NOTES OF THE COURSE ON CHAOTIC DYNAMICAL SYSTEMS

STÉPHANE NONNENMACHER

The aim of this course is to present some properties of low-dimensional dynamical systems, particularly in the case where the dynamics is "chaotic". We will describe several aspects of "chaos", by introducing various modern mathematical tools, allowing us to analyze the long time properties of such systems. Several simple examples, leading to explicit computations, will be treated in detail.

Here are some topics I plan to deal with in these notes. They do not directly correspond to the final table of contents.

(1) Definition of a dynamical system: flow generated by a vector field, discrete time transformation. Poincaré sections vs. suspended flows. Examples: Hamiltonian flow, geodesic flow, transformations on the interval or the 2-dimensional torus.
(2) Ergodic theory: long time behavior. Statistics of long periodic orbits. Probability distributions invariant through the dynamics (invariant measures). "Physical" invariant measure.
(3) Chaotic dynamics: instability (Lyapunov exponents) and recurrence. From the hyperbolic fixed point to Smale's horseshoe.
(4) Various levels of "chaos": ergodicity, weak and strong mixing.
(5) Symbolic dynamics: subshifts on 1D spin chains. Relation (semiconjugacy) with expanding maps on the interval.
(6) Uniformly hyperbolic systems: stable/unstable manifolds. Markov partitions: relation with symbolic dynamics. Anosov systems. Example: Arnold's "cat map" on the 2-dimensional torus.
(7) Complexity theory. Topological entropy, link with statistics of periodic orbits. Partition functions (dynamical zeta functions). Kolmogorov-Sinai entropy of an invariant measure.
(8) Exponential mixing of expanding maps: spectral analysis of some transfer operator. Perron-Frobenius theorem.

Date: 05 november 2009.

Contents

1. What is a dynamical system?

A discrete-time dynamical system (DS) is a transformation rule (function) f on some phase space X, namely a rule

$$X \ni x \mapsto f(x) \in X.$$

The iterates of f will be denoted by $f^n = f \circ f \circ \cdots \circ f$, with time $n \in \mathbb{N}$. The map f is said to be *invertible* if f is a bijection on X (or at least on some subset of it). One can then consider positive and negative iterates: $f^{-n} = (f^{-1})^n$.

A continuous-time dynamical system is a family $(\phi^t)_{t\in\mathbb{R}^+}$ of transformations on X, such that $\phi^t \circ \phi^s = \phi^{t+s}$. If it is invertible (for any $t > 0$), then it is a flow $(\phi^t)_{t\in\mathbb{R}}$.

Very roughly, the dynamical systems theory aims at understanding the long-time asymptotic properties of the evolution through f^n or ϕ^t. For instance:

(1) How numerous are the periodic points ($x \in X$ such that $f^T x = x$ for some $T > 0$). Where are they located on X? Are there more complicated forms of *recurrence*.
(2) More generally, what are the nontrivial *invariant subsets* of X? ($X' \subset X$ is invariant if $f(X') \subset X'$).
(3) is there an invariant probability measure for the map f? (that is a measure μ on X, such that $\mu(A) = \mu(f^{-1}(A))$ for "any" set A). What are the *statistical* properties of the DS w.r.to this measure?
(4) Do small perturbations of f have the same global properties as f? Are they conjugate with f? Is the DS f *structurally stable*?

One would like to *classify* all possible behaviours, that is group the maps f among various equivalence classes.

Definition 1.1. A map $g : Y \to Y$ is *semiconjugate* with f iff there exists an surjective map $\pi : Y \to X$ such that $f \circ \pi = \pi \circ g$. The map f is then called a factor of g. If π is invertible, then f, g are conjugate (isomorphic). One can often analyze a map f by finding a better-understood g of which it is a factor.

In general, the phase space X and the transformation f have some extra structure:

(1) X can be metric space (equipped with a distance function $d(x, y)$), with an associated topology (family of open/closed sets). It is then natural to consider maps f which are *continuous* on X. This is the realm of **topological dynamics**. We will moslty restrict ourselves to X a *compact* (bounded and closed) set.
(2) X can be (part of) a Euclidean space $\mathbb{R}^d$ or a smooth manifold. The map f can then be differentiable, that is near each point x it be approximated by the linear map $df(x)$ sending the tangent space $T_x X$ to $T_{f(x)}X$. This is the realm of **smooth dynamics**. A differentiable flow is generated by a vector field

$$v(x) = \frac{d}{dt}\phi^t(x)|_{t=0} \in T_x X,$$

Generally one starts from the vector field $v(x)$, the flow ϕ^t being obtained by integrating over that field: one notes formally $\phi^t(x) = e^{tv}(x)$. Most physical dynamical systems are of this type.
(3) X can be a measured space, that is it is equipped with a σ-algebra and a measure μ on it[1]. It is then natural to consider transformations which leave μ invariant. This is the realm of **ergodic theory.**
(4) One can then add some other structures. For instance, a metric on X (geometry) is preserved iff f is an isometry. A symplectic structure on X is preserved if f is a canonical (or symplectic) transformation:

[1]A σ-algebra on X is a set $\mathcal{A} = \{A_i\}$ of subsets of X, which is closed under complement and countable union, and contains X. On a topological space X the most natural one is the Borel σ-algebra, which contains all the open sets. A measure μ is a nonnegative function on $\mathcal{A}$ such that $\mu(\bigcup_i A_i) = \sum_i \mu(A_i)$ if the A_i are disjoint. μ is a probability measure if $\mu(X) = 1$.

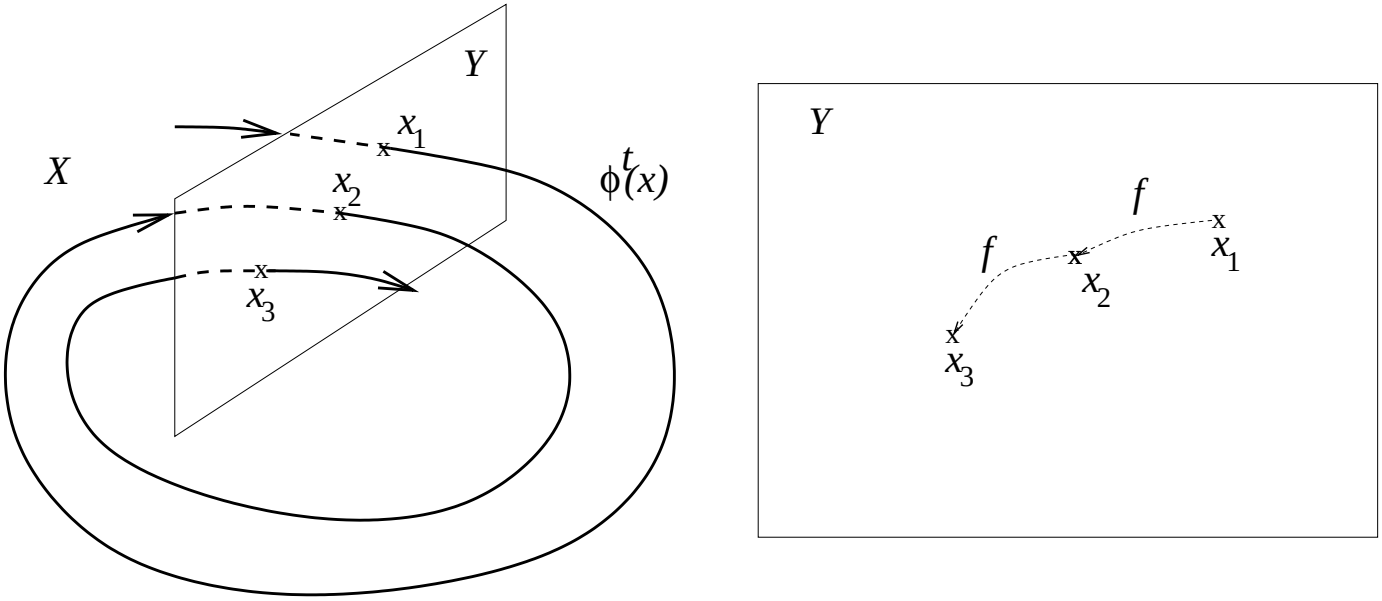

FIGURE 1.1. A Poincaré section Y for the flow ϕ^t on X, and the associated Poincaré map f.

this is the realm of **Hamiltonian/Lagrangian dynamics**, of classical point mechanics. A complex structure is preserved if f is holomorphic. **(complex dynamics**).

These extra structures may also be imposed to the (semi-)conjugacy between two DS. This question is less obvious than it first appears: it appears that requiring smooth conjugacy between two smooth DS is "too restrictive a" condition, as opposed to the notion of continuous (topological) conjugacy. This motivates the following

Definition 1.2. A continuous map $f : X \to X$ on a smooth manifold X is called (C^1-) *structurally stable* if there exists $\epsilon > 0$ such that, for any perturbation $\tilde{f} = f + \delta f$ with $\|\delta f\|_{C^1} \leq \epsilon$, then f and $\tilde{f}$ are topologically conjugate (i.e., there exists a homeomorphism $h : X \to X$ such that $\tilde{f} = h^{-1} \circ f \circ h$).

From maps to flows, and back. In these notes we will mostly consider discrete-time maps $f : X \to X$. From such a map one can easily construct a flow through a *suspension* procedure. Namely, we select a positive function $\tau : X \to \mathbb{R}^+$, called the ceiling function, or first return time. We then consider the product space

$$X_\tau = \left\{(x,t) \in X \times \mathbb{R}^+,\ 0 \leq t \leq \tau(x)\right\}, \quad \text{with the identification } (x,\tau(x)) \equiv (f(x),0).$$

One can easily define a semiflow ϕ^t on X_τ: starting from (x,t_0), take $\phi^t(x,t_0) = (x,t_0+t)$ until $t_0 + t = \tau(x)$, then jump to $(f(x),0)$ and so on. ϕ^t is called a suspended semiflow of the map f. If f is invertible, then ϕ^t is as well (and is a suspended flow). Many dynamical properties of the map f are inherited by the semiflow ϕ^t.

Conversely, a (semi)flow $\phi^t : X \to X$ can often by analyzed through a *Poincaré section*, which is a subset $Y \subset X$ with the following property: for each $x \in X$, the orbit $(\phi^t(x))_{t>0}$ will intersect Y in the future at a discrete set of times. The first time of intersection $\tilde{\tau}(x) > 0$, and the first point of intersection $\tilde{f}(x) \in Y$. Restricting $\tilde{\tau}$, $\tilde{f}$ to Y, we have "summarized" the flow ϕ^t into the first return (Poincaré) map $f : Y \to Y$ and the first return time $\tau : Y \to \mathbb{R}^+$. If X is a n-dimensional manifold, Y is generally a collection of $(n-1)$-dimensional submanifolds, transverse to the flow. Many properties of the flow are shared by f.

2. A gallery of examples

In order to introduce the various concepts and properties, we will analyze in some detail some simple DS, mostly in low (1 or 2) dimension. These examples will already provide a large variety of dynamical behaviours.

2.1. **Contracting map.** A map f defined on a metric space (X, d) is contracting iff for some $0 < \lambda < 1$ one has

$$\forall x, y \in X, \quad d(f(x), f(y)) \leq \lambda\, d(x, y).$$

As a consequence, the iterates of any pair x, y satisfy $d(f^n(x), f^n(y)) \overset{n\to\infty}{\to} 0$.

The **contraction mapping principle** implies that f admits a *unique fixed point* $x_0 \in X$, which is an **attractor**:

$$\forall x \in X, \quad f^n(x) \overset{n\to\infty}{\to} x_0.$$

The **basin** of this attractor is the full space X.

Definition 2.1. On the opposite, a map is said to be **expanding** iff there exists $\mu > 1$ such that, for any close enough point x, y, one has $d(f(x), f(y)) \geq \mu\, d(x, y)$.

2.2. **Linear maps on $\mathbb{R}^d$.** Let $f = f_M : \mathbb{R}^d \circlearrowleft$ be given by an invertible matrix $M \in GL(d, \mathbb{R})$: $f(x) = Mx$. The origin is always a fixed point. What kind of fixed point is it? Is it the only fixed point? The answer depends on the *spectrum* of M. For a real matrix, eigenvalues are either real, or come in pairs $(\lambda, \bar{\lambda})$. Call E_λ the generalized eigenspace (resp. the union of the two generalized eigenspaces of $\lambda, \bar{\lambda}$). These eigenspaces can be grouped into

$$\mathbb{R}^d = E^0 \oplus E^- \oplus E^+, \quad \begin{cases} E^0 & = \bigoplus_{|\lambda|=1} E_\lambda \quad \text{neutral subspace} \\ E^- & = \bigoplus_{|\lambda|<1} E_\lambda \quad \text{stable/contracting subspace} \\ E^+ & = \bigoplus_{|\lambda|>1} E_\lambda \quad \text{unstable/expanding subspace} \end{cases}$$

These 3 subspaces are invariant through the map. The stable subspace E^- is characterized by an exponential contraction (in the future): for some $0 < \mu < 1$,

$$x \in E^- \iff \|M^n x\| \leq C\mu^n \|x\|,\ n > 0.$$

The unstable subspace E^+ is NOT made of the points which escape to infinity, but of the points converging exponentially fast to the origin *in the past*:

$$x \in E^+ \iff \|M^n x\| \leq C\mu^{|n|} \|x\|,\ n < 0.$$

(1) if $E^0 = E^+ = \{0\}$, that is the eigenvalues of M satisfy $\max |\lambda_i| \overset{\text{def}}{=} r(M) < 1$, then 0 is an *attracting fixed point*. Eventhough one may have $\|Mx\| > \|x\|$, the higher iterates satisfy $\|M^n\| \leq C(r(M) + \epsilon)^n$, so the map is *eventually contracting*. The contraction may be faster along certain directions than along others.
(2) on the opposite, if $E^0 = E^- = \{0\}$ the origin is a *repelling fixed point*.
(3) if $E^0 = \{0\}$ but $E^- \neq \{0\}$ and $E^+ \neq \{0\}$, then the map is *hyperbolic*; the origin is called a *hyperbolic fixed point*.
(4) if $E^0 \neq \{0\}$, there exists eigenspaces associated with neutral eigenvalues $|\lambda| = 1$. This leads to the study of rotations on S^1 (see next subsection).

Remark 2.2. The study of linear maps already provides some hints on necessary conditions for a general map to be structurally stable. Inside $GL(n, \mathbb{R})$, contracting/hyperbolic matrices form an open set, meaning that for

each contracting/hyperbolic matrix M, small perturbations $M+\delta M$ will still be contracting/hyperbolic with the same number of unstable/stable directions.

2.3. **Circle rotations.** Let us consider a simple diffeomorphism of the unit circle $S^1 \simeq [0,1)$ which preserves orientation: a rotation by an "angle" $\alpha \in [0,1)$:

$$x \in S^1 \mapsto f(x) = f_\alpha(x) = x + \alpha \bmod 1.$$

This is an isometry. The long-time dynamics qualitatively depends on the value of α:

(1) if $\alpha = \frac{p}{q} \in \mathbb{Q}$, every point is q-periodic.
(2) if $\alpha \notin \mathbb{Q}$, every orbit $\mathcal{O}(x) = \{f^n(x),\, n \in \mathbb{Z}\}$ is dense in S^1. Hence, there is no periodic orbit, but every point x will come back arbitrary close to itself in the future. This is a form of *nontrivial recurrence*. In particular, the only closed invariant subset of S^1 is S^1 itself: this DS is *minimal*. We will also see that this irrational rotation admits a unique invariant probability measure, namely is the Lebesgue measure on S^1: such a map is called *uniquely ergodic*.

Remark 2.3. Any irrational rotation can be approached arbitrarily close (in any C^ktopology) by a rational rotation (and vice-versa). Hence, rotations on S^1 are *not* structurally stable.

2.4. **Expanding maps on the circle.** An interesting smooth noninvertible map on S^1 is the dilation by an positive integer $m \in \mathbb{N}$:

$$S^1 \ni x \mapsto E_m(x) = mx \bmod 1.$$

This map has (topological) degree m: it winds m times around the circle. As opposed to the rotations, the map is *expanding*: for any nearby x, y, one has $d(E_m(x), E_m(y)) = m\, d(x,y)$. Its iterates have the same form: $E_m^n = E_{m^n}$.

Each (small enough) interval I has m disjoint preimages, each of them of length $\frac{|I|}{m}$. As a result, the map E_m preserves the Lebesgue measure on S^1.

E_m has exactly $m-1$ fixed points

$$x_k = \frac{k}{m-1}, \quad k \in \{0, \dots, m-2\}\,.$$

This can be deduced from the study of the lift $\tilde{E}_m$ of E_m on $\mathbb{R}$: the graph of $\tilde{E}_m$ on $[0,1)$ intersects exactly $m-1$ times the shifted diagonals.

Similarly, E_{m^n} has exactly $m^n - 1$ points of period n. The full (countable) set of periodic points is dense on S^1.

2.4.1. *Semiconjugacy of E_m with symbolic dynamics.* The study of these periodic points, and of other dynamical properties, is facilitated when one notices the semiconjugacy between E_m and a simple **symbolic shift**. Consider Σ_m^+ the set of one-sided symbolic sequences $\boldsymbol{x} = x_1 x_2 \cdots$ on the alphabet $x_i \in \{0, 1, \dots, m-1\}$. Each sequence $(x_i) \in \Sigma_m^+$ is naturally associated with a real point $x \in [0,1]$ via the base-m decomposition:

$$\boldsymbol{x} = (x_i)_{i\geq 1} \mapsto \pi(\boldsymbol{x}) = 0 \cdot x_1 x_2 x_3 \cdots = \sum_{i=1}^{\infty} \frac{x_i}{m^i} = x. \tag{2.1}$$

One can easily check that π semiconjugates E_m with the one-sided shift σ on Σ_m^+ (see def. 1.1):

$$E_m \circ \pi = \pi \circ \sigma, \qquad \sigma((x_i)_{i\geq 1}) = (x_{i+1})_{i\geq 1}.$$

This property can be represented by the following commuting diagram:

$$\begin{array}{ccc} \Sigma_2^+ & \stackrel{\sigma}{\to} & \Sigma_2^+ \\ \pi \downarrow & & \pi \downarrow \\ S^1 & \stackrel{E_m}{\to} & S^1 \end{array} \tag{2.2}$$

One says that f is a factor of the shift Σ_2^+. It inherits most of its topological complexity. The defect from being a full conjugacy is due to the (countably many) sequences of the type $x_1x_2\cdots x_n1000\cdots \equiv x_1x_2\cdots x_n0111\cdots$. This defect of injectivity is not very significant when counting periodic points: σ has m^n points of period n, that is one more than E_m (the difference is due to the fixed points $\overline{0} \stackrel{\text{def}}{=} 00000 \equiv \overline{1}$).

It is convenient to equip Σ_m^+ with a (ultrametric) distance function:

$$d(\boldsymbol{x},\boldsymbol{y}) \stackrel{\text{def}}{=} \lambda^{\min\{i,\, x_i\neq y_i\}}, \quad \text{for some } \ 0<\lambda<1.$$

This induces a topology on Σ_m^+, for which the open sets are unions of *cylinders*

$$C_{\epsilon_1\epsilon_2\cdots\epsilon_n} = \left\{\boldsymbol{x}\in\Sigma_m^+ \,|\, x_1=\epsilon_1,\ldots,x_n=\epsilon_n\right\}.$$

The semiconjugacy π is then a continuous (actually, Hölder-continuous) map $\Sigma_m^+ \to S^1$. Through this semiconjugacy, one can easily construct

(1) the periodic points of E_m: an n-periodic point is the image of a n-periodic sequence $\boldsymbol{x} = \overline{x_1x_2\cdots x_n}$.
(2) dense orbits on S^1. (Hint: construct a sequence containing all finite words)
(3) nontrivial (fractal) closed invariant sets. Ex: the 1/3-Cantor set is invariant for E_3, image through π of the subset of sequences $\{(x_i)_{i\geq 1},\ x_i\in\{0,2\}\}$.
(4) nontrivial (fractal) invariant measures. Ex: the push-forward of *Bernouilli measures* on Σ_m^+ (see §4.5.3).

2.4.2. *Structural stability of E_m.* For the linear map E_m one is able to explicitly construct a homeomorphism relating E_m with a C^1 perturbation g_m. Actually, the construction can be done for any expanding map g of topological degree m. The construction proceeds by re-interpreting the semiconjugacy between E_m and Σ_m^+ in terms of a partition of the circle, and then extend this construction to the nonlinear maps g.

The construction of the semiconjugacy (2.1) can be made by introducing a partition of $[0,1]$ into m intervals $\Delta_j = [\frac{j}{m},\frac{j+1}{m}]$, $j=0,\ldots,m-1$. Notice that each such rectangle satisfies $E_m(\Delta_i)=[0,1]$, and the correspondence is 1-to-1. This partition can be refined through the map: for any sequence $\alpha_1\cdots\alpha_n$ we define the set

$$\Delta_{\alpha_1\cdots\alpha_n} = \bigcap_{j=1}^{n} E_m^{-j+1}(\Delta_{\alpha_j}).$$

Notice that E_m^n maps each $\Delta_{\alpha_1\cdots\alpha_n}$ to $[0,1]$ bijectively. Each $\Delta_{\alpha_1\cdots\alpha_n}$ is an interval of the form $[\frac{k}{m^n},\frac{k+1}{m^n}]$, which consists of the points $x\equiv 0,\alpha_1\cdots\alpha_n * **$. This interval is therefore the image of the cylinder $C_{\alpha_1\cdots\alpha_n}$ through π.

Let us now consider an expanding map $g: S^1\to S^1$ of degree $m>1$, and assume (by shifting the origin of S^1) that $g(0)=0$. From the monotonicity of g, one can split $[0,1]$ into m subintervals $\Gamma_0,\ldots,\Gamma_{m-1}$, such that $g(\Gamma_i)=[0,1]$ in a 1-to-1 correspondence. Since g^n is also monotonic and has degree m^n, we can similarly split $[0,1]$ between m^n subintervals $\{\Gamma_{\alpha_1\cdots\alpha_n},\ \alpha_i\in\{0,\ldots,m-1\}\}$; these can also be defined by $\Gamma_{\alpha_1\cdots\alpha_n} = \bigcap_{j=1}^n g^{-j+1}(\Gamma_{\alpha_j})$. The expanding character of g ensures that the lengths of these intervals decreases exponentially with n, so for each inifinite sequence $\boldsymbol{\alpha}$, the intersection $\bigcap_{j\geq 1} g^{-j+1}(\Gamma_{\alpha_j})$ consists in a single point $x=\tilde{\pi}(\boldsymbol{\alpha})\in[0,1]$. We have thus obtained a semiconjugacy between g and Σ_2^+ similar with that between E_m and Σ_2^+.

The maps $\pi, \tilde\pi$ are not invertible, so we cannot directly write down an expression of the form $\pi \circ \tilde\pi^{-1}$. However, the defect of injectivity for π and $\tilde\pi$ is exactly of the same form: it comes from the boundaries of the cylinders $C_{\alpha_1 \cdots \alpha_n}$, mapped by π (resp. $\tilde\pi$) to the intervals $\Delta_{\alpha_1 \cdots \alpha_n}$ (resp. $\Gamma_{\alpha_1 \cdots \alpha_n}$). For any point $x \in S^1$ which is not on the boundary of any interval $\Gamma_{\boldsymbol{\alpha}}$, the preimage $\tilde\pi^{-1}(x) \in \Sigma_m^+$ is unique, so we may define $h(x) \stackrel{\text{def}}{=} \pi(\tilde\pi^{-1}(x))$. On the other hand, if x is the left boundary of some interval $\Gamma_{\boldsymbol{\alpha}}$, we set $h(x)$ to be the left boundary of the corresponding interval $\Delta_{\boldsymbol{\alpha}}$. One check that the map h is well-defined, bijective and is bicontinuous on S^1, and that it satisfies

$$E_m \circ h = h \circ g. \tag{2.3}$$

It thus topologically conjugates E_m with g.

2.4.3. *A variation on the proof of semiconjugacy between E_m and g.* The semiconjugacy equation (2.3) can be solved (the unknown being the map h) by rewriting this equation in terms of a contracting map acting on some appropriate functional space. Consider the space $\mathfrak{C}$ of continuous maps $h : [0,1] \circlearrowleft$ such that $h(0) = 0$, $h(1) = 1$, endowed with the metric $d(h_1, h_2) = \max_x |h_1(x) - h_2(x)|$. We define the following map on $\mathfrak{C}$:

$$\mathcal{F}h(x) \stackrel{\text{def}}{=} \frac{h(g(x)) + j}{m} \text{ if } x \in \Gamma_j,\ j = 0, \ldots, m-1.$$

This amounts to applying the j-th branch of E_m^{-1} on each interval Γ_j, so that $E_m \circ \mathcal{F}h = h \circ g$. It is easy to check that $\mathcal{F}h \in \mathfrak{C}$ (one only needs to check it at the boundaries of the Γ_j). The main property of this map is the following contraction:

$$\forall h_1, h_2 \in \mathfrak{C}, \quad d(\mathcal{F}h_1, \mathcal{F}h_2) \leq \frac{1}{m}\, d(h_1, h_2).$$

The map $\mathcal{F}$ is therefore contracting on $\mathfrak{C}$, and has thus a single fixed point h_0 (which can be obtained by iterating $\mathcal{F}$ infinitely many times). The equation $\mathcal{F}h_0 = h_0$ is obviously equivalent with the semiconjugacy (2.3).

To prove that h is 1-to-1 provided g is expanding, one constructs a semiconjugacy in the other direction ($g \circ \tilde h = \tilde h \circ E_m$) using a similar map $\tilde{\mathcal{F}}$. In that case, the contraction constant for the map $\tilde{\mathcal{F}}$ is given by $\lambda = \max_x g'(x)^{-1} < 1$. The above equation admits a unique solution $\tilde h_0$. One has therefore $E_m \circ h_0 \circ \tilde h_0 = h_0 \circ \tilde h_0 \circ E_m$ for a map $h_0 \circ \tilde h_0$ of degree 1. It is easy to show that one must have $h_0 \circ \tilde h_0 = Id$.

2.5. **More on symbolic dynamics: subshifts.** We have considered the set of one-sided infinite sequences $(x_i)_{i \geq 1}$ on m symbols. One can also let the shift σ act on *two-sided* sequences $(x_i)_{i \in \mathbb{Z}}$. The space of bi-infinite sequences is denoted by Σ_m, and we denote by (Σ_m, σ) the corresponding DS. As opposed to the one-sided shift, the two-sided one is an invertible, bicontinuous map. It has the same number m^n of n-periodic points.

A *subshift* of Σ_m (or Σ_m^+) is a closed, shift-invariant subset $\Sigma \subset \Sigma_m$ (or $\Sigma \subset \Sigma_m^+$). Ex: the 1/3 Cantor set was the image of a subshift of Σ_3^+, made of all sequences containing no symbol $x_i = 1$. This subshift is obviously isomorphic with the shift Σ_2^+.

It is more interesting to consider subshifts defined by forbidding certain combinations of successive symbols. Among this type of subshifts (called subshifts of finite type) we find the **topological Markov chains**. Such a chain is defined by an $m \times m$ matrix $A = (A_{kl})$ with entries given by 0 or 1, called an *adjacency matrix*. A pair $x_i x_{i+1}$ is said to be allowed iff $A_{x_i x_{i+1}} = 1$. The subshift $\Sigma_A^{(+)} \subset \Sigma_m^{(+)}$ is made of all the sequences (x_i) such that all successive pairs $x_i x_{i+1}$ are allowed. (Check that $\Sigma_A^{(+)}$ is a closed, shift-invariant set).

The DS $(\Sigma_A^{(+)}, \sigma)$ is relatively simple to analyze, because its properties are encoded in the $m \times m$ adjacency matrix A. The latter can be conveniently represented by a *directed graph* Γ_A on m vertices: each sequence $(x_i)_i \in \Sigma_A$ corresponds to a trajectory on the graph.

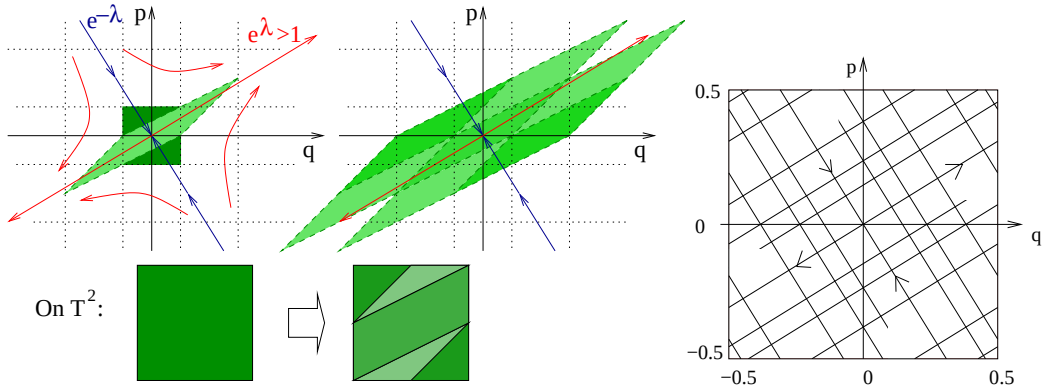

FIGURE 2.1. Arnold's cat map, defined by projecting on $\mathbb{T}^2$ the linear map M. Right: unstable and stable manifolds of the origin.

More generally, a subshift of type k is defined by allowing certain $k+1$-words among $\{0,\dots,m-1\}^{k+1}$, that is certain combinations $x_i x_{i+1}\cdots x_{i+k}$.

Exercise 2.4. Each subshift of type k is conjugate to a certain shift of type 1 (i.e. a topological Markov chain).

Hint: change the alphabet.

Exercise 2.5. Count the number of n-words in Σ_A starting with $x_1 = \epsilon_1$ and ending with $x_n = \epsilon_n$. Count the number of n-periodic points of Σ_A.

Some relevant properties of adjacency matrices. Let A be an $m \times m$ matrix with nonnegative entries. If for any pair (k,l) there exists $n > 0$ such that $(A^n)_{kl} > 0$, then A is called *irreducible*. It means that in the directed graph Γ_A, there exists a path between any pair of vertices (k,l).

A is called *primitive* if there exists $N > 0$ such that all entries $(A^N)_{kl} > 0$. Notice that the same property then holds for any $n \geq N$. It means that for any $n \geq N$, any pair (k,l) of vertices can be connected by a path of length n.

Theorem 2.6. *[Perron]*

Let A be a primitive *$m \times m$ matrix with nonnegative entries. Then A has a positive eigenvalue λ with the following properties:*

i) λ is simple and every other eigenvalue satisfies $|\lambda'| < \lambda$

ii) λ has a positive associated eigenvector v (that is, all components $v_i > 0$), no eigenvector of A associated with an eigenvalue $\lambda' \neq \lambda$ can be only non-negative.

2.6. **Hyperbolic torus automorphisms ("Arnold's cat map").** A linear automorphism on $\mathbb{T}^d = \mathbb{R}^d/\mathbb{Z}^d$ is given by projecting on $\mathbb{T}^d$ the linear map induced on $\mathbb{R}^d$ by a matrix $M \in GL(n,\mathbb{Z})$ (matrix with integer coefficients and $\det M = \pm 1$). The dynamics is simply $\mathbb{T}^d \ni x \mapsto f(x) = Mx \bmod 1$. This map is invertible and smooth.

The automorphism is said to be hyperbolic iff the matrix M is so. Let us restrict ourselves to the dimension $d = 2$, the spectrum is of the form (λ, λ^{-1}), $\lambda \in \mathbb{R}$, $|\lambda| > 1$. The simplest example is given by Arnold's "cat" map (see fig. 2.1)

$$M_{cat} = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}, \quad \lambda = \frac{3+\sqrt{5}}{2}.$$

The corresponding eigenspaces are called $E^{\pm}$.

At each point $x \in \mathbb{T}^2$ the tangent map $df(x) = M$, so all tangent spaces can be decomposed into $T_x\mathbb{T}^2 = E^+_x \oplus E^-_x$, where the unstable/stable subspaces $E^{\pm}_x = E^{\pm}$ are independent of x. They are invariant through the map: $df(x)E^{\pm}_x = E^{\pm}_{f(x)}$. The tangent map $df(x)$ acts on E^-_x (resp. E^+_x) by a contraction (resp. a dilation). We will see later tha these properties define an Anosov diffeomorphism (Definition 6.3).

For each $x \in \mathbb{T}^2$, the projected line $W^-(x) \stackrel{\text{def}}{=} x + E^- \bmod 1$ is called the *stable manifold* of x. It is made of all points $y \in \mathbb{T}^2$ such that $d(f^n(x), f^n(y)) \xrightarrow{n\to+\infty} 0$. Similarly, the projected line $W^+(x) \stackrel{\text{def}}{=} x + E^+ \bmod 1$ is called the *unstable manifold* of x. It is made of all points $y \in \mathbb{T}^2$ such that $d(f^n(x), f^n(y)) \xrightarrow{n\to-\infty} 0$ (see fig. 2.1).

The spitting of $\mathbb{T}^2$ between stable manifolds (or *leaves*) is called the stable foliation. This foliation is invariant: $f(W^-(x)) = W^-(f(x))$.

Exercise 2.7. Show that each stable manifold W^-_x is *dense* in $\mathbb{T}^2$.

Periodic points are given by all rational points, in particular they are dense on $\mathbb{T}^2$.

Exercise 2.8. Count the number of n-periodic points for a hyperbolic automorphism A on $\mathbb{T}^2$.

Due to the property $|\det(M)| = 1$, the automorphism f leaves invariant the Lebesgue measure on $\mathbb{T}^2$ (it is an area-preserving diffeomorphism). Later we will show that f is *ergodic* w.r.to this measure.

2.6.1. *Markov partition for Arnold's cat map.* In this section we will construct a semiconjugacy between Arnold's cat map and a specific topological Markov chain. The construction is less obvious than in the case of the dilations on S^1 (see §2.4). It requires the construction of a Markov partition of $\mathbb{T}^2$, performed as follows (see fig. 2.2).

One first defines two rectangles $R_1, R_2 \subset \mathbb{T}^2$ on the torus, with sides given by some stable or unstable segments. The intersections of the rectangles with their images under f produce 5 connected subrectangles $\Delta_1, \dots, \Delta_5$. By construction, the images of the stable sides of Δ_i are contained in the stable sides of some Δ_j, while the backwards images of the unstable sides of Δ_i are contained in the unstable sides of some Δ_j: the rectangles Δ_i thus form a **Markov partition** of $\mathbb{T}^2$ (see §2.8.1 for a general definition).

We may define an adjacency matrix A_{ij} through the condition $A_{ij} = 1$ iff $f(\Delta_i) \cap \Delta_j$ has nonempty interior. From the above picture, we see that $A = \Delta_1 \cup \Delta_2 \cup \Delta_3$, $f(A) = \Delta_1 \cup \Delta_3 \cup \Delta_4$, and the image of any of the first ones intersects any of the second ones. Similarly, $B = \Delta_4 \cup \Delta_5$, $f(B) = \Delta_2 \cup \Delta_5$. We thus obtain the following ajacency matrix:

$$A = \begin{pmatrix} 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 \end{pmatrix}.$$

From this matrix we construct the (two-sided) topological Markov chain (Σ_A, σ). The Markov property of the partition ensures that for any sequence $\boldsymbol{\alpha} \in \Sigma_A$, the set $\bigcap_{j\in\mathbb{Z}} f^{-j}(\Delta_{\alpha_j})$ is not empty. If the Δ_i were disjoint (see the case of Smale's horseshoe), this set would reduce to a single point. In the present case, one has to take a little care of the boundaries $\partial\Delta_i$, and rather consider the set $\Delta_{\boldsymbol{\alpha}} \stackrel{\text{def}}{=} \bigcap_{n\geq 1} \overline{int\left(\bigcap_{|k|\leq n} f^{-k}(\Delta_k)\right)}$. This set

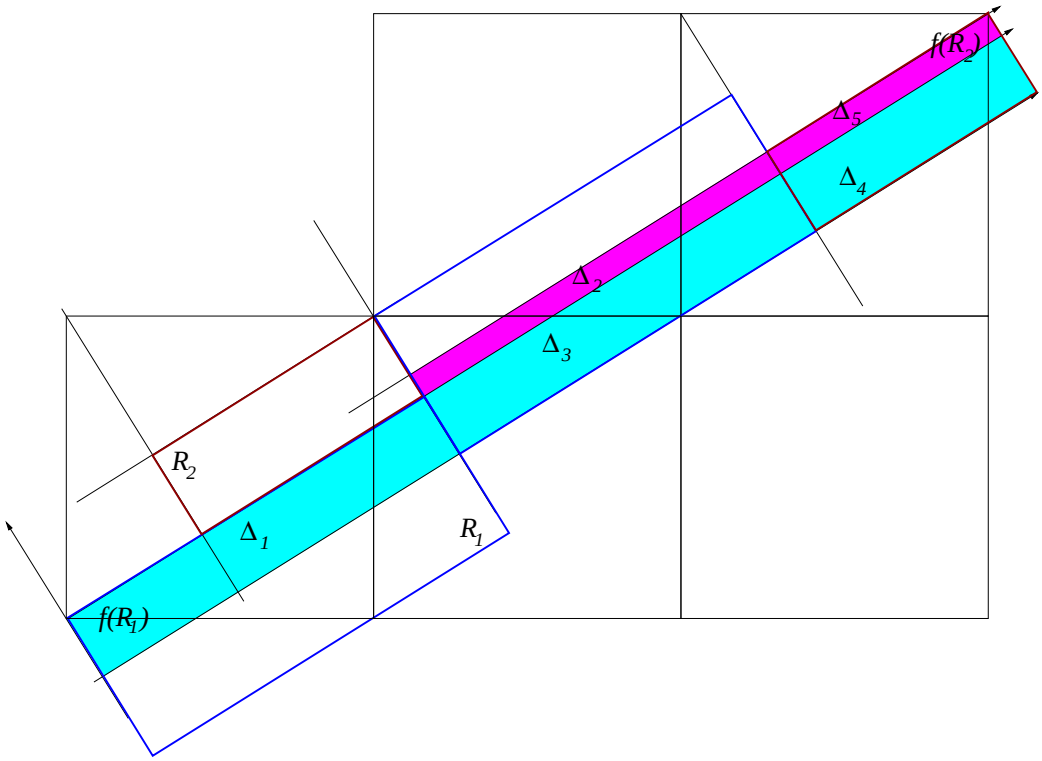

FIGURE 2.2. Adler and Weiss's Markov partition for Arnold's cat map. Two copies of the rectangles R_1, R_2 are shown (thick blue and red lines). Their images $f(R_1)$, $f(R_2)$ are the long rectangles (filled, light blue and pink). The intersections of the latter and the former provide the 5 rectangles $\Delta_1, \dots, \Delta_5$ defining the Markov partition.

consists in a single point, which we denote by $x(\boldsymbol{\alpha}) = \pi(\boldsymbol{\alpha})$. Hence we have obtained a semiconjugacy between the subshift (Σ_A, σ) and f:

$$\begin{array}{ccc} \Sigma_A & \xrightarrow{\sigma} & \Sigma_A \\ \pi \downarrow & & \pi \downarrow \\ \mathbb{T}^2 & \xrightarrow{f} & \mathbb{T}^2 \end{array}$$

One easily checks that all elements of A^2 are positive, showing that A is primitive. One consequence is that the subshift Σ_A (and therefore its factor f) is *topologically mixing* (see Definition 3.13).

Remark 2.9. The Perron-Frobenuis eigenvalue of A is exactly given by $\lambda = \frac{3+\sqrt{5}}{2}$. This is consistent with the fact that the number of periodic orbits for Σ_A has the same exponential growth rate as the number of periodic orbits for f.

2.6.2. *Structural stability of hyperbolic torus automorphisms.* We are interested in small C^1-perturbations of the hyperbolic automorphism M on $\mathbb{T}^2$, that is $g = M + \delta g$, $\|\delta g\|_{C^1} \leq \epsilon$. We want to prove the following property.

Theorem 2.10. *The linear hyperbolic automorphism M is C^1-structurally stable.*

Proof. We first want to solve the semiconjugacy equation

$$h \circ g = M \circ h. \tag{2.4}$$

For this we will use a method very similar to the one presented in §2.4.2. The map g is in the same homotopy class as M[2], so the perturbation δg is $\mathbb{Z}^2$-periodic. Similarly, h must be in the same homotopy class as the

[2]The lift of g on $\mathbb{R}^2$ satisfies $g(x + (1,0)) = g(x) + M((1,0))$, $g(x + (0,1)) = g(x) + M((0,1))$.

identity, so $h = Id + \delta h$, with δh biperiodic. The above equation thus reads:

$$\begin{aligned} & M^{-1} \circ (Id + \delta h) \circ (M + \delta g) = Id + \delta h \\ \iff & M^{-1} \circ \delta g + M^{-1} \circ \delta h \circ g = \delta h. \end{aligned} \tag{2.5}$$

This equation is taken over biperiodic continuous functions $\delta h : \mathbb{R}^2 \circlearrowleft$. The LHS cannot be directly expressed as a single contracting map because M^{-1} has both contracting and expanding directions. To remedy this problem, we will simply decompose δh along the unstable/stable basis in $\mathbb{R}^2$:

$$\delta h(x) = h_+(x)\, e_+ + h_-(x)\, e_-,$$

where $h_\pm : \mathbb{T}^2 \to \mathbb{R}$ are continuous biperiodic. The above equation, projected along e_+, gives

$$\mathcal{F}_+ h_+(x) \stackrel{\text{def}}{=} \lambda^{-1} \delta g_+(x) + \lambda^{-1} h_+ \circ g(x) = h_+(x).$$

The operator $\mathcal{F}_+$ is contracting: $\|\mathcal{F}_+ h_{+,1} - \mathcal{F}_+ h_{+,2}\| \leq \lambda^{-1} \|h_{+,1} - h_{+,2}\|$. As a result, $\mathcal{F}_+$ admits a single fixed point $h_{+,0}$. One easily checks that

$$\|h_{+,0}\| \leq \frac{1}{1-\lambda^{-1}} \|\delta g_+\|.$$

Projecting (2.5) along the stable direction, we get

$$\begin{aligned} & \lambda \delta g_- + \lambda h_- \circ g = h_- \\ & \iff h_- = \lambda^{-1} h_- \circ g^{-1} - \delta g_- \circ g^{-1} \stackrel{\text{def}}{=} \mathcal{F}_- h_-. \end{aligned}$$

Once again, $\mathcal{F}_-$ is contracting and admits a single fixed point $h_{-,0}$, which satisfies

$$\|h_{-,0}\| \leq \frac{1}{1-\lambda^{-1}} \|\delta g_-\|.$$

We have thus constructed a solution h_0 to the semiconjugacy (2.4).

In order to prove that h_0 is invertible, we try to solve the symmetrical equation

$$\begin{aligned} & \tilde{h} \circ M = g \circ \tilde{h} \\ \iff & \delta h \circ M - M \circ \delta h = \delta g \circ (Id + \delta h). \end{aligned} \tag{2.6}$$

The LHS is a linear operator $\mathcal{L}(\delta h)$, which acts separately on the components $h_\pm$ through two operators:

$$\mathcal{L}_+ h_+ = h_+ \circ M - \lambda h_+, \qquad \mathcal{L}_- h_- = h_- \circ M - \lambda^{-1} h_-.$$

These two operators can be easily inverted by Neumann series:

$$H_+ = \mathcal{L}_+ h_+ \iff h_+ = -\lambda^{-1} H_+ + \lambda^{-1} h_+ \circ M = -\lambda^{-1} H_+ - \lambda^{-2} H_+ \circ M - \lambda^{-2} h_+ \circ M^2 = \cdots,$$

so that

$$\mathcal{L}_+^{-1} H_+ = -\sum_{n \geq 0} \lambda^{-1-n} H_+ \circ M^n, \quad \left\|\mathcal{L}_+^{-1} H_+\right\| \leq \frac{\lambda^{-1}}{1-\lambda^{-1}} \|H_+\|.$$

Similarly,

$$\mathcal{L}_-^{-1} H_- = \sum_{n \geq 0} \lambda^{-n} H_- \circ M^{-n-1}, \quad \left\|\mathcal{L}_-^{-1} H_-\right\| \leq \frac{1}{1-\lambda^{-1}} \|H_-\|.$$

Notice that $\mathcal{L}^{-1} = (\mathcal{L}_+^{-1}, \mathcal{L}_-^{-1})$ is not contracting a priori. The equation (2.6) can then be rewritten

$$\delta h = \mathcal{L}^{-1} \mathcal{G} \delta h, \qquad \text{where} \qquad \mathcal{G}\delta h \stackrel{\text{def}}{=} \delta g \circ (Id + \delta h).$$

Now, if δg is small, one has $\|\mathcal{G}\delta h_1 - \mathcal{G}\delta h_2\| = \|\delta g(Id + \delta h_1) - \delta g(Id + \delta h_2)\| \leq \|\delta g\|_{C^1} \|\delta h_1 - \delta h_2\|$, so this operator is very contracting if $\|\delta g\|_{C^1}$ is small. As a result, the full operator $\mathcal{L}^{-1}\mathcal{G}$ will also be contracting, and

admit a unique fixed point $\tilde{h}_0$ solving (2.6). One easily checks that $h_0 \circ \tilde{h}_0$ commutes with M, and must thus be equal to the identity. □

This structural stability is actually a much more general phenomenon among hyperbolic systems.

Theorem 2.11. *Any Anosov diffeomorphism is C^1-structurally stable.*

2.7. **Quadratic maps on the interval.** So far the examples of smooth systems we have given were all linear. We now present a family of simple polynomial maps on $\mathbb{R}$, which has been extensively studied. In spite of its simplicity, it features various interesting dynamical phenomena. These maps depend on a real parameter $\mu > 0$, and are defined by

$$q_\mu(x) \stackrel{\text{def}}{=} \mu x(1-x),\ x \in \mathbb{R}.$$

The study is often restricted to points in the interval $I = [0, 1]$. When varying the parameter μ, the qualitative dynamical features change drastically for some special values; these values are called **bifurcation** values. For instance:

(1) for $0 < \mu < 1$, the map q_μ is contracting. It has a unique fixed point on I (the origin), which is attracting.
(2) for $\mu > 1$, the origin becomes a repelling fixed point (because $q'_\mu(0) > 1$), but q_μ acquires a second fixed point $x_\mu = 1 - 1/\mu$. This latter is attracting for $\mu < 3$.
(3) for $\mu > 3$ the fixed point x_μ becomes repulsive, and an attractive period-2 orbit appears nearby. $\mu = 3$ is the place of a *period-doubling bifurcation*.
(4) For $\mu > 1$, every initial point $x \in \mathbb{R} \setminus I$ will escape to $-\infty$. It is then interesting to investigate the dynamics restricted to the **trapped set** Λ_μ, that is the set of points x which remain forever in I. For $1 < \mu \leq 4$, one has $q_\mu(I) \subset I$, so the trapped set is the full interval. For $\mu > 4$, some points $x \in I$ escape, so the trapped set $\Lambda_\mu \neq I$. so that it escapes to $-\infty$.

Let us describe more precisely the trapped set when $\mu > 4$.

Proposition 2.12. *For $\mu > 4$ the trapped set Λ_μ is a Cantor set[3] in I. The restriction $q_\mu \restriction \Lambda_\mu$ is (topologically) conjugate with the full shift (Σ_2^+, σ).*

Proof. For $a = \frac{1}{2} - \sqrt{\frac{1}{4} - \frac{1}{\mu}}$, $b = \frac{1}{2} - \sqrt{\frac{1}{4} - \frac{1}{\mu}}$), the interval (a, b) is mapped by q_μ outside I, whereas $I_0 = [0, a]$ and $I_1 = [b, 1]$ are mapped bijectively to I. Call f_0, f_1 the inverse branches of q_μ on these two intervals: $f_i : I \to I_i$. These two maps allow to iteratively define a sequence of subintervals indexed by symbolic sequences $\boldsymbol{\epsilon} = \epsilon_1 \cdots \epsilon_n$. We define

$$I_{\epsilon_1 \epsilon_2 \cdots \epsilon_n} = f_{\epsilon_1} \circ f_{\epsilon_2} \circ \cdots \circ f_{\epsilon_n}(I).$$

Observe that

$$I_{\epsilon_1 \cdots \epsilon_n} \subset I_{\epsilon_1 \cdots \epsilon_{n-1}} \subset \cdots \subset I_{\epsilon_1} \subset I, \quad \text{and} \quad q_\mu(I_{\epsilon_1 \cdots \epsilon_n}) = I_{\epsilon_2 \cdots \epsilon_n}.$$

These properties show that the interval $I_{\epsilon_1 \epsilon_2 \cdots \epsilon_n}$ is made of the points $x \in I$ which have the same *symbolic history*up to time n, with respect to I_0, I_1: the point x is in I_{ϵ_1}, then its first iterate $q_\mu(x) \in I_{\epsilon_2}$, and so on: $q_\mu^j(x) \in I_{\epsilon_{j+1}}$, up to finally $q_\mu^n(x) \in I$.

This property shows that for each $n > 0$ the intervals $\left\{I_{|\boldsymbol{\epsilon}|},\ |\boldsymbol{\epsilon}| = n\right\}$ are all disjoint., and their union $I^n = \bigcup_{|\boldsymbol{\epsilon}|=n} I_{\boldsymbol{\epsilon}}$ consists of all points x such that $q_\mu^j(x) \in I$ for all $0 \leq j \leq n$. As a result the trapped set can be

[3] A (topological) Cantor set is a closed set which is perfect (has no isolated points) and is nowhere dense in I.

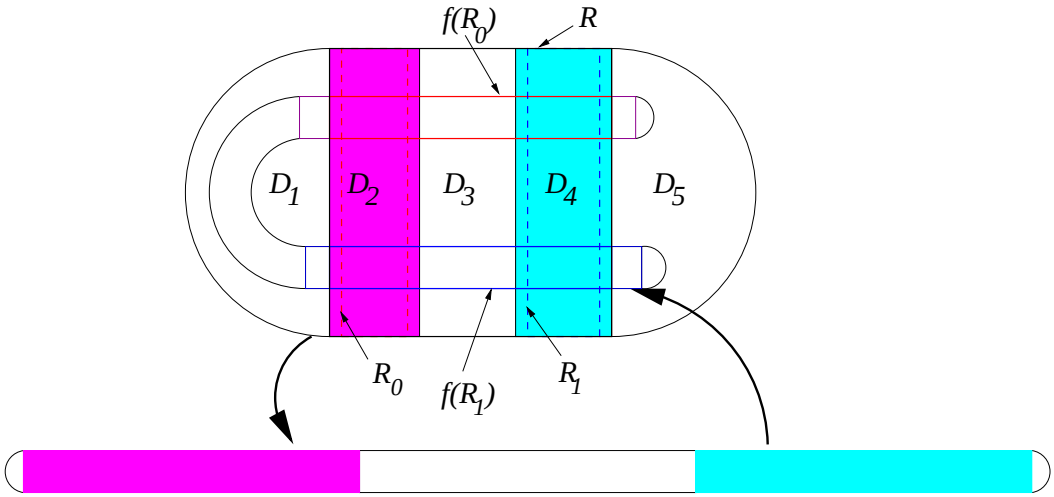

FIGURE 2.3. An example of horseshoe

defined as the closed set

$$\Lambda_\mu = \bigcap_{n\geq 1} I^n.$$

For $\mu > 2+\sqrt{5}$ the maps f_i are contracting: $|f_i'(x)| \leq \lambda_\mu < 1$, $\lambda_\mu = \mu\sqrt{1-\frac{4}{\mu}}$. As a result, the length of the intervals $I_{\boldsymbol{\epsilon}}$ decreases exponentially as $|I_{\boldsymbol{\epsilon}}| \leq \lambda_\mu^{|\boldsymbol{\epsilon}|}$. As a result, for any infinite sequence $\epsilon_1\epsilon_2\cdots$, the set $I_{\boldsymbol{\epsilon}} = \bigcap_n I_{\epsilon_n\cdots\epsilon_1}$ is a nonempty interval of length zero, that is a single point $x = x_{\boldsymbol{\epsilon}} \in I$. The map

$$\pi : \boldsymbol{\epsilon} \in \Sigma_2^+ \mapsto x_{\boldsymbol{\epsilon}} \in \Lambda_\mu$$

is a bijection, which is *bicontinuous* w.r.to the standard topology on Σ_2^+ and the induced topology on Λ_μ. It thus realizes a topological conjugacy between the full shift (Σ_2^+, σ) and the restriction $q_\mu \restriction \Lambda_\mu$. In particular, this shows that the set Λ_μ is a fully disconnected set. The contractivity of the f_i shows that $q_\mu \restriction \Lambda_\mu$ is expanding. The set Λ_μ is then called a *hyperbolic repeller*.

The case $4 < \mu < 2+\sqrt{5}$ is a little more delicate to treat, but the conclusion is the same. □

2.8. **Smale's (linear) horseshoe.** We now construct a 2-dimensional invertible map, which is an analogue of the polynomial maps q_μ ($\mu > 4$) studied in the previous section.

Smale's horseshoe can be defined as an injective (non surjective) map on a "stadium domain" $D \subset \mathbb{R}^2$, split between the two half-circles D_1, D_5 and the central square R is split between three vertical rectangles D_2, D_3, D_4 of height 1 and width $= 1/3$. The main assumptions on f are the following:

(1) $f_{\restriction D_2}$ and $f_{\restriction D_4}$ are similarities, which stretch vertically by a factor $\lambda < 1/2$ and expand horizontally by a factor $\mu > 3$, such that $f(D_2)$ and $f(D_4)$ intersects both D_1 and D_5.
(2) the map $f_{\restriction D_3}$ is nonlinear, $f(D_3)$ is contained in D_1.
(3) $f(D_1)$ and $f(D_5)$ are contained in D_5.

The map f is not surjective on D, but $f : D \to f(D)$ is injective.

The preimage $f^{-1}(R)$ splits into two disjoint rectangles $R_0 \subset D_2$, $R_1 \subset D_4$ of width μ^{-1} and height 1.

The backwards images of each of these rectangles $f^{-1}(R_i)$ is the union of two vertical rectangles of width μ^{-2} and height 1 contained in R_0 and R_1, so that $R_{\epsilon_0} \cap f^{-1}(R_{\epsilon_1})$ is such a rectangle. By iteration, the sets

$R_{\epsilon_0} \cap f^{-1}(R_{\epsilon_1}) \cap f^{-2}(R_{\epsilon_2}) \cap \cdots \cap f^{-n+1}(R_{\epsilon_{n-1}})$ are vertical rectangles of width μ^{-n} and height 1. For each sequence $\boldsymbol{\alpha} \in \Sigma_2^+$, the set

$$R_{\boldsymbol{\alpha}} = \bigcap_{j \geq 0} f^{-j}(R_{\alpha_j})$$

is a vertical segment of height 1 contained in R_{ϵ_0}. The set $H^- \stackrel{\text{def}}{=} \bigcup_{\boldsymbol{\alpha} \in \Sigma_2^+} R_{\boldsymbol{\alpha}} = \bigcap_{j \geq 0} f^{-j}(R)$ is the product of a horizontal Cantor set by the union of two vertical intervals. It is made of points x whose forward trajectory always remains in R.

Similarly, for any $\epsilon_{-n} \cdots \epsilon_{-1}$ the set $R_{\epsilon_{-n} \cdots \epsilon_{-1} \cdot} = \bigcap_{j=1}^{n} f^j(R_{\epsilon_{-j}})$ is a rectangle of height λ^n and width 1 contained in R. For each $\boldsymbol{\epsilon} \in \Sigma_2^-$, the set $R_{\boldsymbol{\epsilon}\cdot} = \bigcap_{j=1}^{\infty} f^j(R_{\epsilon_{-j}})$ is a single horizontal segment (of width 1). The union of these segments $H^+ \stackrel{\text{def}}{=} \bigcup_{\boldsymbol{\epsilon} \in \Sigma_2^-} R_{\boldsymbol{\epsilon}\cdot} = \bigcap_{j \geq 0} f^j(R)$ is the product of a vertical Cantor set by a horizontal segment.

Hence, the intersection $\Lambda = H^+ \cap H^-$ is the product of two Cantor sets. It is made of all points whose (forward and backward) trajectories always remain in R. Let us now take $\boldsymbol{\beta} = \boldsymbol{\epsilon} \cdot \boldsymbol{\alpha} \in \Sigma_2$ a bi-infinite sequence. By construction, the intersection $R_{\boldsymbol{\beta}} = \bigcap_{j \in \mathbb{Z}} f^{-j}(R_{\beta_j})$ is a single point $x_{\boldsymbol{\beta}} = \pi(\boldsymbol{\beta})$, which is characterized by the property

$$f^j(x_{\boldsymbol{\beta}}) \in R_{\beta_j}, \ \forall j \in \mathbb{Z}.$$

The map $\pi : \boldsymbol{\beta} \in \Sigma_2 \mapsto x_{\boldsymbol{\beta}} \in \Lambda$ is a bicontinuous bijection, which conjugates the two-sided full shift (Σ_2, σ) with the (invertible map) $f \restriction \Lambda$.

By construction, at each point $x \in \Lambda$ the linearized map $df(x) = \begin{pmatrix} \mu & 0 \\ 0 & \lambda \end{pmatrix}$ is the same. This shows that each $x \in \Lambda$ is a hyperbolic point, with E_x^+ the horizontal direction (resp. E_s^- the vertical direction). Λ is therefore a *hyperbolic set* (a compact, invariant set such that each $x \in \Lambda$ is hyperbolic, see §6.1).

2.8.1. *Markov partition.* The sets $\mathcal{R}_i = R_i \cap \Lambda$, $i = 0, 1$ are *rectangles* in the usual sense, but also in the sense of the *local product structure* of hyperbolic dynamics (see §6.3): for each $x, y \in \mathcal{R}$, the unique point

$$[x, y] \stackrel{\text{def}}{=} W^-_{loc}(x) \cap W^+_{loc}(y)$$

also belongs to $\mathcal{R}$. Hence, in 2 dimensions the boundaries of $\mathcal{R}$ are made of unstable and stable segments.

Obviously, one has $\Lambda = \mathcal{R}_0 \sqcup \mathcal{R}_1$. From such a partition, one can always obtained refined partitions $\{\mathcal{R}_{\boldsymbol{\alpha}\cdot\boldsymbol{\epsilon}}, \ |\boldsymbol{\alpha}| = |\boldsymbol{\epsilon}| = n\}$. Due to the hyperbolicity of f, it is easy to show that the diameters of the $\mathcal{R}_{\boldsymbol{\alpha}\cdot\boldsymbol{\epsilon}}$ decreases exponentially with n, so that to each bi-infinite sequence $\boldsymbol{\beta}$ will be associated at most a single point $x(\boldsymbol{\beta})$. What is not obvious in general is to determine which sequences $\boldsymbol{\beta}$ are allowed (that is, effectively correspond to a point). The answer is relatively simple provided the rectangles $\mathcal{R}_i$ form a **Markov partition**:

(1) $int(\mathcal{R}_i) \cap int(\mathcal{R}_j) = \emptyset$ if $i \neq j$.
(2) if $x \in int(\mathcal{R}_i)$ and $f(x) \in int(\mathcal{R}_j)$ then $W^+_{\mathcal{R}_j}(f(x)) \subset f(W^+_{\mathcal{R}_i}(x))$
(3) if $x \in int(\mathcal{R}_i)$ and $f(x) \in int(\mathcal{R}_j)$ then $f(W^-_{\mathcal{R}_i}(x)) \subset W^-_{\mathcal{R}_j}(f(x))$.

In the present case, the first property is obvious because $\mathcal{R}_0, \mathcal{R}_1$ are disjoint. Each unstable leaf $W^+_{\mathcal{R}_i}(x)$ consists in the intersection between a horizontal segment of length μ^{-1} and Λ; its image through f is the union of two such segments, one intersecting $\mathcal{R}_0$ all along, the other intersecting $\mathcal{R}_1$ all along, so the second property is OK. Similarly $W^-_{\mathcal{R}_i}(x)$ is a vertical segment of length 1 intersecting Λ, its image is a vertical segment of length λ intersecting Λ, and fully contained in either $\mathcal{R}_0$ or $\mathcal{R}_1$, so the third property is OK.

Lemma 2.13. *The above properties of the partition imply the following "Markov" property:*

if $f^m(\mathcal{R}_i) \cap \mathcal{R}_j \neq \emptyset$ and $f^n(\mathcal{R}_j) \cap \mathcal{R}_k \neq \emptyset$, then $f^{n+m}(\mathcal{R}_i) \cap \mathcal{R}_k \neq \emptyset$.

Exercise 2.14. Describe the unstable and stable manifolds of $x \in \Lambda$, defined by

$$W^{\pm}(x) = \left\{ y \in D, \quad dist\left(f^{\mp n}(x), f^{\mp n}(y)\right) \xrightarrow{n\to+\infty} 0 \right\}.$$

2.9. **Hamiltonian flows.** In this section we add up some more structure on the manifold X. We assume that X is a symplectic manifold, namely it is equipped with a nondegenerate closed antisymmetric two-form ω (X is then necessarily even-dimensional and orientable). The simplest case is the Euclidean space $X = T^*\mathbb{R}^d \simeq \mathbb{R}^{2d}$, with coordinates $x = (q, p)$, and symplectic form $\omega = \sum_{i=1}^d dp_i \wedge dq_i$. A more general example is that of the cotangent bundle $X = T^*M$ over a manifold M. One can then define ω as above in each coordinate chart (q_i, p_i). One checks that the formula is invariant through a change of coordinates $y = \phi(q), \xi = {}^t d\phi^{-1}(y) \cdot p$. Notice that these phase spaces are noncompact.

A Hamiltonian is a function $H(q,p) \in C^\infty(X)$, which represents the "energy" of the particle. It generates a Hamiltonian vector field X_H on X, given by $dH = \omega(\cdot, X_H)$, that is

$$X_H(q,p) = \sum_i \frac{\partial H(q,p)}{\partial p_i}\frac{\partial}{\partial q_i} - \frac{\partial H(q,p)}{\partial q_i}\frac{\partial}{\partial p_i}.$$

This vector field generates a flow, that is trajectories $(q(t), p(t))$ satisfying

$$\dot{q}_i \stackrel{\text{def}}{=} \frac{dq_i}{dt} = \frac{\partial H(q,p)}{\partial p_i}, \qquad \dot{q}_i \stackrel{\text{def}}{=} \frac{dp_i}{dt} = -\frac{\partial H(q,p)}{\partial q_i}.$$

Let us take the differential of these equations:

$$d\dot{q}_i = \frac{\partial^2 H(q,p)}{\partial q_j \partial p_i} dq_j + \frac{\partial^2 H(q,p)}{\partial p_j \partial p_i} dp_j, \quad d\dot{p}_i = -\frac{\partial^2 H(q,p)}{\partial q_k \partial q_i} dq_k - \frac{\partial^2 H(q,p)}{\partial p_k \partial q_i} dp_k,$$

so the variation of $\sum_i dq_i \wedge dp_i$ is simply

$$d\dot{q}_i \wedge dp_i + dq_i \wedge d\dot{p}_i = \left(\frac{\partial^2 H(q,p)}{\partial q_j \partial p_i} dq_j + \frac{\partial^2 H(q,p)}{\partial p_j \partial p_i} dp_j\right) \wedge dp_i - dq_i \wedge \left(\frac{\partial^2 H(q,p)}{\partial q_k \partial q_i} dq_k - \frac{\partial^2 H(q,p)}{\partial p_k \partial q_i}\right) = 0,$$

meaning that the flow preserves the symplectic form (the terms $dp_j \wedge dp_i$ vanish because $\frac{\partial^2 H(q,p)}{\partial p_j \partial p_i}$ is symmetric; idem for $dq_i \wedge dq_k$). As a byproduct, the natural volume element $dvol = \prod dq_i dp_i \simeq \bigwedge_i dq_i \wedge dp_i = \frac{1}{d!}\omega^d$ is also preserved by the flow (Liouville theorem).

The energy of the particle is constant along a trajectory:

$$\dot{H} = \sum_i \frac{\partial H}{\partial q_i}\dot{q}_i + \frac{\partial H}{\partial p_i}\dot{p}_i = 0.$$

It thus makes sense to restrict the dynamics to individual energy shells $H^{-1}(E) = \{(q,p) \in X,\ H(q,p) = E\}$. In the cases where $H^{-1}(E)$ is compact, we are back to the study of a flow on a compact manifold. The Liouville measure $d\mu_E = \delta(H(q,p) - E) dvol$ supported on $H^{-1}(E)$ is flow-invariant.

Geodesic flow on a manifold. A particular case of Hamiltonian flow on a Riemannian manifold (M, g) is provided by the free motion: it corresponds to the Hamiltonian

$$H(q,p) = \frac{\|p\|_g^2}{2} = \frac{1}{2}\sum G_{ij} p_i p_j.$$

(here the metric g acts on the cotangent bundle T^*M, so in coordinates it corresponds to the matrix $G = g^{-1}$, where $g = (g_{ij})$ represents the metrics on TM: $ds^2 = \sum_{ij} g_{ij} dx_i dx_j$). The dynamics on the unit cotangent bundle $H^{-1}(1/2) = S^*M$ (that is, the set of points with unit momenta), is equivalent with the *geodesic flow*, which lives on the space SM of unit velocities. The Liouville measure on S^*M is the lift of the Lebesgue measure on M.

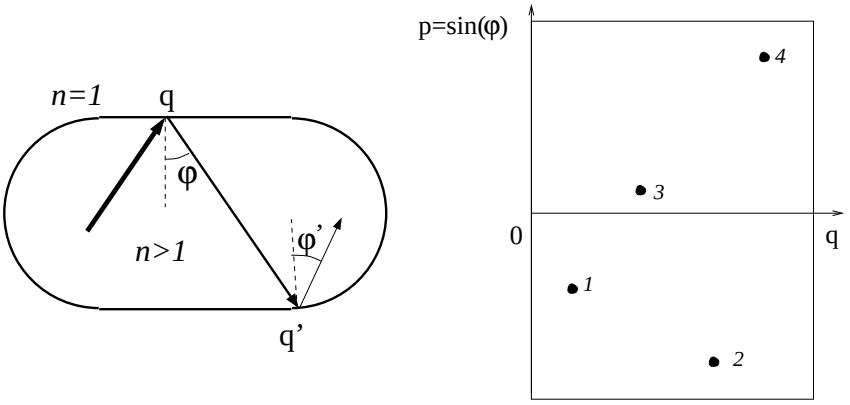

FIGURE 2.4. A Euclidean billiard and its associated billiard map

Depending on the topology of M and the riemannian metric g on it, the dynamical properties of the geodesic flow can be quite diverse. One interesting class of manifolds are the manifolds (M, g) such that the *sectional curvature* K is everywhere negative (each embedded plane locally looks like a saddle). This negativity implies a uniform hyperbolicity of the dynamics, so that the full energy shell S^*M is a hyperbolic set (see §6.1).

Euclidean billiards. Another possibility is to restrict the motion of the free particle inside a bounded region of (M, g), with specular reflection at the boundaries. For instance, a bounded connected domain $D \subset \mathbb{R}^2$ is called a Euclidean billiard. The particle moves with velocity $|\dot{q}| = 1$ along straight lines inside the domain, and is reflected when touching the boundary (if the boundary if C^1, the reflection is well-defined everywhere). The motion of the particle is restricted to the compact phase space S^*D. Its qualitative features only depend on the shape of D. For instance, the billiard flow in the stadium billiard (see fig. 2.4) is known to be ergodic and mixing w.r.to the Liouville measure.

A natural Poincaré section for the billiard dynamics is the bounce map (or billiard map): it only collects the points where the particle bounces on the boundary, as well as the angle $\varphi \in [-\pi/2, \pi/2]$ of the outgoing velocity with the inwards normal vector to the boundary:

$$(s, \sin\varphi) \mapsto (s', \sin\varphi').$$

That this map preserves the symplectic form $\omega = \cos(\varphi) d\varphi \wedge ds$ on the reduced phase space B^*S, where $S \simeq [0, L)$ is the perimeter, and $B^*S = \{s \in S, \sin\varphi \in [-1, 1]\}$ its unit cotangent ball.

2.10. **Gradient flows.** Let (X, g) be a Riemannian manifold, and F a smooth real function on X. The gradient of the function F is the tangent vector given (in local coordinates) by

$$\nabla F(x) = G(x) \begin{pmatrix} \partial F/\partial x_1 \\ \vdots \\ \partial F/\partial x_d \end{pmatrix},$$

where $G = g^{-1}$. This vector is orthogonal to the level sets of F. The flow generated by the vector field ∇F is called the gradient flow of F. The function F decreases along all trajectories, strictly so except at the fixed points, which are the critical points of F.

3. Recurrences in topological dynamics

We will now define some particular long-time properties of a *continuous* map f on a *compact* metric space X. In a first step, we will only consider the topological properties of the dynamics.

3.1. **Recurrences.** Consider an initial point $x \in X$. If its iterates $f^n(x)$ leave a neighbourhood of x for ever (that is, for every $n > N$), then the point x is said to be *non-recurrent*. To better describe this property, it is convenient to introduce the ω-limit set of x (denoted by $\omega(x)$), which is the set of points $y \in X$ such that the forward trajectory $(f^n(x))_{n\geq 0}$ comes arbitrary close to y infinitely many times[4]. If f is invertible, then the α-limit set of x is defined similarly w.r.to the backward evolution.

Exercise 3.1. For each x the set $\omega(x)$ is nonempty, closed and invariant.

Example 3.2. Consider the gradient flow of F on a compact manifold X. For any x, the set $\omega(x)$ consists of fixed points, that is critical points of F. One can show that for each x, the set $\omega(x)$ is either a single point, or an infinite set of points.

Definition 3.3. A point x such that $x \in \omega(x)$ is called recurrent. The set of such points is denoted by $\mathcal{R}(f)$.

Example. A periodic point such that $f^n(x) = x$ for some $n > 0$ is obviously recurrent: the set $\omega(x)$ is then the (finite) periodic orbit.

Example 3.4. Let x_0 be a hyperbolic fixed point of a diffemorphism f. Assume x_0 admits a *homoclinic point*, that is a point $x_1 \neq x_0$ such that $f^n(x_1) \xrightarrow{n\to\pm\infty} x_0$. In that case, $\omega(x_1) = \alpha(x_1) = x_0$. The point x_0 is recurrent, but x_1 is not.

The set of recurrent points $\mathcal{R}(f)$ is invariant w.r.to f, but in general it is not a closed set. For this reason, it is more convenient to use a weaker notion of recurrence:

Definition 3.5. A point $x \in X$ is called *nonwandering* if, for any (small) neighbourhood $U(x)$, there exists arbitrary large $n > 0$ such that $f^n(U(x)) \cap U(x) \neq \emptyset$. (equivalently, $f^n(U(x))$ will intersect $U(x)$ infinitely many times). The set of nonwandering points is denoted by $NW(f)$.

Exercise 3.6. The set $NW(f)$ is closed and invariant. It contains the recurrent points $\mathcal{R}(f)$, as well as the ω- and α-limit sets of all $x \in X$.

The set of nonwandering points is the locus of the "interesting part" of the dynamics. A region of phase space outside $NW(f)$ can only welcome some "transient" dynamics, but after a while the trajectory will leave that region.

One aim of topological dynamics is to understand the structure of closed invariant sets.

Definition 3.7. A closed, invariant set $\emptyset \neq Y \subset X$ is *minimal* if it does not contain any proper subset which is also closed and invariant.

Equivalently, for any $x \in Y$ the orbit $\mathcal{O}^+(x)$ is dense in Y.($\Longrightarrow$every point in a minimal set is recurrent).

Example. The simplest example of minimal set is a *periodic orbit*. On the other hand, it is easy to see that the full circle $S^1 = X$ is minimal for an irrational rotation f_α.

Proposition 3.8. *Any continuous map $f : X \to X$ admits a minimal set $Y \subset X$.*

[4]Equivalently, there is a sequence $(n_k)_{k\geq 1}$ such that $f^{n_k}(x) \overset{k\to\infty}{\to} y$.

The next notion describes whether the dynamics acts "separately" on different parts of X.

Definition 3.9. Let $f : X \to X$ be a continuous map. f is said to be *topologically transitive* if there exists an orbit[5] $\{f^n(x_0),\ n \in \mathbb{N}\}$ which is *dense* in X. Equivalently, for any (nonempty) open sets U, V, there is a time $n \geq 0$ such that $f^n(U) \cap V$ is not empty.

Example 3.10. Irrational rotations on S^1, linear dilations E_m on S^1, hyperbolic automorphisms on $\mathbb{T}^d$, quadratic maps q_μ ($\mu > 4$) on the trapped set Λ_μ, full shifts $\Sigma_m^{(+)}$ are topologically transitive.

A topological Markov chain Σ_A^+ is topologically transitive if the matrix A is irreducible.

3.2. **What is a "chaotic system"?** There is no mathematically precise notion of "chaos". One could consider an irrational translation as being "chaotic", because single trajectories explore the full phase space. Still, under "chaotic" one generally assumes that all (or at least, many) trajectories enjoy a *sensitive dependence to initial conditions*. This property could be phrased as follows: on a subset $X' \subset X$ there exists a distance $\delta > 0$ such that, for any $x \in X'$ and any (small) distance $\epsilon > 0$, there are $y \in X$ and $n \geq 0$ such that $dist(x, y) \leq \epsilon$ and $dist(f^n(x), f^n(y)) \geq \delta$.

This property (which concerns points at *finite* distances) is often replaced by the notion of **Lyapunov exponents**, which concern the growth of *infinitesimal* distances for a *differentiable* map f on a smooth manifold X:

$$\forall x \in M,\ \forall v \in T_xX, \quad \chi(x, v) \stackrel{\text{def}}{=} \limsup_{n\to\infty} \|df^n(x)v\|.$$

Eventhough the two notions are not equivalent, in practice

$$\text{sensitive dependence to initial conditions} \simeq \text{positive Lyapunov exponents.}$$

The next property expresses a stronger form of sensitivity to initial conditions than above.

Definition 3.11. A map (resp. homeomorphism) is *expansive* iff there exists $\delta > 0$ such that, for any two distinct points $x \neq y$, there exists $n \in \mathbb{N}$ (resp. $n \in \mathbb{Z}$) such that $dist(f^n(x), f^n(y)) > \delta$. The largest such δ is called the expansiveness constant of f.

Compared with the previous definition of "sensitivity", we do not need to assume that $x \in X'$, and the future separation is true for any y close to x.

Obviously, the rotations (like any isometry) are not expansive. The other examples (which contain some hyperbolicity) are expansive.

This rather innocent-looking property implies a stronger consequence:

Proposition 3.12. *Let f be an expansive homeomorphism on an (infinite) compact metric space X. Then there exists $x_0 \neq y_0$ such that $dist(f^n(x_0), f^n(y_0)) \stackrel{n\to\infty}{\to} 0$.*

The next property is again a form of recurrence, which looks quite similar with topological transitivity.

Definition 3.13. A continuous map $f : X \to X$ is said to be *topologically mixing* iff for any nonempty open sets U, V, there exists a time $N > 0$ such that *for any* $n \geq N$ one has $f^n(U) \cap V \neq 0$.

This property describes a quite different phenomenon from topological transitivity. Consider a small open set $U \subset X$, and a finite open cover $X = \cup_{j=1}^J V_j$. Topological transitivity tells us that a small open set U will, through the map f, intersect each V_j in the future: the dynamics will carry U through the whole phase space.

[5]If f is a homemorphism and X has no isolated point, this is equivalent to assuming that there is a dense full orbit $\{f^n(x_0),\ n \in \mathbb{Z}\}$.

However, the different parts of phase space can be visited at different times. On the opposite, topological mixing implies the existence of some $N > 1$ such that, for any $n \geq N$, the set $f^n(U)$ intersects all V_j simultaneously. This shows that for such large times, the set $f^n(U)$ has been stretched by the dynamics so that it (roughly) covers the whole phase space.

Example 3.14. The rotations on S^1 are not topologically mixing. Dilations E_m on S^1, hyperbolic automorphisms on $\mathbb{T}^d$, full shifts $\Sigma_m^{(+)}$ are topol. transitive. A topological Markov chain $\Sigma_A^{(+)}$ is topologically mixing if the adjacency matrix A is primitive.

We will see later that the notions of topological transitivity and topological mixing have natural counterparts in the framework of measured dynamical systems, namely ergodicity and mixing. Also, the notion of Lyapunov exponent acquires a crucial role in that framework.

3.3. **Counting periodic points.** In the case where the number or periodic orbits of period n is finite for all $n > 0$, one is interested in counting them as precisely as possible, at least in the limit $n \gg 1$. Such counting is obviously a topological invariant of the system.

For many systems of interest, the number of periodic points grows exponentially with n. It thus makes sense to define the rate

$$p(f) \stackrel{\text{def}}{=} \limsup_{n\to\infty} \frac{1}{n} \log \sharp \operatorname{Fix}(f^n). \tag{3.1}$$

Inspired from methods from number theory, one can use various forms of generating functions to count periodic points.

Definition 3.15. If a map f has finitely many n-periodic points for each n, we can associate to f the *zeta function*

$$\begin{aligned}\zeta_f(z) &\stackrel{\text{def}}{=} \exp \sum_{n\geq 1} \frac{z^n}{n} \sharp \operatorname{Fix}(f^n) \\ &= \prod_{\gamma} (1 - z^{|\gamma|})^{-1} \qquad \text{Euler product} \\ &= \exp z g_f'(z), \qquad g_f(z) = \sum_{n\geq 1} z^n \sharp \operatorname{Fix}(f^n) \quad \text{is a generating function.}\end{aligned}$$

The Euler product on the second line is taken over *primitive* orbits only.

The analytical properties of ζ_f provide informations on the statistics of long periodic orbits. For instance, the radius of convergence for ζ_f is given by $r = \frac{1}{p(f)}$, where ζ_f develops a singularity (usually a pole).

Exercise 3.16. For the SFT Σ_A, show that $\zeta(z) = \frac{1}{\det(1-zA)}$. Assuming A is primitive, compute the asymptotics for $\sharp \operatorname{Fix}(f^n)$.

4. Measured dynamical systems: ergodic theory

So far the only structure we have assumed on phase space is a distance (inducing a topology, that is a notion of continuity), and a differentiable structure (implying that one can linearize the dynamics locally at each point).

In this section we impose an additional structure on the phase space: a probability measure.

4.1. **What is a measure space?** To define measures on X, one must first decide of which subsets of X are measurable. Such sets form a σ-algebra $\mathfrak{U}$ (closed under countable union and complement). A measure μ is a nonnegative σ-additive function on $\mathfrak{U}$: for any coutable family of *disjoint* sets ($A_i \in \mathfrak{U}$), one must have $\mu(\cup_i A_i) = \sum_i \mu(A_i)$. A probability measure satisfies $\mu(X) = 1$. The triplet $(X, \mathfrak{U}, \mu)$ is called a *measure space*. In case $\mu(X) = 1$, it is called a probability space.

The main point of measure theory is the following:

> Null sets (that is sets A such that $\mu(A) = 0$) are totally irrelevant. The complement of a null set is a set of full measure.
>
> A property is said to be true *almost surely* (a.s.), or almost everywhere (a.e.), if it holds on the complement of a null set.

Definition 4.1. A map (or transformation) $T : X \to X$ is said to be measurable iff for any measurable set A, the preimage $T^{-1}(A)$ is also measurable. The measure μ is said to be invariant w.r. to f (or equivalently, T is said to be measure-preserving) iff for any measurable set A, one has $\mu(T^{-1}(A)) = \mu(A)$.

We call $\mathcal{M}(X)$ the set of probability measures on X, and $\mathcal{M}(X, T)$ the set of invariant probability measures. We will see below (Thm. 4.5) that the latter set is nonempty if T is a continuous transformation. Both sets are **compact** w.r.to the weak topology on measures, meaning that from any sequence of probability measure (μ_n) one can extract a subsequence (μ_{n_k}) converging to a measure (resp. an invariant measure) μ.

The sets $\mathcal{M}(X)$, $\mathcal{M}(X, T)$ are obviously **convex**.

Two measure spaces $(X, \mathfrak{U}, \mu)$ and $(Y, \mathfrak{B}, \nu)$ are said to be isomorphic iff there exists subsets $X' \subset X$, $Y' \subset Y$ of full measure, and measure-preserving bijection $\psi : X' \to Y'$. From such an isomorphism, one easily defines the notion of isomorphy between transformations $S : X \to X$ and $T : Y \to Y$.

Let us be more specific with our measure spaces. Since our space X is already equipped with a topology, the most natural σ-algebra on it is the Borel σ-algebra $\mathfrak{B}$, which contains all open and all closed sets. From now on we will exclusively consider this σ-algebra. A measure μ on $\mathfrak{B}$ is called a Borel measure. A point $x \in X$ is called an *atom* if $\mu(\{x\}) > 0$. On Euclidean space (or by extension, on a Riemannian manifold), the measure inherited from the metric structure is the *Lebesgue measure*.

The probability spaces we will encouter are all *Lebesgue spaces*: they are isomorphic with some interval $[0, a]$ equipped with the Lebesgue measure, plus at most countably many atoms.

Remark 4.2. If X is a domain on $\mathbb{R}^d$ or a Riemannian manifold, one should not confuse the notion "Lebesgue space" with the fact that μ is absolutely continuous w.r.to the Lebesgue measure on X: the isomorphism f is by no means required to be continuous! For instance, the 1/3-Cantor set $\mathcal{C}$ equipped with its standard Bernouilli measure is a Lebesgue space, eventhough its measure looks "fractal". Also, the unit square $[0, 1]^2$ equipped with the Lebesgue measure is isomorphic with the unit interval $[0, 1]$ equipped with Lebesgue (Exercise).

The first major result of ergodic theory concerns recurrence properties (now expressed in terms of measurable sets).

Theorem 4.3. *[Poincaré recurrence theorem]*

Assume T is a measure-preserving transformation on the probability space $(X, \mathfrak{U}, \mu)$. Consider $A \subset X$ a measurable set. Then, for (μ-)almost every $x \in A$, the trajectory $\{T^n(x),\, n \geq 0\}$ will visit A infinitely many times.

Proof. Consider the set

$$B = \{x \in A,\, T^n(x) \notin A,\, \forall n > 0\} = A \setminus \bigcup_{n>0} T^{-n}(A).$$

That set is measurable, and $T^{-k}(B)$ contains points such that $T^k(y) \in A$ but $T^{k+n}(y) \notin A$ for any $n > 0$, hence the $T^{-k}(B)$ are all disjoint. On the other hand, they have the same measure as B, so deduces that $\mu(B) = 0$. □

If we now assume that X is a metric space, μ is a Borel measure and $T : X \circlearrowleft$ is continuous and preserves μ, we deduce that ($\mu-$)almost every point x is recurrent (in the topological sense). As a result, the *support*[6] of the measure μ is contained in the closure of the recurrence set, which is itself contained in the nonwandering set.

One has $\mu(X \setminus \operatorname{supp}\mu) = 0$, and any set of full measure is dense in $\operatorname{supp}\mu$. By definition, any nonempty open set $A \subset \operatorname{supp}\mu$ has positive measure.

4.1.1. *Observables on a measure space.*

Remark 4.4. On a measure space $(X, \mathfrak{U}, \mu)$ the natural "observables", or "test functions" are measurable functions $f : X \to \mathbb{R}$, preferably with some bounded growth: in general one requires them to belong to some Banach space $f \in L^p(X, \mu)$ $(1 \leq p \leq \infty)$. To check whether $f \in L^p(X, \mu)$ one only needs to control f on a set of full measure[7]. Among the Banach spaces the Hilbert space $L^2(X, \mu)$ will play a particular rôle.

For some refined properties (e.g. exponential mixing), one often needs to require stronger regularity properties on the observables.

4.2. **Existence of invariant measures.** Since the following section will deal with invariant measures, the first relevant question concerning a given transformation T is:

> Given a measurable map T on X, does it always admit an invariant measure?

In full generality, the answer is NO. A simple example is provided by the following map on $S^1 \equiv (0, 1]$:

$$f(x) = x/2,\ x \in (0, 1].$$

This map is discontinuous at the origin. The following theorem shows that continuity of T is a sufficient condition to insure the existence of some invariant measure.

Theorem 4.5. *[Krylov-Bogolubov]Let $T : X \circlearrowleft$ be continuous on the compact metric space X. Then there exists a T-invariant Borel probability measure μ on X.*

Proof. The proof uses some compactness arguments. For any function $f \in C(X)$, we define the *Birkhoff averages*

$$f_n = \frac{1}{n} \sum_{j=0}^{n-1} f \circ T^j, \qquad n \geq 1. \tag{4.1}$$

[6] $\operatorname{supp}\mu$ *is the intersection of all closed sets of full measure. Equivalently, its complement is the union of all null open sets.*
[7] More precisely, the elements of L^p are equivalence classes of functions, $f \sim g$ iff $f(x) = g(x)$ almost everywhere.

Fix a point $x \in X$, and consider a dense countable set $(\varphi^m)_{m\geq 1}$ in $C(X)$. For each φ^m, the sequence $((\varphi^m)_n(x))_{n\geq 0}$ is bounded, so it admits a convergent subsequence. By the diagonal trick, we can extract a subsequence n_k such that

$$\forall m \geq 1, \qquad \lim_{k\to\infty} (\varphi^m)_{n_k}(x) = J(\varphi^m) \quad \text{exists.}$$

By density of (φ^m) inside $C(X)$, this limit exists as well for any continuous function φ, and defines a linear, bounded, positive functional $J(\bullet)$ on the space of continuous functions. By the Riesz representation theorem, $J(\varphi) = \int \varphi \, d\mu$ where $\mu \in \mathcal{M}(X)$. Besides, we have

$$\forall n, \varphi, \qquad (\varphi \circ T)_n(x) = \frac{1}{n}\sum_{j=1}^{n} \varphi \circ T^j(x) = \varphi_n(x) + \frac{\varphi \circ T^n(x) - \varphi(x)}{n},$$

so that $J(\varphi) = J(\varphi \circ T)$, or equivalently $\int \varphi \, d\mu = \int (\varphi \circ T) \, d\mu$ for any continuous φ. This last property makes sense because T is continuous, and is equivalent with the invariance of μ. □

4.3. **Ergodicity.**

4.3.1. *Formal definition.* The notions of ergodicity and mixing describe the asymptotic properties of the action of a transformation on observables: this action can be expressed through the operator $U_T(f) \stackrel{\text{def}}{=} f \circ T$ on $L^p(\mu)$. From the invariance of μ, this operator is an isometry on $L^p(\mu)$: $\|U_T(f)\|_p = \|f\|_p$. If T is invertible, the inverse $U_T^{-1} = U_{T^{-1}}$ is also an isometry; in particular, U_T is then a unitary operator on the space $L^2(\mu)$.

A function f is said to be essentially invariant through T iff the set $\{x \in X,\ f(T(x)) = f(x)\}$ has full measure. A measurable set $A \subset X$ is invariant through T iff $T^{-1}(A) = A$, and *essentially invariant* iff $\mu\left(T^{-1}(A)\Delta A\right) = 0$ [8].

We start by giving a formal definition of the notion of ergodicity. A more "physical" definition will be given in the following section.

Definition 4.6. A measure-preserving transformation $T : X \circlearrowleft$ on a probability space $(X, \mathfrak{U}, \mu)$ is ergodic (w.r.to the invariant measure μ) iff any (essentially) invariant measurable set A has measure zero or unity.

Proposition 4.7. *T is ergodic iff any (essentially) invariant function $f \in L^p(X, \mu)$ is constant almost everywhere. Ergodicity can thus be expressed as a spectral statement for the operator U_T on L^p: T is ergodic iff* $\ker(U_T - 1)$ *is one-dimensional.*

We have already seen that a measure-theoretic form of recurrence holds for any measure-preserving transformation. Ergodicity, on the other hand, is the measure-theoretic counterpart of topological transitivity: it implies that any set A of positive measure will, in the course of evolution, visit the full phase space (up to a null set). But the statement can be made much more quantitative: each region of phase space is visited at an asymptotically precise frequency, which is proportional to its μ-volume.

We will denote by $\mathcal{M}_e(X, T)$ the set of ergodic invariant probability measures.

Proposition 4.8. *$\mathcal{M}_e(X, T)$ exactly consists of the **extremal** points in the convex set $\mathcal{M}(X, T)$, that is the measures which cannot be expressed as a convex combination of two different measures.*

Proof. Assume μ is not ergodic, so that there exists $A \subset X$ invariant with $0 < \mu(A)$, $0 < \mu(\complement A)$. μ is then the linear combination of the two invariant measures $\frac{\mu\restriction A}{\mu(A)}$, $\frac{\mu\restriction \complement A}{\mu(\complement A)}$, so it is not extremal.

[8] $A\Delta B = A \setminus B \cup B \setminus A$) is the symmetric difference between the sets A, B.

On the opposite, assume μ is ergodic, and $\mu = p\mu_1 + (1-p)\mu_2$, with $\mu_1, \mu_2 \in \mathcal{M}(X,T)$ and $\mu_1 \neq \mu_2$. The two measures μ_i are absolutely continuous w.r.to μ, in particular $d\mu_1 = \rho_1\, d\mu$, $\rho_1 \in L^1(\mu)$. Call $E \stackrel{\text{def}}{=} \{x \in X,\ \rho_1(x) < 1\}$. The identity $\mu_1(E) = \mu(T^{-1}E)$ implies $\mu_1(E \setminus T^{-1}E) = \mu_1(T^{-1}E \setminus E)$, that is

$$\int_{E\setminus T^{-1}E} \rho_1\, d\mu = \int_{T^{-1}E\setminus E} \rho_1\, d\mu.$$

From the assumption $\rho_1 < 1$ on E, one deduces that $\mu(E \setminus T^{-1}E) = \mu(T^{-1}E \setminus E) = 0$, meaning that E is essentially invariant. From the ergodicity assumption, we must have $\mu(E) = 0$ or $\mu(E) = 1$. In the latter case, $\mu_1(X) = \mu_1(E) < 1$, a contradiction. Therefore, $\mu(E) = 0$. The same proof shows that the set $F \stackrel{\text{def}}{=} \{x \in X,\ \rho_1(x) > 1\}$ is null. Hence, $\mu = \mu_1$. □

The convexity of $\mathcal{M}(X,T)$ has a stronger consequence:

Theorem 4.9. *[Ergodic decomposition] Every invariant Borel measure μ can be decomposed into a (possibly countable) convex combination of ergodic invariant measures. There exists a Borel probability measure τ_μ on the set $\mathcal{M}_e(X,T)$, such that*

$$\mu = \int_{\mathcal{M}_e(X,T)} m\, d\tau_\mu(m). \tag{4.2}$$

4.3.2. *Birkhoff averages.* The initial goal of ergodic theory was the study of the Birkhoff averages (or *time averages*) f_n of an observable f. If f is invertible, we may as well define the average in the past direction, $f_n^- = (f^{-1})_n$. Ergodic theory wants to determine whether, and in which sense these averages admit well-defined limits when $n \to \infty$.

The easiest analysis of this problem uses a "quantum-like" analysis (in the sense of "operator theory on L^2").

Theorem 4.10. *[Von Neumann] Assume the transformation T preserve the measure μ on X. For any $f \in L^2(\mu)$, the Birkhoff averages f_n converges in L^2 to a function $\bar f \in L^2(\mu)$. The latter is invariant through T, and one has $\int f\, d\mu = \int \bar f\, d\mu$.*

If T is invertible, then f_n^- converges (in L^2) towards the same function $\bar f$. The function $\bar f$ is called the ergodic mean of f.

Proof. Due to the isometry of U_T, the Hilbert space $\mathcal{H} = L^2$ splits orthogonally between the invariant subspace $\mathcal{H}_0 = \ker(U_T - 1)$ and $\mathcal{H}_1 = \mathrm{Ran}(U_T - 1)$. As a result, one has

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} U_T^j = \Pi_0,$$

where Π_0 us the orthogonal projector on $\mathcal{H}_0$ (the limit holds in the strong operator topology). As a consequence, for any initial observable $f \in \mathcal{H}$, the time averages f_n converge (in L^2) towards $\bar f \stackrel{\text{def}}{=} \Pi_0 f$. If U_T is unitary, one easily checks that f_n^- has the same limit. Notice that the function $\bar f$ is an element of L^2, so it is defined a.e. □

Corollary 4.11. *[Von Neumann] Assume the transformation T is ergodic w.r.to the invariant measure μ. Then the ergodic average $\bar f$ is constant a.e.:*

$$\bar f(x) = \int f(x)\, d\mu(x) \qquad \mu - a.e.$$

The converse also holds.

The convergence of the *time* averages f_n towards an essentially constant function $\bar f$ equal to the *space* average of f is indeed what physicists have in mind by "ergodicity". Still, the convergence described in the above corollary

(in the L^2 sense: $\|f_n - \bar f\|_2 \overset{n\to\infty}{\to} 0$) is rather "weak". A more "physical" type of convergence is expressed by the following theorem.

Theorem 4.12. *[Birkhoff Ergodic Theorem] For any observable $f \in L^1(\mu)$, the limit*

$$\bar f(x) = \lim_{n\to\infty} f_n(x) \quad \textit{exists for a.e.}x,$$

is in L^1 and is T-invariant, satisfying $\int f\,d\mu = \int \bar f\,d\mu$. (if $f \in L^2$, this limit is the same as in the Von Neumann theorem).

If T is invertible, then $f_n^-(x)$ also converges to $\bar f(x)$ a.e.

Proof. This proof uses some "nontrivial" measure theory. Consider the sub σ-algebra $\mathcal{I}$ made of μ-invariant sets, and its restriction $\mu_{\mathcal{I}}$ on $\mathcal{I}$. From an observable f one constructs the signed measure $f\mu$, and its restriction $(f\mu)_{\mathcal{I}}$ on the σ-algebra $\mathcal{I}$. This restriction is absolutely continuous w.r.to $\mu_{\mathcal{I}}$, and we call its Radon-Nikodym derivative $f_{\mathcal{I}} = \left[\frac{(f\mu)_{\mathcal{I}}}{\mu_{\mathcal{I}}}\right]$. This is a function which is $\mathcal{I}$-measurable, hence T-invariant. Our aim is to show that $f_n(x) \to f_{\mathcal{I}}(x)$ a.e.

Define the increasing sequence of functions $F_n(x) \overset{\text{def}}{=} \max_{k\le n} k f_k(x)$. For a given $x \in X$, the sequence $(F_n(x))_{n\ge 1}$ is either bounded, or it diverges; the latter case defines the (invariant) set A_f. From the obvious $F_{n+1} - F_n \circ T = f - \min(0, F_n \circ T) \downarrow f$, so by dominated convergence one has

$$0 \le \int_{A_f} (F_{n+1} - F_n)\,d\mu \overset{n\to\infty}{\to} \int_{A_f} f\,d\mu = \int f_{\mathcal{I}}\,d\mu_{\mathcal{I}}.$$

Starting from some observable $\varphi \in L^1(\mu)$, we apply the above reasoning to $f = \varphi - \varphi_{\mathcal{I}} - \epsilon$. Obviously $f_{\mathcal{I}} \equiv -\epsilon < 0$, so the above inequality shows that $\mu(A_f) = 0$. One obviously has $f_n \le \frac{F_n}{n}$, so for any $x \notin A_f$ (that is, for μ-a.e. x) one has

$$\limsup_n f_n(x) = \limsup_n \varphi_n - \varphi_{\mathcal{I}} - \epsilon \le \limsup \frac{F_n(x)}{n} \le 0,$$

and hence $\limsup_n \varphi_n(x) \le \varphi_{\mathcal{I}}(x) + \epsilon$. a.e. Since this holds for any $\epsilon > 0$, we have $\limsup_n \varphi_n(x) \le \varphi_I(x)$ a.e. Applying the same reasoning to the observable $-\varphi$, we get $\liminf_n \varphi_n \ge \varphi_{\mathcal{I}}$ a.e. The two inequalities show that $\lim \varphi_n(x) = \varphi_{\mathcal{I}}(x)$ a.e. □

We end up this section on a connexion with topological dynamics. As we had noticed above, ergodicity is a measure-theoretic analogue of topological transitivity (∃ a dense orbit). We see below that this analogue is actually much more precise.

Proposition 4.13. *If $T : X \circlearrowleft$ is a continuous map, ergodic w.r.to μ, then the orbit of μ-a.e. point is dense in* $\operatorname{supp}\mu$.

4.4. **Mixing.** We now come to stronger chaotic properties.

Definition 4.14. A measure-preserving transformation $T : X \circlearrowleft$ on a probability space $(X, \mathfrak{U}, \mu)$ is *mixing* (w.r.to the invariant measure μ) iff for any measurable sets A, B, one has

$$\lim_{n\to\infty} \mu(T^{-n}(A) \cap B) = \mu(A)\,\mu(B). \tag{4.3}$$

Equivalently, for any bounded measurable functions f, g, one has

$$\lim_{n\to\infty} \int f(T^n(x))\,g(x)\,d\mu(x) = \int f(x)\,d\mu(x) \int g(x)\,d\mu(x).$$

This mixing property characterizes how the statistical **correlations** between two subsets A, B (resp. two observables f, g) evolve with time: mixing means that the correlations decay when the time $n \to \infty$. The system becomes "quasi-Markovian" in the long-time limit.

By a standard approximation procedure, one can show that

Proposition 4.15. *T is mixing iff, for any complete system Φ of functions in $L^2(\mu)$ and any $f, g \in \Phi$, one has*

$$\lim_{n\to\infty} \int f(T^n(x))\, g^*(x)\, d\mu(x) = \int f(x)\, d\mu(x) \int g^*(x)\, d\mu(x).$$

This property is at the heart of what is often understood by "chaos". It shows that, for any initial set A , each long time iterate $T^n(A)$ meets all regions of phase space. Split X into N components B_j of positive measures, and considers an initial set A of positive measure. Then, mixing means that for n large enough, the long time iterate $T^n(A)$ meets all sets B_j, and it does so approximately in a μ-distributed way.

Definition 4.16. This measure-theoretic notion is more precise than the corresponding topological notion.

Proposition 4.17. *If a continuous map f is mixing w.r.to an invariant measure μ, then it is topologically mixing on* $\operatorname{supp}\mu$*. (the converse is not necessarily true, but counterexamples are "pathological").*

Definition 4.18. A measure-preserving transformation T is weak mixing w.r.to the measure μ iff for any two measurable sets A, B one has

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \left|\mu(T^{-j}(A) \cap B) - \mu(A)\,\mu(B)\right| = 0.$$

Equivalently, there exists a set $J \subset \mathbb{N}$ of density one, such that

$$\lim_{J \ni n\to\infty} \mu(T^{-n}(A) \cap B) = \mu(A)\,\mu(B).$$

This notion appears less natural than mixing. It has the advantage to be easily expressible in terms of the isometry U_T:

Proposition 4.19. *Let T be an invertible measure-preserving transformation. T is weakly mixing w.r.to μ iff the isometry $U_T : L^2(\mu) \circlearrowleft$ has no eigenvalue except unity, which is simple.*

The 3 properties defined so far notions are clearly embedded:

Proposition 4.20. *Mixing implies weak mixing, which implies ergodicity.*

Proof. That mixing implies weak mixing is obvious. Assume $A \in \mathfrak{U}$ is invariant. Then, one has $0 = \mu(A \cap \complement A) = \mu(A)\mu(\complement A)$, so $\mu(A) = 0$ or $\mu(A) = 1$. □

4.5. **Examples of ergodic and mixing transformations.** We can now scroll our list of examples and study their measure-theoretic properties w.r.to some "natural" invariant measures. Quite often, mixing or ergodicity will be easier to prove from the "observable" point of view than the "subset" point of view.

4.5.1. *Rotations on S^1.* A natural invariant measure is the Lebesgue measure μ_L on S^1. We find the same dichotomy as in §2.3:

(1) if the angle α is rational, μ_L is not ergodic. Besides, the map R_α admits many other invariant measures.
(2) if α is irrational, μ_L is ergodic. To see this, we expand any function $f \in L^2(S^1)$ in Fourier series, and check whether the function can be invariant. Actually, one can prove that R_α is **uniquely ergodic**: μ_L is its unique invariant measure.

It is relevant at this stage to introduce a more constraining notion, which applies only to continuous maps.

Definition 4.21. A continuous map $T : X \circlearrowleft$ is **uniquely ergodic** iff it admits a unique (Borel) invariant measure.

Remark 4.22. The unique measure is then automatically ergodic.

One can also characterize unique ergodicity from the behaviour of Birkhoff averages.

Proposition 4.23. *A map $T : X \circlearrowleft$ is uniquely ergodic iff for any continuous observable $f \in C(X)$, the Birkhoff averages f_n converge **uniformly** to a constant when $n \to \infty$. (that constant is equal to $\int f\, d\mu$, where μ is the unique invariant measure).*

Let us turn back to the irrational translations. Any continuous function on S^1 can be approximated by a trigonometric polynomial[9] $f^{(K)}(x) = \sum_{|k|\leq K} \hat{f}_k\, e_k(x)$, so we only need to prove uniform convergence of Birkhoff averages for such polynomials. By linearity, we only need to prove it for each individual Fourier mode e_k, $k \in \mathbb{Z} \setminus 0$.

$$\forall n \geq 1, \qquad (e_k)_n(x) = \frac{1}{n} \sum_{j=0}^{n-1} e_k(x + j\alpha) = \frac{1}{n}\, \frac{1 - e_k(n\alpha)}{1 - e_k(\alpha)}\, e_k(x)$$
$$\Longrightarrow \|(e_k)_n\|_\infty \leq \frac{1}{n}\, \frac{2}{|1 - e_k(\alpha)|} \overset{n\to\infty}{\to} 0.$$

Remark 4.24. The irrational translation R_α is not weakly mixing. Indeed, any Fourier mode e_k is an eigenvector of U_{R_α}, with eigenvalue $e_k(\alpha)$. The absence of mixing reminds us of the fact that R_α is not topologically mixing.

Irrational translation flow on $\mathbb{T}^2$. One can suspend the irrational rotations R_α on S^1, using the constant function $\tau(x) = 1$ as ceiling function: the flow obtained is equivalent with the translation flow $T^t_\alpha : (x, y) \mapsto (x + \alpha t, y + t)$ on $\mathbb{T}^2$. This flow is also uniquely ergodic: for any nontrivial Fourier mode $e_{\mathbf{m}}$, $\mathbf{m} = (m_1, m_2) \neq (0, 0)$, one has

$$\forall x \in \mathbb{T}^2, \qquad \frac{1}{T} \int_0^T dt\, e_{\mathbf{m}}(T^t_\alpha(x)) = \frac{1}{T} \int_0^T dt\, e^{2i\pi(m_1\alpha + m_2)t}\, e_{\mathbf{m}}(x) = \frac{1}{T}\, \frac{e_{\mathbf{m}}(\alpha, 1) - 1}{m_1\alpha + m_2}\, e_{\mathbf{m}}(x) \overset{T\to\infty}{\to} 0, \tag{4.4}$$

so Prop. 4.23 (generalized to flows) implies unique ergodicity of T^t_α, the unique invariant measure being Lebesgue.

4.5.2. *Linear dilations on S^1.* We have already noticed that each E_m leaves the Lebesgue measure μ_L invariant, since for a short enough interval I the preimage $E_m^{-1}(I)$ consists in m intervals of length $\frac{|I|}{m}$.

Proposition 4.25. *The map E_m is mixing w.r.to μ_L.*

Proof. We use Proposition 4.15 applied to the Fourier basis $\{e_k,\, k \in \mathbb{Z}\}$ of $L^2(S^1)$. For any two Fourier modes e_k, e_l, we have

$$\forall n \geq 0, \quad \int \bar{e}_k\, e_l \circ E_m^n\, d\mu_L = \int \bar{e}_k\, e_{lm^n}\, d\mu_L = \delta_{k,lm^n}.$$

For any fixed $(k, l) \neq (0, 0)$, this integral vanishes for n large enough, that is converges to $\int \bar{e}_k\, d\mu_L \int e_l\, d\mu_L$. □

One can also show the mixing by using the the topological semiconjugacy (2.1) between E_m and the full shift Σ^+_m (see Ex.4.27 below).

[9] we denote by $e_k(x) = e^{2i\pi kx}$ the k-th Fourier mode on S^1.

Exponential mixing. The above proof shows the decay of correlations for any two observables $f, g \in L^2(S^1)$. By requiring more regularity of the observables, one is often able to have a better control on the speed of decay of the correlations. In the present case, one can easily show that correlations between C^p observables ($p > 0$) decay exponentially. Indeed, for any $f \in C^p(S^1)$ the Fourier coefficients $\hat{f}_k$ decay as

$$\forall k \neq 0, \quad |\hat{f}_k| \leq \frac{\|f\|_{C^p}}{|k|^p},$$

therefore the above computations show that for $f, g \in C^p$ one has

$$\left| \int f\, g \circ E_m^n \, dx - \hat{f}_0 \hat{g}_0 \right| = \left| \sum_{l \neq 0} \hat{f}_{-lm^n} \hat{g}_l \right| \leq m^{-np} \sum_{l \neq 0} \frac{\|f\|_{C^p} \|g\|_{C^p}}{|l|^{2p}},$$

showing that the exponential decay of correlations for C^p observables.

One proof proceeds by using the spectral analysis of a **transfer operator** associated with the dynamics. Above we have introduced the operator $U_T : f \mapsto f \circ T$, which is an isometry on any L^p. This operator is not very appropriate to deal with regular observables, since the function $f \circ T$ is generally less singular than f. It is then more convenient to consider the *dual operator* $\mathcal{L}_T$, defined by

$$\int f\, U_T g \, dx = \int (\mathcal{L}_T f)\, g \, dx,$$

which gives (in the case of an expanding map T on S^1):

$$\mathcal{L}_T f(x) = \sum_{y : Ty = x} \frac{f(y)}{|T'(y)'|} = \frac{1}{m} \sum_{j=0}^{m-1} f\left(\frac{x+j}{m} \right).$$

The correlations w.r.to the Lebesgue measure can be directly expressed in terms of this transfer operator:

$$\int f(x)\, g(T^n(x))\, dx = \int (\mathcal{L}_T^n f)(x)\, g(x)\, dx.$$

The crucial advantage of $\mathcal{L}_T$ is that its spectrum is quasicompact on any C^p, $p > 1$: it has a simple eigenvalue unity, no other eigenvalue on S^1, and the rest of the spectrum lies in some smaller disk $\{|\lambda| \leq r_p\}$, with $r_p < 1$. As a result, for any $f \in C^p$ one has the following expansion:

$$\mathcal{L}_T^n f = \langle 1, f \rangle 1 + r_p^n \mathcal{R}_n f, \quad \|\mathcal{R}_n f\|_{C^p} \leq C \|f\|_{C^p},$$

and therefore

$$\int (\mathcal{L}_T^n f)(x)\, g(x)\, dx = \langle 1, f \rangle \langle 1, g \rangle + \mathcal{O}_{f,g}(r_p^n).$$

In the present case, on can show that $r_p = m^{-p}$: the smoother the observables, the faster the decay.

4.5.3. *Full shift* $\Sigma_m^{(+)}$. One can easily construct shift-invariant probability measures ν on Σ_m^+:

Definition 4.26. Consider a **probability distribution** $\boldsymbol{p} = \{p_0, \ldots, p_{m-1}\}$, satisfying $p_k \geq 0$, $\sum_{k=0}^{m-1} p_k = 1$. To this distribution is associated a single Borel probability measure $\nu_{\boldsymbol{p}}$ on Σ_m^+, which takes the following weights on cylinders:

$$\forall n \geq 1, \ \forall \epsilon_1 \cdots \epsilon_n, \qquad \nu_{\boldsymbol{p}}(C_{\epsilon_1 \epsilon_2 \cdots \epsilon_n}) = \prod_{i=1}^{n} p_{\epsilon_i}. \tag{4.5}$$

This measure is obviously shift-invariant. It is called the Bernoulli measure associated with the distribution $\boldsymbol{p}$. Its statistical meaning is obvious: at each time the particle has a probability p_i to be in the slot i, without any dependence on its past position.

Let us come back to the dilation E_m for a moment. Any σ-invariant measure can then be pulled-back through π to a measure on S^1:

$$\mu = \pi^*\nu \Longleftrightarrow \forall A \subset S^1,\ \mu(A) = \nu(\pi^{-1}(A)).$$

The measure μ is then automatically invariant through E_m:

$$\forall A \subset \mathbb{T}^2, \quad \mu(E_m^{-1}(A)) = \nu(\pi^{-1}(E_m^{-1}(A)) = \nu(\sigma^{-1}(\pi^{-1}(A)) = \nu(\pi^{-1}(A)) = \mu(A).$$

Exercise 4.27. The Lebesgue measure μ_L is the pull-back through π of the Bernoulli measure $\mu_{\boldsymbol{p}_{\max}}$ with $\boldsymbol{p}_{\max} = \left\{\frac{1}{m}, \ldots, \frac{1}{m}\right\}$. The topological semiconjugacy π is a measure-theoretic *isomorphism* between (E_m, μ_L) and $(\Sigma_m^+, \nu_{\boldsymbol{p}_{\max}})$.

Proposition 4.28. *The full one-sided shift Σ_m^+ is mixing w.r.to any Bernoulli measure $\mu_{\boldsymbol{p}}$.*

Proof. We use the fact that the cylinders $\{C_{\boldsymbol{\epsilon}}\}$ generate the topology on Σ_m^+, and therefore the Borel σ-algebra. For any two cylinders $C_{\boldsymbol{\alpha}}, C_{\boldsymbol{\beta}}$ of lengths $m > 0$, we have

$$\forall n > m, \quad \nu_{\boldsymbol{p}}(C_{\boldsymbol{\alpha}} \cap \sigma^{-n} C_{\boldsymbol{\beta}}) = \nu_{\boldsymbol{p}} \left(\bigcup_{x_{m+1}, \ldots, x_n} C_{\alpha_1 \cdots \alpha_m x_{m+1} \cdots x_n \beta_1 \cdots \beta_m} \right) = \nu_{\boldsymbol{p}}(C_{\boldsymbol{\alpha}})\, \nu_{\boldsymbol{p}}(C_{\boldsymbol{\beta}}).$$

Equivalently, the two cylinders have become *statistically independent* of each other after time m. □

We have exhibited a whole family of mixing (so, in particular, ergodic) probability measures on Σ_m^+. Exactly the same construction can be performed on the two-sided full shift Σ_m. The following property shows that these measures are really different from one another.

Proposition 4.29. *Any two Bernoulli measures $\nu_{\boldsymbol{p}} \neq \nu_{\boldsymbol{p}'}$ are singular with one another*[10].

This result is a particular case of a more general one:

Proposition 4.31. *If $\mu \neq \nu$ are two ergodic probability measures for a transformation T, then they are mutually singular.*

Proof. The Lebesgue decomposition theorem states that for any pair $\mu, \nu \in \mathcal{M}(X)$, the measure μ can be uniquely split into $\mu = p\mu_1 + (1-p)\mu_2$, where μ_1 is absolutely continuous w.r.to ν, while μ_2 is singular w.r.to ν. Since μ and ν are invariant, this decomposition is also invariant: $\mu_1, \mu_2 \in \mathcal{M}(X,T)$. Because μ is extremal, one must have $p = 0$ or $p = 1$. □

In Prop.4.29 we notice an apparent paradox: the measures $\mu_{\boldsymbol{p}}$ varies continuously (in the weak-* sense) with $\boldsymbol{p}$, but no matter how "close" two measures $\mu_{\boldsymbol{p}} \neq \mu_{\boldsymbol{p}'}$ are, they are supported on complementary sets. To get some "feeling" on the nature of these sets, let us concentrate on the case $m = 2$. For any $\alpha \in [0,1]$, define the subset $F_\alpha \stackrel{\text{def}}{=} \left\{ \boldsymbol{\epsilon} \in \Sigma_2^+,\ \frac{\#\{\epsilon_i = 0,\, i=1,\ldots,n\}}{n} \stackrel{n\to\infty}{\to} \alpha \right\}$. It consists of sequences which have a well-defined asymptotic frequency to be equal to zero, and this frequency is α. Obviously, $F_\alpha \cap F_\beta = \emptyset$ if $\alpha \neq \beta$, and $\bigsqcup_{\alpha\in[0,1]} F_\alpha \subset \Sigma_2^+$. We claim that $\mu_{(\alpha,1-\alpha)}(F_\alpha) = 1$. Equivalently, with $\mu_{(\alpha,1-\alpha)}$-probability 1, a point $\boldsymbol{\epsilon} \in \Sigma_2^+$ has asymptotic frequency α. This property explicitly shows that $\mu_{(\alpha,1-\alpha)}$ and $\mu_{(\beta,1-\beta)}$ are singular.

10

Proposition 4.30. *Two measures μ, ν are singular with one another iff, there exists a subset $A \subset X$ such that $\mu(A) = 1$ while $\nu(A) = 0$.*

4.5.4. *Linear toral automorphisms.* The situation of the linear automorphisms M on $\mathbb{T}^d$ (with a matrix $M \in GL(d,\mathbb{Z})$ is quite similar with the case of E_m above. First of all, the Lebesgue measure μ_L is invariant due to $|\det(M)| = 1$.

Proposition 4.32. *A hyperbolic linear automorphism M is mixing w.r.to μ_L.*

Proof. Once more, we use Prop.4.15 and the Fourier basis $\{e_{\mathbf{m}},\ \mathbf{m} \in \mathbb{Z}^d\}$ of $L^2(\mathbb{T}^d)$. For any $(\mathbf{m},\mathbf{n}) \neq (0,0)$, we have

$$\int \bar{e}_{\mathbf{m}}(x)\, e_{\mathbf{n}}(M^n x)\, d\mu_L(x) = \int \bar{e}_{\mathbf{m}}(x)\, e_{{}^tM^n\mathbf{n}}(x)\, d\mu_L(x) = \delta_{\mathbf{m},{}^tM^n\mathbf{n}}.$$

Since the orbits $\{{}^tM^n\mathbf{n},\ n \in \mathbb{Z}\}$ all go to infinity due to hyperbolicity, this integral vanishes provided n is large enough, showing the mixing, and therefore the ergodicity □

A "more intuitive" proof of mixing, which will be easier to generalize to nonlinear Anosov maps, uses the unstable manifolds of M. Let us restrict ourselves to the 2-dimensional case. To prove mixing, it is sufficient to show (4.3) with A, B two rectangles aligned with the unstable and stable directions. For n large, $M^{-n}(A)$ is another rectangle, very elongated along the stable direction (of length $l_-\lambda^n$), and very thin (of length $l_+\lambda^{-n}$) along the unstable one. It is thus a "thickening" of a long stable segment of length $l_-\lambda^n$. Since the stable direction is irrational, stable segments become dense on $\mathbb{T}^2$ in the limit of large length; more precisely, the measure carried by the segment converges to the Lebesgue measure on $\mathbb{T}^2$ (this statement is equivalent with (4.4), namely the unique ergodicity of irrational translation flows).

Remark 4.33. It is also possible to show that correlations decay exponentially for smooth enough observables. The proof is a little more subtle than in §4.5.2:

$$\left| \int f \circ M^{-n}\, g \circ M^n\, dx - \hat{f}_0 \hat{g}_0 \right| = \left| \sum_{\mathbf{m}\neq 0} \hat{f}_{-{}^tM^{-n}\mathbf{m}} \hat{g}_{{}^tM^n\mathbf{m}} \right| \leq \sum_{\mathbf{m}\neq 0} \frac{\|f\|_{C^p}\, \|g\|_{C^p}}{|{}^tM^{-n}\mathbf{m}|^p\, |{}^tM^n\mathbf{m}|^p}.$$

To estimate the RHS, one needs to decompose the vector $\mathbf{m}$ into the dual stable/unstable basis: $\mathbf{m} = m_+\mathbf{e}_+ + m_-\mathbf{e}_-$. First consider the sector $\{|m_-| \leq |m_+|\}$. In that sector,

$$\left|{}^tM^{-n}\mathbf{m}\right| \left|{}^tM^n\mathbf{m}\right| \simeq \left(\lambda^{-n}|m_+| + \lambda^n|m_-|\right)\left(\lambda^n|m_+| + \lambda^{-n}|m_-|\right) = m_+^2 + m_-^2 + \lambda^{2n}|m_+m_-| \simeq |\mathbf{m}|^2 + \lambda^{2n}|m_+m_-|.$$

Since $\mathbf{m}$ is on a hyperbola intersecting $\mathbb{Z}^2 \setminus 0$, the product $|m_+m_-|$ is uniformly bounded from below by some $c > 0$. Hence, for $p \geq 3$ the sum on this sector is bounded from above by

$$\frac{C}{\lambda^{2np}} \sum_{\mathbf{m}\in sector} \frac{1}{|\mathbf{m}|^p} \leq \frac{C'}{\lambda^{2np}}.$$

The other sectors are treated similarly. This proves the exponential decay of correlations when $p \geq 3$.

4.5.5. *Markov measures on topological Markov chains.* We now consider a topological Markov chain $\Sigma_A^{(+)}$ defined by an adjacency matrix A. Since the particle cannot jump from i to any j, one cannot produce an invariant measure as simple as (4.5). Instead, one can impose some *statistical weights* on the transitions $i \to j$, that is replace the adjacency matrix by a Markov matrix Π, such that Π_{ij} is the probability to jump from $i \to j$. This matrix has the following properties:

(1) $\Pi_{ij} = 0$ iff $A_{ij} = 0$
(2) $\forall i$, $\sum_j \Pi_{ij} = 1$ (the matrix Π is *stochastic*).

This matrix provides a statistical complement onto the mere topological information given by A. The resulting system is a Markov chain (or Markov process). Let us assume that Π is *primitive*. To this process is then

associated a natural invariant measure, called the Markov measure, which can be constructed as follows. By stochasticity, Π admits 1 as (largest) eigenvalue, associated with the right eigenvector $\mathbf{1} = (1, 1, \cdots)$. It also has a (unique) left eigenvector with positive entries, which we can normalize as $\boldsymbol{p} = \boldsymbol{p}_\Pi = (p_0, p_1, \ldots, p_{m-1})$ with $\sum_i p_i = 1$. We can now define a measure on Σ_A^+ as follows:

$$\forall n \geq 1,\ \forall \epsilon_0 \epsilon_1 \cdots \epsilon_n, \qquad \nu_\Pi(C_{\epsilon_0 \epsilon_1 \epsilon_2 \cdots \epsilon_n}) = p_{\epsilon_0} \Pi_{\epsilon_0 \epsilon_1} \cdots \Pi_{\epsilon_{n-1} \epsilon_n}.$$

One easily checks that this measure satisfies

(1) the compatibility condition $\nu_\Pi(C_{\epsilon_1 \cdots \epsilon_n}) = \sum_{\epsilon_{n+1}} \nu_\Pi(C_{\epsilon_1 \cdots \epsilon_n})$ (from the stochasticity)
(2) the shift-invariance $\nu_\Pi(C_{\epsilon_1 \cdots \epsilon_n}) = \sum_{\epsilon_0} \nu_\Pi(C_{\epsilon_0 \epsilon_1 \cdots \epsilon_n})$ (from $\boldsymbol{p}\Pi = \boldsymbol{p}$).

The measure ν_Π is called the Markov measure of the Markov chain Π. It can obviously be extended to an invariance measure of the two-sided subshift Σ_A. The support of ν_Π is the full subshift Σ_A, and ν_Π has no atomic component.

Proposition 4.34. *Assume Π is a primitive stochastic matrix associated with the adjacency matrix A. Then the shift (Σ_A, σ) is mixing w.r.to ν_Π.*

Proof. As in the case of the Bernoulli measure on the full shift, we consider correlations between cylinders $C_{\boldsymbol{\alpha}}, C_{\boldsymbol{\beta}}$ of length m. For any $n > m$, a little algebra shows that

$$\nu_\Pi(C_{\boldsymbol{\alpha}} \cap \sigma^{-n} C_{\boldsymbol{\beta}}) = \nu_{\boldsymbol{p}} \left(\bigcup_{x_{m+1}, \ldots, x_n} C_{\alpha_1 \cdots \alpha_m x_{m+1} \cdots x_n \beta_1 \cdots \beta_m} \right) = \nu_{\boldsymbol{p}}(C_{\boldsymbol{\alpha}})\, \nu_{\boldsymbol{p}}(C_{\boldsymbol{\beta}}) \frac{(\Pi^{n-m})_{\alpha_m \beta_1}}{p_{\beta_1}}.$$

Since Π is primitive and stochastic, its large powers satisfy $\Pi^N = \mathbf{1} \otimes \boldsymbol{p} + \mathcal{O}(\mu^N)$ for some $0 < \mu < 1$, so in particular $\left(\Pi^N\right)_{ij} = p_j + \mathcal{O}(\mu^N)$. This shows that we have *exponential mixing* when considering cylinders. □

5. Complexity and Entropies

So far we have described dynamical systems by providing some *qualitative* properties, like topological transitivity/mixing, ergodicity/mixing... These properties describe the recurrence properties of the system, in a more or less precise way. Among the examples of smooth maps we gave, it occured that the strongest form of chaos (strong mixing) occured for the case of hyperbolic systems. However, no direct connection between positive Lyapunov exponents and mixing has been established yet.

In this section we will define and analyze the **entropies** associated with either a topological DS (X,T) or a measured DS (X,T,μ). These entropies provide a **quantitative estimate of the complexity of the DS.** The entropy associated with a DS (X,T) or (X,T,μ) is a nonnegative (possibly infinite) number. The "more complex" the system, the larger its entropy. All systems we will consider have finite entropies.

The measure-theoretic entropy was introduced in the 1950' by Khinchin and Kolmogorov, by extending the Shannon entropy (of information theory) to the DS setting. Its use was then simplified drastically due to a remark by Sinai; for this reason, this measure-theoretic entropy is often called the Kolmogorov-Sinai entropy of (X,T,μ), and denoted by $H_{KS}(\mu,T)$. We will analyze it in §.

The topological entropy associated with a continuous map (X,T) was introduced in 1965 by Adler-Konheim-McAndrew, by adapting the definition of the measure-theoretic entropy to the topological framework. The more "intuitive" definition (given below) is due to Dinaburg and Bowen.

5.1. Measure-theoretic (Kolmogorov-Sinai) entropy.

5.1.1. *Entropy associated with a partition.* The idea of entropy comes from a "rough" description of the phase space X associated with a finite partition $\mathcal{P} = \{P_1, \dots, P_K\}$ of X. This description is done by the following "experimental" scheme: the "observer" has only a rough camera at his disposal, which is only able to detect whether the point belongs to P_1, or to P_2 etc. Each P_k is a "pixel" of the observer's camera.

Let us assume that the statistical distribution of the points on X is given by the probability measure μ. One first wants to define a measure of the **uncertainty** present before performing the observation, or equivalently the **quantity of information gained** after this observation. This uncertainty is necessarily related with the a priori probability of finding the particle in P_1, or in P_2 etc, that is with the probability distribution $\boldsymbol{p} = \Big\{ p_k \stackrel{\text{def}}{=} \mu(P_k),\ k = 1, \dots, K \Big\}$. If the measure μ is fully supported in P_1, there is no uncertainty: the observer will certainly observe the particle to be in P_1. On the other hand, if the measure μ is equally distributed among all subsets P_i (that is, $p_k = \mu(P_k) = \frac{1}{K}$ for all k), then the uncertainty of the observation is "maximal". The uncertainty (*entropy*) associated with this observation is therefore a function $H(\mu,\mathcal{P}) = H(\mathcal{P}) = H(\boldsymbol{p})$ which satisfies the following properties:

(1) $H(\boldsymbol{p}) \geq 0$, with equality iff some $p_i = 1$.
(2) $H(\boldsymbol{p})$ depends continuously of $\boldsymbol{p}$, and is symmetric in the p_i.
(3) $H(\boldsymbol{p})$ is maximal for $\boldsymbol{p} = \boldsymbol{p}_{\max} = \left\{ \frac{1}{K}, \cdots, \frac{1}{K} \right\}$
(4) if one element has a zero probability (say, $p_K = 0$), then the entropy is the same as $H(\{p_1, p_2, \dots, p_{K-1}\})$.

These conditions do not constrain the function H completely. To do so, we need to impose a condition relative to the use of a second partition $\mathcal{Q} = \{Q_1, \dots, Q_L\}$ (that is, a second measuring device). The observer has now the possibility to first "$\mathcal{Q}$-observe", and then $\mathcal{P}$-observe. If the point has been observed to be in Q_1, then the uncertainty before performing the $\mathcal{P}$-observation will be given by $H\left(\left\{ \frac{\mu(Q_1 \cap P_1)}{\mu(Q_1)}, \cdots \frac{\mu(Q_1 \cap P_K)}{\mu(Q_1)} \right\} \right)$. One then

averages over the output Q_1, to define the **conditional entropy** of the $\mathcal{P}$-observation, taking into accout the $\mathcal{Q}$-observation:

$$H(\mu, \mathcal{P}|\mathcal{Q}) = \sum_{l=1}^{L} \mu(Q_l)\, H\left(\left\{\frac{\mu(Q_1 \cap P_1)}{\mu(Q_1)}, \cdots \frac{\mu(Q_1 \cap P_K)}{\mu(Q_1)}\right\}\right).$$

From the two partitions $\mathcal{P}, \mathcal{Q}$ one can naturally define their *common refinement*, namely the partition

$$\mathcal{P} \vee \mathcal{Q} = \{P_k \cap Q_l,\, k = 1, \ldots, K,\; l = 1, \ldots, L\},$$

meaning that the observations $\mathcal{P}, \mathcal{Q}$ are performed simultaneously. It seems therefore natural to assume the following property:

$$H(\mu, \mathcal{P} \vee \mathcal{Q}) = H(\mu, \mathcal{Q}) + H(\mu, \mathcal{P}|\mathcal{Q}),$$

Together with the previous conditions, this last property fully constrains the entropy function (up to a scalar factor, which we set to unity): it must be given by

$$H(\mu, \mathcal{P}) = \sum_{k=1}^{K} \eta(\mu(P_k)), \quad \text{with} \quad \eta(s) \stackrel{\text{def}}{=} -s \log s. \tag{5.1}$$

This is exactly the definition given by Boltzmann (in statistical mechanics), and by Shannon (in information theory).

The function $\eta(s)$ is concave on $[0, 1]$, which has important consequences:

(1) if the partition $\mathcal{P}$ is *finer* than $\mathcal{Q}$ (or $\mathcal{P}$ is a refinement of $\mathcal{Q}$), meaning that each P_k is contained in some Q_l (or for short $\mathcal{P} \vee \mathcal{Q} = \mathcal{P}$), then $H(\mathcal{P}) \geq H(\mathcal{Q})$.
(2) the entropy is *subadditive*: for any two partitions $\mathcal{P}, \mathcal{Q}$, one has

$$H(\mathcal{P} \vee \mathcal{Q}) \leq H(\mathcal{P}) + H(\mathcal{Q}),$$

with equality iff $\mathcal{P}$ and $\mathcal{Q}$ are statistically independent.
(3) for any partition $\mathcal{P} = \{P_1, \ldots, P_K\}$, one has $H(\mathcal{P}) \leq \log K$, with equality iff all $\mu(P_k) = \frac{1}{K}$.

5.1.2. *Taking the dynamics into account.* So far we have described the entropy associated with the "observations at time 0", namely without taking any dynamics into account. Assume that after performing the $\mathcal{P}$-observation, the observer lets the system evolve under the map T (which could be the stroboscopic map of a continuous flow), then it observes the system again (with the same partition $\mathcal{P}$), lets it evolve again, observe etc.. At time n, he will have performed n observations, by recording which elements P_k the particle belongs to at the times $0, 1, \ldots n-1$. Each initial point $x \in X$ has been associated with a finite string $x_0 x_1 \cdots x_{n-1}$, defined by its "symbolic history" between the times 0 and $n-1$:

$$\forall j = 0, \ldots n-1, \quad T^j(x) \in P_{x_j} \Longleftrightarrow x \in P_{x_0 \cdots x_{n-1}} \stackrel{\text{def}}{=} P_{x_0} \cap T^{-1} P_{x_1} \cap \cdots \cap T^{-n+1} P_{x_{n-1}}.$$

Through this mechanism, the dynamics T naturally creates successive refinements $\mathcal{P}^{\vee n} = \left\{P_{x_0 \cdots x_{n-1}}\right\}$ of the initial partition $\mathcal{P}$. Hence, the uncertainty associated with this refined partitions *increases* with n: $H(\mathcal{P}^{\vee n}) \geq H(\mathcal{P}^{\vee (n-1)})$.

Now the question asked by the observer is the following: what is the *rate of increase* of the entropies $H(\mathcal{P}^{\vee n})$ associated with these successive refinements? By subadditivity, we can define

$$\lim_{n \to \infty} \frac{H(\mathcal{P}^{\vee n})}{n} = \inf_{n \geq 1} \frac{H(\mathcal{P}^{\vee n})}{n} \stackrel{\text{def}}{=} H(T, \mathcal{P})$$

as the entropy associated with the map (T, μ) and the initial partition $\mathcal{P}$.

Another way to define $H(\mu, T, \mathcal{P})$ is to consider the following question: if I know the symbolic history of the particle from time 1 up to time n, how much new information do I obtain by performing the observation at time 0? In other terms, what is the entropy of the partition $\mathcal{P}$, conditioned by the partition $T^{-1}\mathcal{P}^{\vee n}$? One can indeed show that

$$H(T, \mathcal{P}) = \lim_{n\to\infty} H(\mathcal{P}|T^{-1}\mathcal{P}^{\vee n}). \tag{5.2}$$

Proof.

$$\begin{aligned} H(\mathcal{P}^{\vee n}) &= H(\mathcal{P} \vee T^{-1}\mathcal{P}^{\vee n-1}) = H(T^{-1}\mathcal{P}^{\vee n-1}) + H(\mathcal{P}|T^{-1}\mathcal{P}^{\vee n-1}) \\ &= H(\mathcal{P}^{\vee n-1}) + H(\mathcal{P}|T^{-1}\mathcal{P}^{\vee n-1}) \\ &= H(\mathcal{P}) + \sum_{j=1}^{n} H(\mathcal{P}|T^{-1}\mathcal{P}^{\vee j-1}). \end{aligned}$$

The sequence $H(\mathcal{P}|T^{-1}\mathcal{P}^{\vee j-1})$ is decreasing w.r.to j since the "denominator" gets refined. Dividing by n and taking the limit, the limit of the Cesaro mean on the RHS is identical with the limit (5.2). □

Remark 5.1. If the map T is invertible, one also has

$$H(T, \mathcal{P}) = \lim_{n\to\infty} H(T^{-n}\mathcal{P}|\mathcal{P}^{\vee n}),$$

that is, the asymptotic information provided by the observation at time n, taking into account the observations at times 0 through $n-1$.

The entropy $H(T, \mathcal{P})$ inherits properties of $H(\mathcal{P})$, plus some others:

(1) $0 \le H(T, \mathcal{P}) \le H(\mathcal{P})$
(2) $H(T, \mathcal{P}) = H(T, T^{-1}\mathcal{P})$. If T is invertible, $H(T, \mathcal{P}) = H(T, T\mathcal{P})$
(3) $H(T, \mathcal{P}) = H(T, \mathcal{P}^{\vee n})$ for any $n \ge 0$. If T is invertible, $H(T, \mathcal{P}) = H(T, \bigvee_{-n}^{n} T^{-j}\mathcal{P})$
(4) $H(T, \mathcal{P} \cap \mathcal{Q}) \le H(T, \mathcal{P}) + H(T, \mathcal{Q})$
(5) $|H(T, \mathcal{P}) - H(T, \mathcal{Q})| \le H(\mathcal{P}|\mathcal{Q}) + H(\mathcal{Q}|\mathcal{P}) \stackrel{\text{def}}{=} d_R(\mathcal{P}, \mathcal{Q})$ (the *Rokhlin distance* between the partitions $\mathcal{P}, \mathcal{Q}$).

5.1.3. *Optimizing over the initial partition.* So far our definition of the entropy is associated with some initial partition ("observation") $\mathcal{P}$. A crucial step was achieved by Kolmogorov and Sinai, who understood that the partition could be chosen with some freedom. They first defined the KS entropy as the supremum over all finite initial partitions:

$$H_{KS}(T, \mu) = H_{KS}(T) \stackrel{\text{def}}{=} \sup_{\mathcal{P}} H(T, \mathcal{P}).$$

In this form, the quantity is hardly computable. The crucial step (due to Sinai) was to show that this supremum is attained if the initial partition $\mathcal{P}$ *generate*s the full σ-algebra.

Definition 5.2. A partition $\mathcal{P}$ is said to be *generating* (w.r.to (T, μ)) iff, for any finite partition $\mathcal{Q}$ and any $\epsilon > 0$, there exists $n \ge 0$ such that $\mathcal{P}^{\vee n}$ is finer than some partition $\mathcal{Q}_n$ which is at ϵ-distance from $\mathcal{Q}$ (w.r.to the Rokhlin metric). In the case of μ a nonatomic Borel measure, $\mathcal{P}$ is a generator if the diameter of $\mathcal{P}^{\vee n}$ goes to zero as $n \to \infty$.

For T invertible, $\mathcal{P}$ is generating iff for any finite $\mathcal{Q}$ and any $\epsilon > 0$, $\bigvee_{-n}^{n} T^{-j}\mathcal{P}$ is finer than some partition $\mathcal{Q}_n$ which is ϵ-close from $\mathcal{Q}$.

Notice that the atomic parts of an invariant measure are not very interesting from the complexity point of view: any atomic ergodic measure (supported on a periodic orbit) has zero entropy.

Theorem 5.3. *[Sinai] Let $\mathcal{P}$ be a generating partition for (T,μ). Then $H_{KS}(T) = H(T,\mathcal{P})$.*

This result allows one to directly compute the KS entropy for many transformations (see below). It was crucial in the development of the entropy as a *central measure-theoretic invariant of the dynamics*.

We now provide some properties of $H_{KS}(T,\mu)$:

(1) The entropy is **additive w.r.to a time rescaling**: for any $k \in \mathbb{N}$, $H_{KS}(T^k) = kH_{KS}(T)$. If T is invertible, $H_{KS}(T^{-1}) = H_{KS}(T)$.
(2) If (S,ν) and (T,μ) are measure-theoretically isomorphic, then $H_{KS}(S,\nu) = H_{KS}(T,\mu)$. The entropy is thus **an invariant of the measure-theoretic dynamical system**.
(3) If $A \subset X$ is invariant and $\mu(A) > 0$, then

$$H_{KS}(T,\mu) = \mu(A)H_{KS}(T \restriction A, \mu_A) + \mu(\complement A)H_{KS}(T \restriction \complement A, \mu_{\complement A}). \tag{5.3}$$

(4) The entropy is **affine**. If μ,ν are two probability measures preserved by T, then for any $\alpha \in [0,1]$ the proba measure $\alpha\mu + (1-\alpha)\nu$ is also invariant. One then has

$$H_{KS}\left(T, \alpha\mu + (1-\alpha)\nu\right) = \alpha H_{KS}(T,\mu) + (1-\alpha)H_{KS}(T,\nu).$$

This property extends to the (possibly noncountable) ergodic decomposition (4.2). It shows that one only needs (in principle) to compute the entropies of *ergodic* invariant measures.

5.1.4. *Alternative definitions of the entropy.* From its definition, the entropy $H(T,\mathcal{P})$ represents the average exponential decay rate of the weights $\left\{\mu(P_{\alpha_0\cdots\alpha_{n-1}}),\ |\boldsymbol{\alpha}| = n\right\}$. If all these elements have weights $\mu(P_{\alpha_0\cdots\alpha_{n-1}}) \leq Ce^{-\beta n}$, uniformly w.r.to n then $H(T,\mathcal{P}) \geq \beta$. This result can be made more quantitative in the case of ergodic transformations.

For any $x \in X$, let us call $P^{\vee n}(x)$ the unique element $P_{\alpha_0\cdots\alpha_{n-1}}$ of $\mathcal{P}^{\vee n}$ containing x (equivalently, the point x has the symbolic history $\alpha_0\cdots\alpha_{n-1}$). The function

$$I_n(x) \stackrel{\text{def}}{=} -\log\left(\mu\left(P^{\vee n}(x)\right)\right)$$

is called the information function w.r.to the partition $\mathcal{P}^{\vee n}$. We see that the entropy of the partition $\mathcal{P}^{\vee}$ is nothing by the μ-average of I_n. The following theorem give a more precise result in case of transformations measures.

Theorem 5.4. *[Shannon-McMillan-Breiman] Let T be ergodic w.r.to the invariant measure μ, and let $\mathcal{P}$ be a finite partition. Then,*

$$\lim_{n\to\infty} \frac{I_n(x)}{n} = H(T,\mathcal{P}) \qquad \text{for a.e. point } x \text{ and in } L^1(\mu).$$

As a result, for any $\epsilon > 0$ there exists $n_0 > 0$ such that, for any $n \geq n_0$ there exists $S_n \subset \mathcal{P}^{\vee n}$ such that $\mu(S_n) \geq 1-\epsilon$ and, for any partition element $C \in S_n$,

$$H(T,\mathcal{P}) - \epsilon \leq \frac{-\log\mu(C)}{n} \leq H(T,\mathcal{P}) + \epsilon.$$

In case T is a continuous map on X, one can also define the KS entropy independently of any partition, using a "geometric" approach which will be useful in the case of the topological entropy. For this, we start by defining a time-dependent distance on X:

Definition 5.5. Let $d(x,y)$ be the distance on the compact metric space X. Then, for any time $n \geq 0$, we call

$$d_n(x,y) \stackrel{\text{def}}{=} \max_{0\leq j\leq n-1} d(T^j(x), T^j(y)). \tag{5.4}$$

This is also a distance on X, sometimes called the Bowen distance. The time-n *Bowen ball* of radius δ around x is defined by $B(x,\delta,n) = \{y,\ d_n(x,y) < \delta\}$: it consists of the points y whose trajectory stays δ-close to the trajectory of x at least up to time $n-1$.

Assume that the observer cannot distinguish points at distances $\leq \epsilon$. Then, two points x, y in the same (ϵ, n)-ball cannot be distinguished when performing n successive measurements. We will see below that the topological entropy is defined using these balls. It is also remarkable that the KS-entropy can also be defined in terms of these balls, through a notion of "local KS entropy".

Theorem 5.6. *[Brin-Katok] Let $T : X \circlearrowleft$ be continuous map preserving a probability measure μ. Then, for almost every x the limit*

$$\lim_{\delta\to 0} \limsup_{n\to\infty} \frac{-\log \mu\left(B(x,\delta,n)\right)}{n} \stackrel{\text{def}}{=} H_{KS}(T,\mu,x) \quad \textit{exists.}$$

Furthermore, the function $H_{KS}(T,\mu,x)$ is in $L^1(\mu)$ and is T-invariant (the $\limsup_n$ could be replaced by $\liminf_n$ a.e.).

Finally, the (global) KS entropy is given by

$$H_{KS}(T,\mu) = \int H_{KS}(T,\mu,x)\, d\mu(x).$$

5.2. **Topological entropy.** A few years after the "invention" of the KS entropy, Adler-Kronheim-McAndrew invented a similar notion in the case of a topological dynamical system (X,T), which they called the *topological entropy* of the continuous map T. Like its measure-theoretic ancestor, this entropy can be defined in several ways. We will provide the most popular definition(s) given by Bowen. It uses the refined distances d_n (5.4).

Definition 5.7. Take any $\epsilon > 0$ and $n \in \mathbb{N}$. Then a subset $A \subset X$ is said to be (n,ϵ)-*spanning* if for any $x \in X$ there exists $y \in A$ such that $d_n(x,y) \leq \epsilon$. Equivalently, the union of balls $\bigcup_{y\in A} B(y,\epsilon,n)$ covers X.

Since X is compact, there exists finite (n,ϵ)-spanning sets, but these sets still need to be ϵ-dense when ϵ is small. Call $Span(n,\epsilon)$ be the *minimal* cardinal of an (n,ϵ)-spanning set.

We also call $cov(n,\epsilon)$ the minimum cardinal of a covering of X by sets of d_n-diameter $\leq \epsilon$.

Definition 5.8. A subset $B \subset X$ is said to be (n,ϵ)-*separated* iff for any $x \neq y \in B$, one has $d_n(x,y) > \epsilon$.

Obviously, such a set must be finite. Call $Sep(n,\epsilon)$ the *maximal* cardinal of an (n,ϵ)-separated set.

Lemma 5.9. *These numbers are related by $cov(n,2\epsilon) \leq span(n,\epsilon) \leq sep(n,\epsilon) \leq cov(n,\epsilon)$.*

These quantities count the number of orbits segments of length n which a distinguishable by an ϵ-observer. The following ϵ-entropy measures the exponential growth rate of this number (w.r.to n):

$$H_\epsilon(T) \stackrel{\text{def}}{=} \limsup_{n\to\infty} \frac{-\log sep(n,\epsilon)}{n}.$$

Proposition 5.10. *In the above formula one could replace $sep(n,\epsilon)$ by $span(n,\epsilon)$ or $cov(n,\epsilon)$. In the case of $cov(n,\epsilon)$ one can replace $\limsup_n$ by $\lim_n$.*

Proof. The first statement derives from the above Lemma. The second one from the fact that $cov(n,\epsilon)$ is submultiplicative:

$$cov(n+m,\epsilon) \leq cov(n,\epsilon)\, cov(m,\epsilon).$$

□

Definition 5.11. Let $T : X \circlearrowleft$ be a continous map. Then the topological entropy of T is defined by

$$H_{top}(T) \stackrel{\text{def}}{=} \lim_{\epsilon \to 0} H_\epsilon(T).$$

The topological entropy can also be defined as the limit of the "lower" ϵ-entropy $\underline{H}_\epsilon(T) \stackrel{\text{def}}{=} \liminf_{n\to\infty} \frac{-\log sep(n,\epsilon)}{n}$, where *sep* could be replaced by *span* or *cov*.

Proposition 5.12. *The entropy $H_{top}(T)$ is independent of the distance function $d(\cdot,\cdot)$ generating the topology on X. This means that H_{top} is a topological invariant.*

The topological entropy enjoys a few common properties with its measure-theoretic counterpart:

(1) For any $k \in \mathbb{N}$, $H_{top}(T^k) = kH_{top}(T)$. If T is invertible, $H_{top}(T^{-1}) = H_{top}(T)$.
(2) Let $(A_i)_{i=1,\dots,I}$ be closed invariant subsets of X. Then
$$H_{top}(T, X) = \max_{1\le i\le I} H_{top}(T, A_i).$$
Notice the difference with (5.3).
(3) If T is a topological factor of S (that is, $T \circ \pi = \pi \circ S$ for some surjection π), then $H_{top}(T) \le H_{top}(S)$.

There exists an alternative of $H_{top}(T)$ (which is actually the historical definition) which somehow resembles the definition of the Kolmogorov-Sinai entropy. It is based on **finite open covers** of X (which take the place of finite partitions $\mathcal{P}$). Let us call $\mathfrak{U} = \{U_1, \dots, U_K\}$ an open cover. A subcover is a subset of $\mathfrak{U}$ which is still a cover. For any such cover, we call $N(\mathfrak{U})$ the cardinal of the smallest subcover of $\mathfrak{U}$.

Definition 5.13. For any open cover $\mathfrak{U}$, the entropy of this cover is defined by $H(\mathfrak{U}) \stackrel{\text{def}}{=} \log N(\mathfrak{U})$.

Any cover $\mathfrak{U}$ can be refined through the dynamics (as partitions were refined). For any time n, the refined cover $\mathfrak{U}^{\vee n}$ is made of the open sets

$$U_{\alpha_0 \cdots \alpha_{n-1}} = U_{\alpha_0} \cap T^{-1}U_{\alpha_1} \cap \cdots \cap T^{-n+1}U_{\alpha_{n-1}}, \qquad \alpha_i = 1, \dots, K.$$

The entropies $H(\mathfrak{U}^{\vee n})$ are subadditive, so the limit

$$H(T, \mathfrak{U}) = \lim_{n\to\infty} \frac{H(\mathfrak{U}^{\vee n})}{n} \qquad \text{exists.}$$

Proposition 5.14. *The topological entropy can be defined by*

$$H_{top}(T) = \sup_{\mathfrak{U}} H(T, \mathfrak{U}),$$

where $\mathfrak{U}$ ranges over all open covers of X.

Proof. If $\mathfrak{U}$ is an open cover with Lebesgue number δ[11], then $N(\mathfrak{U}^{\vee n}) \le span(n, \delta/2)$. On the other hand, if $\operatorname{diam}(\mathfrak{U}) \le \epsilon$, then $sep(n, \epsilon) \le N(\mathfrak{U}^{\vee n})$. □

Remark 5.15. If we compare this definition of H_{top} with that of the KS entropy, we already feel that the former, which only "counts" elements of the cover, will majorize the latter, which equips each element of the partition with some probability weight: indeed, the former corresponds to an equidistributed weight across the elements of the cover. This feeling will be confirmed by the variational principle, thm 5.20.

[11]This means that $\forall x \in X$, the ball $B(x, \delta)$ is fully contained in some element $U_k \in \mathfrak{U}$

We recall that a homeomorphism T is **expansive** iff there exists $\delta > 0$ such that any two different orbits $(T^n x)$, $(T^n y)$ are eventually separated by at least δ. This property is equivalent with the existence of finite open covers *generating* the topology, namely covers $\mathfrak{U}$ such that for any bi-infinite sequence $\boldsymbol{\alpha}$ the set $U_{\boldsymbol{\alpha}} = \bigcap_{i=-\infty}^{\infty} T^{-i} U_i$ contains at most one point.

Proposition 5.16. *Assume $T : X \circlearrowleft$ is an expansive homeomorphism, with expansivity constant $\delta_0 > 0$. Then,*

-if the finite open cover $\mathfrak{U}$ is a generator for T, then $H_{top}(T) = H(T, \mathfrak{U})$.

-for any $\delta < \delta_0/2$, one has$H_{top}(T) = H_\delta(T)$.

Notice the similarity between this property and Sinai's theorem 5.3 for $H_{KS}(T, \mu)$. As a consequence, any expansive homeomorphism has a finite topological entropy. Expansivity also allows us to connect the topological entropy with the counting of long periodic orbits (see (3.1)).

Proposition 5.17. *Let $T : X \circlearrowleft$ be an expansive homeomorphism. Then $H_{top}(T) \geq p(T)$.*

Proof. Let $\delta_0 > 0$ be the expansivity constant of T. Take any $0 < \epsilon < \delta_0$. For any$n \geq 1$ and any $x \neq y \in Fix(T^n)$, the trajectories $(T^j x)$, $(T^j y)$ are δ_0-separated, which means that $d_n(x, y) \geq \delta_0$. Hence, the set $Fix(T^n)$ is (n, ϵ)-separated. Hence, $|Fix(T^n)| \leq sep(n, \epsilon)$ for any n. □

5.2.1. *Topological entropy of smooth maps.* In the case of a smooth map f on a D-dimensional manifold, the points cannot separate arbitrarily fast: the rate of separation is bounded by the Lipshitz constant of f, $C_{Lip}(f) = \sup_{x \neq y} \frac{d(f(x), f(y))}{d(x,y)}$. As a result, the topological entropy is bounded by

$$H_{top}(f) \leq \max\left(0, D \log C_{Lip}(f)\right). \tag{5.5}$$

This estimate is much refined by taking into account the Lyapunov exponents of the system.

Theorem 5.18. *[Odselets] For any C^1 transformation (T, μ), there exists a.e. defined functions $k_i(x)$, $\chi_i(x)$ and subspaces $E_i(x) \subset T_x X$ of dimensions $k_i(x)$, such that $\sum_i k_i(x) = d$, $T_x X = \bigoplus_i E_i(x)$ the subspaces $E_i(x)$ form a T-invariant "folation", and*

$$\forall v \in E_i(x), \qquad \lim_{n \to \infty} \frac{1}{n} \log \|dT^n(x) v\| = \chi_i(x).$$

The $\chi_i(x)$ are called the Lyapunov exponents.

In the case of an Anosov system, we had a decomposition $T_x X = E^+(x) \oplus E^-(x)$ at each point. The connection with the Oselets decomposition is $E^+(x) = \bigoplus_{i:\chi_i(x)>0} E_i(x)$.

Theorem 5.19. *[Ruelle inequality] For any C^1 transformation (T, μ), the KS entropy satisfies the following bound:*

$$H_{KS}(T, \mu) \leq \int \sum_{\chi_i(x)>0} k_i(x) \chi_i(x)\, d\mu(x).$$

In the case of an Anosov system preserving the Lebesgue measure, this bound is reached only for $\mu = \mu_L$.

5.3. **Variational principle.** There exists a deep connection between the topological and measure-theoretic entropies, taking the form of a variational principle:

Theorem 5.20. *Let $T : X \circlearrowleft$ be a continuous map on the compact metric space X. Then*

$$H_{top}(T) = \sup\left\{H_{KS}(T, \mu),\ \mu \in \mathcal{M}(X, T)\right\}. \tag{5.6}$$

Proof. The proof of the inequality

$$H_{top}(T) \leq \sup_{\mu} H_{KS}(T,\mu) \tag{5.7}$$

proceeds by explicitly constructing measures of large entropy, based on (n,ϵ)-separated sets. Fix $\epsilon > 0$. For any $n \geq 1$ we consider a (n,ϵ)-separated set $E_n \subset X$ of maximal cardinal, and its associated uniform measure $\nu_n = \frac{1}{|E_n|}\sum_{x\in E_n}\delta_x$. We average the latter through the map:

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

(the measure μ_n is also uniform measure over $\bigcup_{j=0}^{n-1} T^j E_n$). Up to extracting a subsequence, we may assume that $\limsup_{n\to\infty}\frac{1}{n}\log|E_n| = \lim_{n\to\infty}\frac{1}{n}\log|E_n|$, and consider any accumulation point μ of the subsequence μ_n. That limit measure is automatically T-invariant. We want to prove that

$$\lim_{n\to\infty}\frac{1}{n}\log|E_n| \leq H_{KS}(T,\mu). \tag{5.8}$$

Let $\mathcal{P}$ be a partition of X of diameter $< \epsilon$, and such that $\mu(\partial\mathcal{P}) = 0$. Each element of the partition $\mathcal{P}^{\vee n}$ then satisfies $\nu_n(C) = 0$ or $\nu_n(C) = \frac{1}{|E_n|}$, which implies $H(\nu_n, \mathcal{P}^{\vee n}) = \log|E_n|$. By playing a bit with the set $\{0,\dots,n-1\}$ and using the subadditivity and the subaffineness[12] of the entropy, one gets

$$\frac{q}{n}H(\nu_n,\mathcal{P}^{\vee n}) \leq H(\mu_n,\mathcal{P}^{\vee q}) + \frac{2q^2}{n}\log|\mathcal{P}|.$$

Fix some $0 \leq k < q < n$, and us Euclidean divison by q to decompose the set $S = \{k, k+1, \dots, k+q(a(k)-1)+(q-1)\}$ into q-packets, where we used $a(k) \stackrel{\text{def}}{=} \left[\frac{n-k}{q}\right]$. Call S' the complement of S in $\{0,\dots,n-1\}$. The refinement $\mathcal{P}^{\vee n}$ can be decomposed accordingly:

$$\bigvee_{j=0}^{n-1} T^{-i}\mathcal{P} = \left(\bigvee_{r=0}^{a(k)-1} T^{-rq-k}\mathcal{P}^{\vee q}\right) \vee \bigvee_{i\in S'} T^{-i}\mathcal{P}.$$

The subadditivity of the entropy implies that

$$H(\nu_n,\mathcal{P}^{\vee n}) \leq \sum_{r=0}^{a(k)-1} H(T_*^{(rq+k)}\nu_n, \mathcal{P}^{\vee q}) + H(\nu_n, \mathcal{P}^{\vee S'}).$$

Since $|S'| \leq 2q$, the last term of the RHS is bounded by $2q\log|\mathcal{P}|$. Summing over $k = 0,\dots,q-1$, we get

$$\frac{q}{n}H(\nu_n,\mathcal{P}^{\vee n}) \leq \frac{1}{n}\sum_{k=0}^{q-1}\sum_{r=0}^{a(k)-1} H(T_*^{(rq+k)}\nu_n,\mathcal{P}^{\vee q}) + \frac{2q^2}{n}\log|\mathcal{P}|.$$

The sum on the RHS runs once through the whole set $\{0,1,\dots,n-1\}$; the subaffineness on then yields $H(\mu_n,\mathcal{P}^{\vee q})$. Because μ doesn't charge $\partial\mathcal{P}$, it does not charge either $\mathcal{P}^{\vee q}$. One has then $H(\mu_n,\mathcal{P}^{\vee q}) \stackrel{n\to\infty}{\to} H(\mu,\mathcal{P}^{\vee q})$. We thus get

$$\lim_n \frac{\log|E_n|}{n} \leq \frac{1}{q}H(\mu,\mathcal{P}^{\vee q}),$$

so we get (5.8) by taking $q\to\infty$.

To get the reverse inequality, we start from an invariant measure and a partition $\mathcal{P} = \{P_1,\dots,P_K\}$, and construct a family of (n,ϵ)-separated sets. There exists compact sets $Q_i \subset P_i$ such that $\mathcal{P}$ is close to the partition $\mathcal{Q} = \{Q_1,\dots,Q_K, Q_0 = X\setminus\cup_i Q_i\}$, so that

$$H(T,\mu,\mathcal{P}) \leq H(T,\mu,\mathcal{Q}) + 1.$$

[12]that is, $\alpha H(\nu,\mathcal{P}) + (1-\alpha)H(\mu,\mathcal{P}) \leq H(\alpha\nu + (1-\alpha)\mu, \mathcal{P})$ for any pair of probability measures and $\alpha \in [0,1]$

The family $\mathfrak{Q} = \{Q_0 \cup Q_1, \dots, Q_0 \cup Q_K\}$ is an open cover of X. On the one hand, the refinements of this cover and of the partition $\mathcal{Q}$ are related by $|\mathcal{Q}^{\vee n}| \leq 2^n |\mathfrak{Q}^{\vee n}|$, so that

$$H(\mu, \mathcal{Q}^{\vee n}) \leq \log |\mathfrak{Q}^{\vee n}| + n \log 2.$$

On the other hand, the covers $\mathfrak{Q}$ and $\mathfrak{Q}^{\vee n}$ are maximal (they don't admit strict subcovers): each element $C \in \mathfrak{Q}^{\vee n}$ contains a point x_C which is not in any other elements. If δ is the Lebesgue number of the cover $\mathfrak{Q}$[13], it is also the Lebesgue number of the cover $\mathfrak{Q}^{\vee n}$ w.r.to the distance d_n. This implies that the ball $B(x_C, \delta, n)$ is also contained in C. As a result, the discrete set $\{x_C : C \in \mathfrak{Q}^{\vee n}\}$ is (n, δ)-separated. Sending $n \to \infty$, we thus find

$$H(T, \mu, \mathcal{Q}) \leq H_{top}(T) + \log 2 \Longrightarrow H(T, \mu, \mathcal{P}) \leq H_{top}(T) + \log 2 + 1 \Longrightarrow H_{KS}(T, \mu) \leq H_{top}(T) + \log 2 + 1.$$

The same inequality holds if we replace T by any power T^k. Taking $k \to \infty$ and using $H(T^k) = kH(T)$, we get

$$H_{KS}(T, \mu) \leq H_{top}(T).$$

□

Proposition 5.21. *For any* expansive *map T, the supremum in (5.6) is reached: there exists at least one invariant measure of maximal entropy.*

Proof. For an expansive map, Prop. 5.16shows that if $\epsilon > 0$ is small enough and the sets E_n are maximal (n, ϵ)-separated sets, we have $H_{top}(T) = \limsup_{n \geq 1} \frac{\log |E_n|}{n}$. The measure μ constructed in the first part of the proof of the variational principle then satisfies $H_{KS}(\mu) = H_{top}(T)$. □

5.4. **A few examples of computing entropies.**

5.4.1. *Irrational translations.* If $\alpha = \frac{p}{q} \in \mathbb{Q}$, we know that $(R_\alpha)^q = Id$. For any invariant measure μ we have $H_{KS}(R_\alpha, \mu) = \frac{1}{q} H_{KS}(Id, \mu) = 0$. The latter property is obviously due to $\mathcal{P}^{\vee n} = \mathcal{P}$ for any finite partition $\mathcal{P}$ and any $n > 0$.

If $\alpha \notin \mathbb{Q}$, we know that the unique invariant measure of R_α is the Lebesgue measure μ_L. Call $\mathcal{P}$ the partition of S^1 into two intervals of length $\frac{1}{2}$. Then, it is easy to check that the refined partition $\mathcal{P}^{\vee n}$ is made of exactly$2n$ intervals. Hence, $H(\mathcal{P}^{\vee n}) \leq \log(2n)$, which implies that $H(R_\alpha, \mathcal{P}) = 0$. Besides, since the interval ends fill the circle densely as $n \to \infty$, the partition $\mathcal{P}$ is generating, so $H_{KS}(R_\alpha, \mu_L) = H_{top}(R_\alpha) = 0$.

Proposition 5.22. *More generally, the topological entropy of any* isometry *vanishes.*

5.4.2. *Linear dilations on S^1.* We consider the map E_m on S^1 ($m \in \mathbb{N}$, $m \geq 2$) , which preserves the Lebesgue measure. The partition $\mathcal{P}$ into intervals $[\frac{j}{m}, \frac{j+1}{m})$ is obviously generating, since $\mathcal{P}^{\vee n}$ is made of the very small intervals $[\frac{j}{m^n}, \frac{j+1}{m^n})$. We obviously have

$$H(\mathcal{P}^{\vee n}) = n \log |m|,$$

so that $H_{KS}(E_m, \mu_L) = \log m$.

The Lipschitz bound (5.5) on the topological entropy shows that we must have $H_{top}(E_m) \leq \log m$. The variational principle shows that we indeed have $H_{top}(E_m) = \log m$ (another proof of $H_{top} \geq \log m$ proceeds by noticing that On the other hand, the m^n points $\left\{\frac{j}{m^n}\right\}$ form a (n, ϵ)-separated set for any $\epsilon < \frac{1}{2m}$).

We also notice that $H_{top}(T) = p(T)$, the exponential growth rate of periodic orbits (compare with Prop.5.17).

[13]for any $x \in X$, the ball $B(x, \delta)$ is contained in some element $Q_0 \cup Q_i$ of the cover.

5.4.3. *Full shifts.* The easiest computations of the entropy are performed in the case of symbolic dynamics, namely subshifts of finite type.

Let us start from the full shift $(\Sigma_m^{(+)}, \sigma)$ perserving a (nonatomic) Bernoulli measure $\mu_{\boldsymbol{p}}$ (see (4.5)). The cylinders $\mathcal{P} \overset{\text{def}}{=} \{C_{\epsilon_0}, \epsilon_0 = 0, \dots, m-1\}$ form a generating partition. We then have

$$\begin{aligned} H(\mu_{\boldsymbol{p}}, \mathcal{P}^{\vee n}) &= -\sum_{|\boldsymbol{\epsilon}|=n} p_{\epsilon_0} \cdots p_{\epsilon_{n-1}} \left(\log p_{\epsilon_0} + \log p_{\epsilon_1} + \ldots + \log p_{\epsilon_{n-1}}\right) \\ &= -n \sum_{\epsilon_0} p_{\epsilon_0} \log p_{\epsilon_0}, \end{aligned}$$

so that

$$H_{KS}(\mu_{\boldsymbol{p}}) = -\sum_{\epsilon_0} p_{\epsilon_0} \log p_{\epsilon_0}.$$

We see that the entropy of Bernoulli measures is maximized for the uniform distribution $\boldsymbol{p}_{\max} = \left\{\frac{1}{m}, \dots, \frac{1}{m}\right\}$ (which corresponds to the Lebesgue measure of E_m through the semiconjugacy): this maximum takes the value $H_{KS}(\mu_{\boldsymbol{p}_{\max}}) = \log m$.

One can show that this value is the topological entropy of the shift $(\Sigma_m^{(+)}, \sigma)$ (see the following section).

5.4.4. *Subshifts of finite type.* In §4.5.5 we have constructed an invariant measure ν_Π associated with an irreducible stochasotic matrix Π supported by some topological chain (Σ_A, σ). Once again, the simplest generating partition consists in the cylinders $\mathcal{P} \overset{\text{def}}{=} \{C_{\epsilon_0}, \epsilon_0 = 0, \dots, m-1\}$. From the definition of ν_Π we get

$$\begin{aligned} H(\nu_\Pi, \mathcal{P}^{\vee n}) &= -\sum_{\boldsymbol{\epsilon}} p_{\epsilon_0} \Pi_{\epsilon_0\epsilon_1} \cdots \Pi_{\epsilon_{n-2}\epsilon_{n-1}} \left(\log p_{\epsilon_0} + \log \Pi_{\epsilon_0\epsilon_1} + \cdots + \log \Pi_{\epsilon_{n-2}\epsilon_{n-1}}\right) \\ &= -\sum_{\epsilon_0\epsilon_{n-1}} p_{\epsilon_0} (\Pi^{n-1})_{\epsilon_0\epsilon_{n-1}} \log p_{\epsilon_0} - \sum_{\epsilon_0\epsilon_1\epsilon_{n-1}} p_{\epsilon_0} \Pi_{\epsilon_0\epsilon_1} (\Pi^{n-2})_{\epsilon_1\epsilon_{n-1}} \log \Pi_{\epsilon_0\epsilon_1} - \cdots \\ &- \sum_{\epsilon_0\epsilon_1\epsilon_{n-1}} p_{\epsilon_0} (\Pi^{n-2})_{\epsilon_0\epsilon_{n-2}} \Pi_{\epsilon_{n-2}\epsilon_{n-1}} \log \Pi_{\epsilon_{n-2}\epsilon_{n-1}}. \end{aligned}$$

From the Perron-Frobenius theorem we already know that $\left(\Pi^N\right)_{ij} = p_j + \mathcal{O}(\mu^N)$ for some $0 < \mu < 1$. As a result, the above terms can be simplified into

$$\begin{aligned} H(\nu_\Pi, \mathcal{P}^{\vee n}) &= -\sum_{\epsilon_0} p_{\epsilon_0} \log p_{\epsilon_0} - \sum_{\epsilon_0\epsilon_1} p_{\epsilon_0} \Pi_{\epsilon_0\epsilon_1} \log \Pi_{\epsilon_0\epsilon_1} - \ldots - \\ &- \sum_{\epsilon_0\epsilon_1\epsilon_{n-1}} p_{\epsilon_0} p_{\epsilon_{n-2}} \Pi_{\epsilon_{n-2}\epsilon_{n-1}} \log \Pi_{\epsilon_{n-2}\epsilon_{n-1}} + \mathcal{O}(n\mu^{n/2}) \\ &= -\sum_{\epsilon_0} p_{\epsilon_0} \log p_{\epsilon_0} - (n-1) \sum_{\epsilon_0\epsilon_1} p_{\epsilon_0} \Pi_{\epsilon_0\epsilon_1} \log \Pi_{\epsilon_0\epsilon_1} + \mathcal{O}(n\mu^{n/2}). \end{aligned}$$

We thus have shown that

$$H_{KS}(\nu_\Pi) = -\sum_{\epsilon_0\epsilon_1} p_{\epsilon_0} \Pi_{\epsilon_0\epsilon_1} \log \Pi_{\epsilon_0\epsilon_1}.$$

ν_Π is only a particular invariant measure of the subshift (Σ_A, σ). Among these measures which ones do maximize the entropy? Equivalently, which Markov matrix Π supported on the adjacency matrix A induces a maximal complexity?

The answer is relatively simple, it is provided by the Shannon-Parry measure. We have assumed that A is a primitive adjacency matrix. Let λ be its maximal (PF) eigenvalue, and $\boldsymbol{q}, \mathbf{v}$ be respectively left and right eigenvectors associated with λ, normalized such that $\langle \boldsymbol{q}, \mathbf{v} \rangle = 1$. Take V to be the diagonal matrix with $V_{ii} = v_i$. Then, define the matrix

$$\Pi_{SP} \overset{\text{def}}{=} \lambda^{-1} V^{-1} A V, \qquad (\Pi_{SP})_{ij} = \frac{A_{ij} v_j}{\lambda v_i}$$

One easily checks that it is stochastic, and is supported by A (without accidental zeros). The vector $\boldsymbol{p} = \boldsymbol{q}V = (q_1 v_1, \dots q_n v_n)$ is the positive left eigenvector of Π_{Per}. The Markov measure $\mu_{\Pi_{SP}}$ is called the Shannon-Perry measure for the topological Markov chain $(\Sigma_A^{(+)}, \sigma)$. Through a direct computation one finds that it has the entropy

$$H_{KS}(\mu_{\Pi_{SP}}, \sigma) = \log \lambda.$$

On the other hand, this value is also the topological entropy of (Σ_A, σ).

Proposition 5.23. *For any subshift $X \subset \Sigma_m$, the topological entropy is given by the asymptotics of number of n-words in X:*

$$H_{top}(X, \sigma) = \lim_{n\to\infty} \frac{1}{n} \log |W_n(X)|,$$

where $W_n(X)$ is the set of n-words, that is of nonempty n-cylinders $C_{\alpha_0 \cdots \alpha_{n-1}}$.

Proof. The shift is an expansive homeomorphism, and the open cover $\mathfrak{U} = \{C_i,\ i = 0, \dots, m-1\}$ is generating. The refined cover $\mathfrak{U}^{\vee n}$ is made of all cylinders of length n (including the empty ones). Hence, the smallest subcover is provided by the set of nonempty n-cylinders $W_n(X)$. □

In the case of the topological Markov chain, we have $|W_n(\Sigma_A)| = \sum_{\epsilon_0 \epsilon_{n-1}} (A^{n-1})_{\epsilon_0 \epsilon_{n-1}} \sim \lambda^{n-1} q_{\epsilon_0} v_{\epsilon_{n-1}}$. We notice that, not only does $H_{top}(\Sigma_A)$ majorize the periodic orbit counting rate (as explained in Prop. 5.17), but it is actually equal to it:

$$H_{top}(\Sigma_A) = p(\Sigma_A).$$

5.4.5. *Hyperbolic automorphisms on $\mathbb{T}^d$.* In the simplest case of $\mathbb{T}^2$, we have called $\lambda > 1$ the large eigenvalue of the hyperbolic matrix M.

Proposition 5.24. *The topological entropy $H_{top}(M) = \log \lambda$. It is also equal to $H_{KS}(\mu_L)$.*

Proof. One uses a distance adapted to the stable/unstable directions, namely $d(0, x) = \max(x_s, x_u)$. In that case, ϵ-balls are parallelograms ("rectangles") aligned with the axes, with sides of length 2ϵ. A refined ball $B(x, \epsilon, n)$ will have a length $2\epsilon\lambda^{-n}$ along the unstable direction, and 2ϵ along the stable direction. As a result, such balls have volume $C\epsilon^2\lambda^{-n}$, so one needs $\geq \lambda^n (C\epsilon^2)^{-1}$ such balls to cover $\mathbb{T}^2$. On the other hand, one can indeed cover $\mathbb{T}^2$ using $\sim \lambda^n (c\epsilon^2)^{-1}$ such balls, so $H_{top}(M) = \log \lambda$.

If we start from a partition made of such adapted rectangles, the Lebesgue measure of all refined rectangles decay like $\leq C\lambda^{-n}$. This shows that $H_{KS}(\mu_L) \geq \log \lambda$. The reverse inequality is due to the variational principle.

Once more, we notice that $H_{top}(M) = p(M)$. □

This result can be extended to higher-dimensional tori.

Theorem 5.25. *For any $d \geq 2$, let $M \in GL(d, \mathbb{Z})$ be a hyperbolic matrix, which is identified with the corresponding invertible map it induces on $\mathbb{T}^d$. Then its topological entropy is given by*

$$H_{top}(M) = \sum_{|\lambda_i|>1} \log |\lambda_i|,$$

where the eigenvalues $\{\lambda_1, \dots, \lambda_d\}$ are counted with multiplicity. This entropy is also equal to $H_{KS}(M, \mu_L)$.

6. Hyperbolic dynamical systems

6.1. **Hyperbolic set.** Let X be a smooth Riemannian manifold, and $U \subset X$ a nonempty open subset. From now on $f : U \to X$ will always be a C^1 diffeomorphism on its image $f(U) \subset X$. An invariant closed set $\Lambda \subset U$ is said to be a hyperbolic set iff there exists $C > 0$, $\lambda \in (0,1)$ and families of subspaces $E^\pm_x \subset T_xX$ such that, for any $x \in \Lambda$,

(1) $T_xX = E^-_x \oplus E^+_x$ (in particular, $E^-_x \cap E^+_x = 0$),
(2) $\|df^{\pm n}_x v\| \leq C\lambda^n \|v\|$ for any $v \in E^\mp_x$ and $n \geq 0$,
(3) $df_x E^\pm_x = E^\pm_{f(x)}$.

The subspace E^-_x (resp. E^+_x) is called the stable (resp. unstable) subspace at the point x. The distributions $\{E^\pm_x,\ x \in \Lambda\}$ are called the (un)stable distributions of $f_{\restriction\Lambda}$.

Proposition 6.1. *Being a hyperbolic set does not depend on the particular metric on X.*

The simplest form of hyperbolic set is made of a single hyperbolic point $\Lambda = \{x\}$ (e.g. in §2.2). In that case, the subspaces $E^\pm_x$ have a spectral interpretation (they are sums of generalized eigenspaces of the map $df_x : T_xX \circlearrowleft$).

The subsets $E^\pm_x$ are defined by the behaviour of df^n_x in the limits $n \to \pm\infty$. They can be obtained by a limit process, so there is a priori no reason for these spaces to depend smoothly on x. Yet, one can easily prove the following

Proposition 6.2. *Let Λ be a hyperbolic set of f. Then the subspaces $E^\pm_x$ depend continuously on $x \in \Lambda$. This implies that $\dim E^\pm_x$ is locally constant, and that the subspaces $E^\pm_x$ are uniformly transverse on Λ (the "angle" between $E^\pm_x$ is uniformly bounded from below).*

6.1.1. *Stable and unstable manifolds.* From the above definition, a hyperbolic point x is characterized by the (un)stable subspaces $E^\pm_x$ (and their push-forwards through f or f^{-1}). These subspaces describe the linearized, or infinitesimal dynamics near the orbit of x. The *Hadamard-Perron theorem* shows that there exist "nonlinear, macroscopic extensions" of these subspaces, in the form of **stable and unstable manifolds**. These manifolds can be defined by first defining *locally:* they are (small) embedded disks $x \in W^\pm_{x,loc} \subset X$ satisfying the following properties:

(1) $f(W^-_{x,loc}) \subset W^-_{f(x),loc}$, $f^{-1}(W^+_{x,loc}) \subset W^+_{f^{-1}(x),loc}$.
(2) $T_xW^\pm_{x,loc} = E^\pm_x$.

Any point in either of these manifolds will approach the trajectory of x exponentially fast, respectively in the past or future time directions:

$$\forall y \in W^\pm_{x,loc},\ \forall n \geq 0, \qquad d(f^{\mp n}(y), f^{\mp n}(x)) \leq C\lambda^n d(y,x).$$

Proof. The proof of the *Hadamard-Perron theorem* is easier to formulate when x is a hyperbolic fixed point. Let us then sketch the construction of $W^+_{x,loc}$. One considers, in a local coordinate chart near x, the family of Lipschitz functions $\phi : E^+_x \to E^-_x$, such that $\phi(0) = 0$ and with a uniform bound $Lip(\phi) \leq L$ (the functions need to be defined only in some ϵ-ball of the origin in E^+_x). The graph of such a ϕ is a submanifold of X, of dimension $\dim E^+_x$, which is "close" to the unstable subspace E^+_x. One defines a *graph transform* $\phi \mapsto \mathcal{F}(\phi)$ on this family of functions, by $\mathrm{graph}\mathcal{F}(\phi) \stackrel{\text{def}}{=} f(\mathrm{graph}\phi)$. For ϵ small enough, this transform is well-defined, and it is *strictly contracting* (w.r.to $d(\phi,\psi) = \sup_{x \in B_\epsilon} |\phi(x) - \psi(x)|$). From the contracting mapping principle, its iterates converge to a unique fixed point ϕ_0. The graph of this function provides $W^+_{x,loc}$. One can show that ϕ_0 is as not only Lipschitz, but is as smooth as the map f, and that $d\phi_0(0) = 0$.

The proof in the case of an arbitrary point $x \in \Lambda$ is a little more cumbersome, but more or less goes along the same lines (technically one brings back all points $f^n(x)$ to the origin, mapping f to a family of local transformation f_n preserving the origin). It uses the fact that the $E^{\pm}_x$ are uniformly transverse, and the tangent maps df_x are equicontinuous. □

The local (un)stable manifolds can then be extended through the dynamics to form the full (un)stable manifolds:

$$W^{\pm}_x = \bigcup_{n\geq 0} f^{\pm n}\left(W^{\pm}_{f^{\mp n}(x),loc}\right). \tag{6.1}$$

These global manifolds can also be defined topologically:

$$W^{\pm}_x = \left\{ y \in X,\ d(f^n(x), f^n(y)) \overset{n\to\mp\infty}{\to} 0 \right\}.$$

In general each manifold $W^{\pm}_x$ is immersed in X in a complicated way.

Definition 6.3. If $\Lambda = X$ (that is, the full phase space is a hyperbolic set), the map f is said to be an Anosov diffeomorphism (an example is given by the hyperbolic toral automorphisms of §2.6). In the 2-dimensional case, the local (un)stable manifolds are segments (say, of length ϵ) issued from x along the (un)stable directions. The full (un)stable manifolds $W^{\pm}_x$ are the full straight lines issued from x, of slopes given by the (un)stable eigenvectors. These immersed lines form a dense set in $\mathbb{T}^2$.

Remark 6.4. Anosov diffeomorphisms represent a rather "rare" type of hyperbolic set. Indeed, a smooth manifold X carrying an Anosov diffeomorphism must be above to carry two nontrivial continuous distributions $\{E^{\pm}_x,\ x \in X\}$, which is a strong topological constraint. For instance, it is notoriously impossible to construct a nonvanishing 1-dimensional distribution $\{V_x,\ x \in X\}$ on the 2-sphere. Actually the set of known Anosov diffeomorphisms (up to topological conjugacy) is reduced to the hyperbolic toral automorphisms and some automorphisms on some nilmanifolds.

Example 6.5. A more "typical" hyperbolic set is Smale's horseshoe Λ described in §2.8. Such a set doesn't fill up the whole manifold, but is a "fractal" subset of it. The local (un)stable manifolds $W^{\pm}_{x,\epsilon}$ are horizontal and vertical segments. The full unstable manifold W^{+}_x can be obtained from eq.(6.1), since $f(D) \subset D$. It will be an immersed smooth line. On the other hand, f^{-1} is only defined on $f(D)$, so the formula (6.1) should be understood by restricting each $W^{\pm}_{f^n(x),\epsilon}$ on the domain of definition of f^{-n}. The final manifold W^{-}_x is a countable union of vertical segments.

Remark 6.6. The above definition of a hyperbolic set assumes that one already knows the existence of (un)stable subspaces with the given properties. For a given dynamical system, these subspaces must be constructed by a limiting procedure. It is thus desirable to have a less "precise" definition for a hyperbolic set, which is easier to check. This definition is given in terms of **invariant cone fields**, which we now define.

Definition 6.7. A (closed) cone $C \subset \mathbb{R}^d$ is a closed subset invariant by dilation: $\xi \in C \Longrightarrow t\xi \in C$, $\forall t \in \mathbb{R}$. A cone field on a manifold X (or on some subset $\Lambda \subset X$) is a continuous family of cones $\mathcal{C} = \{C_x \subset T_x X,\ x \in \Lambda\}$. Assuming $\Lambda \subset X$ is invariant w.r.to a continuous map f, a cone field $\mathcal{C}$ on Λ is said to be invariant through f iff

$$\forall x \in \Lambda, \quad df_x(C_x) \subset C_{f(x)},$$

and strictly invariant iff

$$\forall x \in \Lambda, \quad df_x(C_x) \subset \mathrm{int} C_{f(x)} \cup \{0\}.$$

Assuming we already know the (un)stable distribution $E^{\pm}_x$ on Λ, two families of cone fields are relevant. They consist of cones aligned "close to" the unstable, resp. stable directions. They are often called, respectively

"horizontal" and "vertical" cone fields. They can be defined in terms of some parameter $\gamma > 0$:

$$\forall x \in \Lambda, \qquad C_x^+ \stackrel{\text{def}}{=} \left\{\xi^+ + \xi^-, \quad \xi^\pm \in E_x^\pm,\ \left\|\xi^-\right\| \leq \gamma \left\|\xi^+\right\|\right\},$$
$$C_x^- \stackrel{\text{def}}{=} \left\{\xi^+ + \xi^-, \quad \xi^\pm \in E_x^\pm,\ \left\|\xi^+\right\| \leq \gamma \left\|\xi^-\right\|\right\}.$$

Lemma 6.8. *If γ is small enough, one can show that $\mathcal{C}^+$ is strictly invariant through f, and $\mathcal{C}_x^-$ is strictly invariant through f^{-1}.*

From the knowledge of the distributions $E_x^\pm$, we were able to construct strictly invariant cone fields. On the opposite, the presence of a hyperbolic set can be characterized by the presence of horizontal/vertical cone fields defined in terms of an "approximate" decomposition of T_xX.

Theorem 6.9. *A compact f-invariant set is hyperbolic if there exists $\lambda \in (0,1)$ such that, for any $x \in \Lambda$, there exists a decomposition $T_xX = F_x^+ \oplus F_x^-$ (in general not invariant), a family of horizontal cones $C_x^+ \supset F_x^+$, a family of vertical cones $C_x^- \supset F_x^-$ associated with this decomposition, such that $\mathcal{C}^+$ is strictly invariant through f, $\mathcal{C}^-$ is strictly invariant through f^{-1}, and*

$$\|df_x\xi\| \geq \lambda^{-1}\|\xi\|, \quad \forall \xi \in C_x^+,$$
$$\left\|df_x^{-1}\xi\right\| \geq \lambda^{-1}\|\xi\|, \quad \forall \xi \in C_{f(x)}^-.$$

Proof. Like in the construction of (un)stable manifolds, this theorem also proceeds by an iterative contruction of the (un)stable distributions $E_x^\pm$, through a contracting mapping principle. Once more, the proof is easier to formulate if we assume that x is a fixed point. The map $df_x : T_xX \circlearrowleft$ leaves invariant the horizontal cone $C_x^+ = \{\xi^+ + \xi^-, \quad \xi^\pm \in F_x^\pm,\ \|\xi^-\| \leq \gamma\|\xi^+\|\}$, so it can be iterated. The family of cones $(df_x)^n(C_x^+)$ is strictly decreasing (in the sense of the inclusion). One can then define $E_x^+ \stackrel{\text{def}}{=} \lim_{n\to\infty}(df_x)^n(C_x^+)$, which is obviously invariant through df_x. A priori, it is a (closed) cone in T_xX. There remains to show that this cone is actually a subspace of dimension $\dim F_x^+$. This can be done by showing (using the hyperbolicity assumptions) that E_x^+ is also given by the limit of $(df_x)^n(F_x^+) \subset (df_x)^n(C_x^+)$. □

6.1.2. *Closing and Shadowing Lemmas.*

Definition 6.10. An ϵ-periodic $(x_i)_{i\in\mathbb{Z}}$ is a sequence such that $\forall i \in \mathbb{Z}$, $d(f(x_i), x_{i+1}) \leq \epsilon$. A periodic ϵ-orbit is an ϵ-orbit such that $x_{i+n} = x_i$.

Theorem 6.11. *[Anosov Closing Lemma] Let Λ be a hyperbolic set for $f : U \to X$. Then there exists a neighbourhood $V \supset \Lambda$ and $C, \epsilon_0 > 0$ such that, for any $\epsilon \leq \epsilon_0$ and any ϵ-periodic orbit $(x_0, \ldots, x_{n-1}) \subset V$, then there is a n-periodic point $y \in U$ such that $d(f^i(y), x_i) \leq C\epsilon$ for all i.*

In other words, hyperbolicity implies that any approximate periodic orbit close to Λ can be "corrected" to make it a true periodic one. One says that the approximate orbit (x_i) is *$C\epsilon$-shadowed* by the true orbit $(f^i(y))$.

This result can be proved also for non-periodic orbits:

Theorem 6.12. *[Shadowing Lemma] Let Λ be a hyperbolic set for $f : U \to X$. Then there exists a neighbourhood $V \supset \Lambda$ such that, for any $\delta > 0$, there exist $\epsilon > 0$ such that any ϵ-orbit $(x_i) \subset V$ is δ-shadowed by an orbit in U.*

Proof. Both the Closing Lemma and the Shadowing Lemma can be proven similarly, using the Contracting mappin principle. The sketch is easier in the case of the closing Lemma, so we present it. We define the map

$$F : (z_0, z_1, \ldots, z_{n-1}) \mapsto (f(z_{n-1}), f(z_0), \ldots, f(z_{n-2}))$$

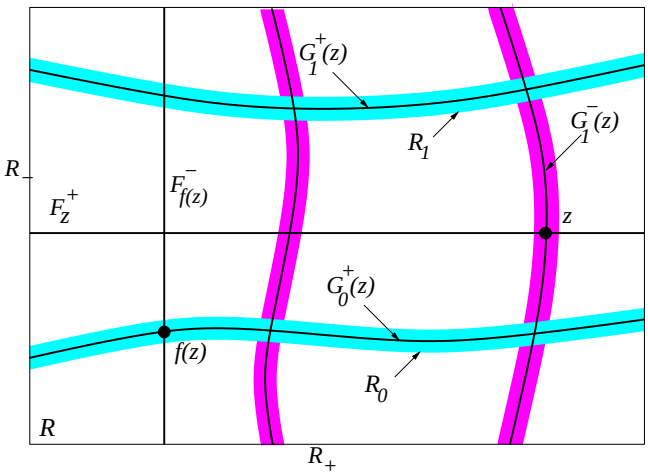

FIGURE 6.1. A rectangle containing a (nonlinear) horseshoe.

in some neighbourhood of the pseudo-orbit $(x_0, \dots, x_{n-1})$. The objective is to find a fixed point of this map. Using local adapted coordinates centered on x_i, the map $f(z_i)$ can be represented as the sum of a linear map $L_i = df_{x_i}$, and a "small" nonlinear correction N_i, acting on $\delta z_i = z_i - x_i$:

$$f(z_i) - x_{i+1} = L_i \delta z_i + N_i(\delta z_i).$$

If $\delta z_i = \mathcal{O}(\epsilon)$, then the correction $\|N_i(\delta z_i)\| = \mathcal{O}(\epsilon)$, but its differential is also small:

$$\|N_i(\delta z_i) - N_i(\delta z_i')\| \leq C\epsilon \|\delta z_i - \delta z_i'\|.$$

This sequence of maps yields a global decomposition $F = L + N$. The linear map L is hyperbolic (because all L_i are so), so that $(Id - L)$ is invertible, with a uniformly bounded inverse. A solution $F(z) = z$ also satisfies $(Id - L)^{-1} \circ N(z) = z$. Now, if ϵ is small enough, the (nonlinear) map $(Id - L)^{-1} \circ N$ is contracting, and admits a unique fixed point (y_i). One can finally easily check that $\|y_i - x_i\| = \mathcal{O}(\epsilon)$, with a constant independent of n. □

6.2. **Horseshoes and transverse homoclinic points.** In §2.8 we have constructed a "linear" horseshoe, namely one such that all unstable subspaces are horizontal, an all stable ones a vertical. One can then directly prove that the invariant set Λ for this horseshoe is a hyperbolic set.

We now give the definition of a nonlinear horseshoes. Take U an open subset of $\mathbb{R}^{n_+ + n_-}$, and a diffeomorphism $f : U \to \mathbb{R}^{n_+ + n_-}$. A set $R \subset U$ of the form $R = R_+ \times R_-$, where $R_\pm$ are disks in $\mathbb{R}^{n_\pm}$, is called a rectangle. We call $F_x^+ = R_+ \times \{x_-\}$ the horizontal fiber at x, and $F_x^- = \{x_+\} \times R_-$ the vertical fiber.

The rectangle R contains a horseshoe for f if

(1) $f(R) \cap R$ contains at least two connected components $R_0, R_1, \dots, R_{m-1}$
(2) if $z \in R$ and $f(z) \in R_i$, then the sets $G_i^+(z) \stackrel{\text{def}}{=} f(F_z^+) \cap R_i$ and $G_i^-(z) \stackrel{\text{def}}{=} f^{-1}(F_{f(z)}^- \cap R_i)$ are connected, and the restriction of π_+ to $G_i^+(z)$ (resp. of $\pi_-$to $G_i^-(z)$) are bijective.
(3) there are $\alpha, \lambda \in (0,1)$ such that, if $z, f(z) \in R$ then df_z preserves the horizontal α-cone C^+, $df_{f(z)}^{-1}$ preserves the vertical α-cone C^-, and

$$\|df_x \xi\| \geq \lambda^{-1} \|\xi\|, \quad \forall \xi \in C_x^+,$$
$$\left\|df_{f(x)}^{-1} \xi\right\| \geq \lambda^{-1} \|\xi\|, \quad \forall \xi \in C_{f(x)}^-.$$

The intersection $\Lambda^f = \bigcap_{n\in\mathbb{Z}} f^n(R)$ (obviously a closed invariant set) is called a horseshoe.

Theorem 6.13. *The horseshoe Λ^f is a hyperbolic set. If $f(R)\cap R$ has m components, then $f_{\restriction\Lambda^f}$ is topologically conjugate to the full two-side shift Σ_m on m symbols.*

Corollary 6.14. *A diffeomorphism f containing a m-horseshoe has topological entropy $H_{top}(f) \geq H_{top}(f_{\restriction\Lambda^f}) = \log m$.*

Proposition 6.15. *The definition of a horseshoe shows that, if $g : U \to X$ is a small C^1-perturbation of f, then the intersection $\Lambda^g = \bigcap_{n\in\mathbb{Z}} g^n(R)$ is also a hyperbolic set, and $g_{\restriction\Lambda^g}$ is topologically conjugated with (Σ_m, σ), and hence with $f_{\restriction\Lambda^f}$. This shows that horseshoes are (locally) structurally stable.*

A horseshoe is not only an ad hoc construction. Its structure appears very often, for instance in the vicinity of a **transverse homoclinic intersection**.

Definition 6.16. Let x_0 be a hyperbolic T-periodic point. A point $y \in U$ is called homoclinic for x_0 if $y \neq x_0$ and $y \in W^+_{x_0} \cap W^-_{x_0}$. It is called transversely homoclinic if $W^+_{x_0}$ and $W^-_{x_0}$ intersect transversely at y.

A homoclinic point converges to the periodic orbit $\mathcal{O}(x_0)$ both in the past and in the future time directions. Homoclinic points appear very naturally as soon as the (un)stable manifolds of x_0 "spread out" across the compact phase space. Homoclinic points somehow correspond to the "next order of complexity" after periodic points. Notice that a hyperbolic point is *not recurrent*. As we will see shortly, they still belong to the nonwandering set, because there exist periodic points arbitrarily close to a homoclinic point. Among all homoclinic points, the *transverse* ones are "generic": a nontransverse intersection will become transverse through a small perturbation of the dynamics.

The main interest of homoclinic points lies in the following

Theorem 6.17. *Let x_0 be a hyperbolic periodic point of a diffeomorphism $f : U \to X$, and let y be a transverse homoclinic point of x_0. Then, for any $\epsilon > 0$, the union of the ϵ-neighbourhoods of the orbits $\mathcal{O}(x_0)$ and $\mathcal{O}(y)$ contains a horseshoe of f.*

As a consequence, transverse homoclinic points represent a **source of complexity**.

Proof. For simplicity let us assume that x_0 is a fixed point, and consider local coordinates such that $W^{\pm}_{x_0,loc}$ span the horizontal, resp. vertical coordinate axes. Let V be a small neighbhourhood of x_0, so that $df_x \approx df_{x_0}$ for $x \in V$. The points $f^n(y)$ converge exponentially fast towards $\mathcal{O}(x_0)$ in the two limits $n \to \pm\infty$. Call $W^{\pm}_{x_0,y}$ small disks of $W^{\pm}_{x_0}$ containing y. Because the $W^+_{x_0,y}$ is transverse to $W^-_{x_0}$, the *inclination lemma* implies that when $n \gg 1$, the image $f^n(W^{\pm}_{x_0,y})$ contains a disk $W^+_{x_0,f^n(y)}$ which is very close to the horizontal axis (in particular, its tangent planes lie in some horizontal cone field $\mathcal{C}^+$), and stretches across V. Similarly, for $n \gg 1$ the backwards images $f^{-n}(W^-_{x_0,y})$ contains a disk $W^-_{x_0,f^{-n}(y)}$ which is very close to the vertical axis and stretches across V.

Let us consider a "vertical rectangle" $R \subset V$ containing x_0. It also contains some positive iterate $f^n(y)$. Its "short horizontal sides" are small enough, so that for some $k > 0$ the image $f^k(R)$ is still in V, but is "horizontal" (close to $W^+_{x_0,loc}$) and contains some negative iterate $f^{-m}(y)$, see Fig. 6.2. In particular, R contains $f^{-m-k}(y)$. Then, for $N = k + n + m$ the image $f^N(R)$ will be a very elongated "rectangle" aligned along a part of $W^+_{x_0}$ containing x_0 and the iterates $f^j(y)$ for $j \in (-\infty, n]$; in particular, it contains the disks $W^+_{x_0,loc}$ and $W^+_{x_0,f^n(y)}$. Like these two disks, $f^N(R) \cap R$ will stretch horizontally "across R". This shows that $f^N(R) \cap R$ contains at least two connected components, and that the image of a horizontal fiber $F^+_z \subset R$ through f^N is a union of

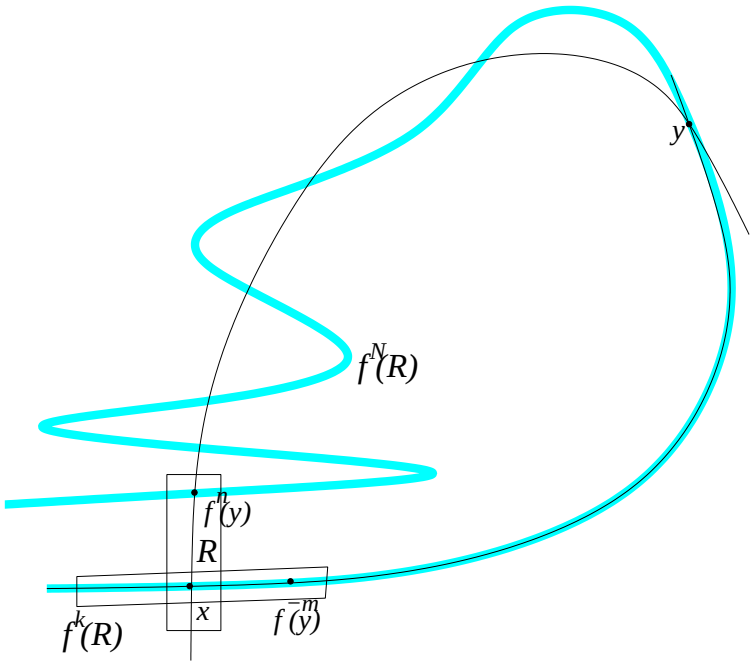

FIGURE 6.2. Construction of a horseshoe near a homoclinic point.

connected components projecting well on the horizontal axis. Similarly, a vertical fiber in $f^k(R)$ will be mapped by f^{-N} onto a union of "close to vertical" connected components in $f^k(R)$, which project well on the vertical axis. Finally, the preservation of a family of horizontal/vertical cones and the accompanying stretching property is ensured by the local hyperbolic dynamics in V, and the fact that most of the trajectory from R to $f^N(R)$ is done inside V. □

The theorem can be extended to the case of *heteroclinic points* connecting two periodic orbits $\mathcal{O}(x_0)$, $\mathcal{O}(x_1)$, namely $y \in W^-_{x_0} \cap W^+_{x_1}$.

6.3. **Locally maximal hyperbolic sets.** The following theorem is a benchmark of the (seemingly paradoxical) duality between hyperbolicity (that is, *instability* of individual trajectories) and *stability* of the hyperbolic structure itself w.r.to perturbations of the dynamics.

Theorem 6.18. *Let Λ be a hyperbolic set for $f : U \to X$. Then, there exists an open neighbourhood $V \supset \Lambda$ such that, for any map g sufficiently C^1-close to f, the set*

$$\Lambda_V^g \stackrel{\text{def}}{=} \bigcap_{n\in\mathbb{Z}} g^n \bar{V} \qquad \textit{is a hyperbolic set for } g.$$

The proof of the above theorem also uses invariant cone fields.

Corollary 6.19. *Any C^1-perturbation of an Anosov diffeomorphism is still an Anosov diffeomorphism.*

Any closed invariant subset of a hyperbolic set is automatically a hyperbolic set. Theorem 6.18 (applied to the case $g = f$) shows that a hyperbolic set can be "locally extended", since $\Lambda_V^f \supset \Lambda$. It allows us to select "nice" hyperbolic sets according to their "stability" with respect to local extensions:

Definition 6.20. Let Λ be a hyperbolic set for f. If there exists an open neighbourhood $V \supset \Lambda$ such that $\Lambda = \Lambda_V^f$, then Λ is said to be *locally maximal*.

Theorem 6.21. *The horseshoe Λ is locally maximal. For any closed invariant subset $S \subset \Lambda$ and any open neighbourhood $V \supset S$, there exists a locally maximal hyperbolic set $\tilde{S}$ such that $S \subset \tilde{S} \subset V$.*

Proof. We will only give a hint at the structure of the set $\tilde{S}$ from a simple example, which consists in a symbolic-dynamics interpretation of Thm. 6.17. Consider the fixed point $x = \bar{0}$ in the shift (Σ_2, σ), and the homoclinic point $y = \bar{0} \cdot 1\bar{0}$. The union $S = \{x_0\} \cup \mathcal{O}(y)$ is a closed invariant set. Yet, it is not locally maximal: any neighbourhood V of S contains, for N large enough, the union of cylinders $\tilde{V}_N = C_{0_N \cdot 0_N} \cup C_{10_{N-1} \cdot 0_N} \cup C_{010_{N-2} \cdot 0_N} \cup \ldots \cup C_{0_N \cdot 0_{N-1} 1}$. The largest invariant set contained in $\tilde{V}_N$, that is $\Lambda^f_{\tilde{V}}$, is the set Λ_{2N-1} made of bi-infinite sequences such that each 1 is separated by at least $2N-1$ zeroes. □

Locally maximal hyperbolic sets have a nice geometric characterization. Let show how (un)stable manifolds of a hyperbolic set allow to construct "local adapted coordinate frames" through a *local product structure*. For two points $x, y \in \Lambda$ at a small distance, the uniform transversality of $E^{\pm}_x$ and $E^{\pm}_y$ and the continuity of the distributions imply that the local manifolds $W^-_{x,loc}$ and $W^+_{y,loc}$ must intersect at a *single* point, which is usually denoted by $[x, y]$, and the intersection is transverse. This point is homoclinic to the trajectories $\mathcal{O}(x)$ and $\mathcal{O}(y)$. The hyperbolic set Λ is said to have a local product structure if there is $\delta > 0$ such that for any $x, y \in \Lambda$, the intersection $W^-_{x,loc} \cap W^+_{y,loc}$ contains at most one point, which belongs to Λ, and if $d(x, y) \leq \delta$ it indeed contains a single point $[x, y]$.

Hence, for every $x \in \Lambda$, there exists a neighbourhood $U(x)$ such that $U(x) \cap \Lambda$ is a "rectangle" built around the "axes" $W^{\pm}_{x,loc}$:

$$U(x) \cap \Lambda = \left\{ [y, z],\ y \in W^+_{x,loc},\ z \in W^-_{x,loc} \right\}.$$

Proposition 6.22. *A hyperbolic set Λ is locally maximal iff it has a local product structure.*

This product structure allows one to prove several interesting results on the global properties of the dynamics on Λ.

Corollary 6.23. *Let Λ be a locally maximal hyperbolic set for $f : U \to X$. Then the periodic points are dense in the nonwandering set $NW(f_{\restriction \Lambda})$.*

References

[1] Michael Brin et Garrett Stuck, Introduction to Dynamical Systems, Cam bridge University Press, 2002

[2] Anatole Katok et Boris Hasselblatt, Introduction to the modern theory of dynamical systems, Cambridge University Press, 1995

[3] Peter Walters, An introduction to ergodic theory, Springer, 1982

E-mail address: snonnenmacher@cea.fr