Barreira, Pesin and Schmeling ([BPS], see page 83). The proof presented here follows [LY]. Since the key argument (step 4, pages 62/63) is very succinct in [LY], it is slightly developed here. LS Lemma comes from [LS].

Invariant measures for non-hyperbolic linear automorphisms of the torus have been studied by Lindenstrauss and Schmidt [LiS]. Their result is surprising: if m is an ergodic measure which is not Lebesgue, then the global leaves parallel to W_{u+1} form a measurable partition and the corresponding conditional measures are made of Dirac measures and uniform measures on circles.

Folding entropy was introduced by Ruelle [Ru2] in connection with non-equilibrium statistical mechanics. The theorem page 69 (in the general C^2 case) comes from [Shu]. See [QXZ], [Liu] and [Shu] for background, motivations and history.

Margulis-Ruelle inequality comes from [Ru1]. Details of the proof page 73 are in [L], pages 30–32. Pesin Stable Manifold Theorem comes from [P1], BPS Lemma from [BPS], Appendix, Mañé' Lemma from [M]. Pesin formula page 85 is due to Pesin ([P2]). Mañé has a simpler approach relying on the analysis of $B(x, n, \varepsilon, (\ell(x))^{-1})$. Here Pesin's formula is obtained as a corollary of the dimension formulas, an artificial feature of our presentation.

Berg [B] showed that the Lebesgue measure is the only measure of maximal entropy for linear ergodic endomorphisms of the torus. This is one example of a characterization of some particular invariant measure by a *variational principle.* Here we show a slightly stronger result, that $h_1 = m_1 \ln \lambda_1$ already characterize Lebesgue measure. See [EL] for more relations between partial entropies in the linear case.

The considerations of Chapter 10a are classical and often useful in problems of analysis. See Doob [D] for Chapter 10b.

B K. Berg, Convolutions of invariant measures, maximal entropy, *Math. Systems Theory,* **3** (1969), 146–150.

BK M. Brin & A. Katok, On Local Entropy, *Geometric dynamics (Rio de Janeiro, 1981*, *Springer Lect. Notes Math.*, **1007** (1983), 30–38.

BW K. Burns & A. Wilkinson, A note on stable holonomy between centers, *in preparation,* http://www.math.northwestern.edu/ wilkinso/papers/c1hol0611.pdf

BP L. Barreira & Ya. Pesin. *Nonuniform Hyperbolicity: Dynamics of Systems with Nonzero Lyapunov Exponents,* Encyclopedia of Mathematics and Its Applications, **115** (2007), Cambridge University Press.

BPS L. Barreira, Ya. Pesin & J. Schmeling, Dimension and product structure of hyperbolic measures, *Ann. Math.* **149** (1999), 755–783.

D J.L. Doob, *Classical potential theory and its Probabilistic counterpart,* Classics in Mathematics, Sringer (2001).

EL M. Einsiedler & E. Lindenstrauss, Rigidity properties of $\mathbb{Z}^d$ actions on tori and solenoids, *Electron. Res. Announc. Amer. Math. Soc.* **9** (2003), 99–110.

ELW M. Einsiedler, E. Lindenstrauss & T. Ward, *Entropy in dynamics* , *in preparation*, http://www.uea.ac.uk/menu/acad_depts/mth/entropy/

K A.N. Kolmogorov, A new metric invariant of transitive dynamical systems and automorphisms of Lebesgue spaces, *Doklady Akad Nauk SSSR*, **119** (1958), 861–864 (in Russian).

KH A. Katok & B. Hasselblatt, *Introduction to the Modern Theory of Dynamical Systems,* Encyclopedia of Mathematics and its Applications, **54**, Cambridge University Press, (1995).

L F. Ledrappier, Quelques Propriétés des exposants caractéristiques. *Ecole d'été de Probabilités de Saint-Flour XII - 1982*, **1097** (1984), 305–393.

LS F. Ledrappier & J.-M. Strelcyn, A proof of the estimation from below in Pesin's entropy formula, *Ergod. Th. Dynam. Sys.*, **2** (1982), 203–219.

LY F. Ledrappier & L.-S. Young, The metric entropy of diffeomorphisms. Part I and II, *Ann. Math.* **122** (1985), 509–539 and 540–574.

LiS E. Lindenstrauss & K. Schmidt, Invariant sets and mesasures of non expansive group automorphisms, *Israel J. Math.*, **144** (2004), 29–60.

Liu P.-D. Liu, Ruelle inequality relating entropy, folding entropy and negative Lyapunov exponents, *Commun. Math. Phys.*, **240** (2003), 531–538.

M R. Mañé, A proof of Pesin formula, *Ergod. Th. Dynam. Sys.* **1** (1981), 95–102.

O V. I. Oseledets, A multiplicative ergodic theorem. Liapunov characteristic numbers for dynamical systems. *Transactions Moscow Math. Soc.*, **19** (1968), 197–221.

P1 Ya. Pesin, Families of invariant manifolds corresponding to non-zero characteristic exponents, *Mathematics USSR, Isz.*, **10: 6** (1976), 1261–1305.

P2 Ya. Pesin, Characteristic exponents and smooth ergodic theory, *Russian math. surveys,* **32:4** (1977), 55–114.

QXZ M. Qian, J.-S. Xie & S. Zhu. *Smooth Ergodic Theory for Endomorphisms*. Lect. Not. Math. **1978**, Springer-Verlag, 2009.

Pa W. Parry, *Entropy and generators in ergodic theory,* W.A. Benjamin Inc. New York - Amsterdam (1969)

Ro1 V.A. Rokhlin, On the fundamental ideas of measure theory, *Translations A.M.S. (series 1)* **10** (1962), 1–54.

Ro2 V.A. Rokhlin, Lectures on the entropy of measure preserving transformations, *Russian math. surveys* **22** (1967), 1–52.

Ru1 D. Ruelle, An inequality for the entropy of differentiable maps, *Boletim Soc. Mat. Bras.*, **9** (1978), 83–87.

Ru2 D. Ruelle, Positivity of the entropy production in non-equilibrium statistical mechanics, *J. Stat. Phys.*, **85** (1996) 1–23.

Sha C. Shannon, The mathematical theory of communication, *Bell system Technical J.* **27** (1948), 379–423, 623–656.

Shu L. Shu, Dimension theory for invariant measures of endomorphisms, *preprint*.

Si Ya. Sinai, On the concept of entropy fo ra dynamic system, *Dokl. Akad. Nauk. SSSR*, **124** (1959) 768–771.

W P. Walters, *An introduction to ergodic theory,* Springer Graduate texts in mathematics, **79** (2000).