NOTES OF THE COURSE ON CHAOTIC DYNAMICAL SYSTEMS (PRELIMINARY VERSION)

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. A tentative plan (not necessarily chronological) follows.

(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.
(9) Structural stability vs. bifurcations. Examples: logistic map on the interval/nonlinear perturbation of the "cat map".

Date: 05 october 2009.

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 It is said to be *invertible* if f is a bijection on X (or at least on some subset. It then has 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 many periodic points $x \in X$ ($f^T x = x$ for some $T > 0$). Where are they? More complicated forms of *recurrence*.
(2) invariant sets. $X' \subset X$ is (forward-)invariant iff $f(X') \subset X'$.
(3) is there an invariant (probability) measure? ($\mu(A) = \mu(f^{-1}(A))$ for "any" set A). *Statistical* properties w.r.to this measure?
(4) *structural stability* of a DS: do "small perturbations" of f have the same behaviour as f? Are they conjugate with f?

Like always in maths, one would like to *classify* all possible behaviours, that is group the maps f into to some natural equivalence relation.

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 well-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, with an associated topology (family of open/closed sets). It is then natural to consider maps f which are *continuous* on X. **Topological dynamics** ($\supset$symbolic dynamics). We will often 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 f can be approximated by the linear map $df(x)$ sending the tangent space $T_x X$ to $T_{f(x)} X$. **Smooth dynamics**. A differentiable flow is generated by a vector field
$$v(x) = \frac{d}{dt} f^t(x)|_{t=0} \in T_x X,$$
Generally one starts from the field $v(x)$, the flow $f^t(x)$ being obtained by integrating over the field $v(x)$: one notes formally $f^t(x) = e^{tv}(x)$. 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. **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. **Hamiltonian/Lagrangian dynamics**. A complex structure is preserved if f is holomorphic. **Complex dynamics**.

[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 $\mu(X) = 1$.

These extra structures may also be imposed to the (semi-)conjugacy between two DS. This question is less obvious than it first appears: we will see that requiring smooth conjugacy between two smooth DS is "too strong a" condition, as opposed to the relevant 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 $\|\delta f\|_{C^1} \le \epsilon$, then f and $\tilde{f}$ are topologically conjugate (i.e., there exists a homeomorphism $h : X \to X$ such that $\tilde{f} = h^{-1} \circ f \circ h$). (It is called *strongly* structurally stable if one can choose h close to the identity).

1.1. **From maps to flows, and back.** So far we have only considered examples given by discrete-time maps $f : X \to X$. From such a map one can easily construct a flow through a suspension procedure. Select a positive function $\tau : X \to \mathbb{R}^+$, called the ceiling function, or first return time. Then, consider the product space

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

One can easily produce a flow 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. Many dynamical properties of the map f are inherited by the flow 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 (1-2) dimensions. They will already show a large variety of behaviours.

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

$$\forall x, y, \quad dist(f(x), f(y)) \le \lambda\, dist(x, y) \Longrightarrow dist(f^n(x), f^n(y)) \overset{n\to\infty}{\to} 0.$$

Contraction mapping principle: (provided X is complete) this implies that f admits a unique fixed point $x_0 \in X$, which is an **attractor** (at exponential speed):

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

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 $\mu > 0$ such that, for any close enough point x, y, one has $dist(f(x), f(y)) \ge \mu\, dist(x, y)$.

2.2. **Linear maps on $\mathbb{R}^d$.** Let $f = f_A$ be given by an invertible matrix $A \in GL(d, \mathbb{R})$: $f(x) = Ax$. The origin is always a fixed point. What kind of fixed points? This depends on the spectrum of A. For a real matrix, eigenvalues are either real, or come in pairs $(\lambda, \bar{\lambda})$. We call E_λ the generalized eigenspace (resp. the union of the two generalized eigenspaces of $\lambda, \bar{\lambda}$). They can be split into

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

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

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

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

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

(1) if $E^0 = E^+ = \{0\}$ that is the eigenvalues of A satisfy $\max |\lambda_i| \stackrel{\text{def}}{=} r(A) < 1$, then 0 is an *attracting fixed point*. Eventhough one may have $\|A\| > 1$, the iterates satisfy $\|A^n\| \leq C(r(A) + \epsilon)^n$, and are 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 $|\lambda| = 1$. If $\lambda \notin \mathbb{R}$, this leads to the study of rotations on S^1.

Remark 2.2. The study of linear maps already provides some hints on structural stability. Inside $GL(n, \mathbb{R})$, contracting/hyperbolic matrices form an open set, meaning that for each contracting/hyperbolic A, small enough perturbations $A + \delta A$ will still be contracting/hyperbolic with the same number of unstable/stable directions.

2.3. **Circle rotations.** The 1-dimensional phase spaces already show interesting features. 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 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 see that this transformation admits a unique invariant measure (on the Borel σ-algebra), namely is the Lebesgue measure on S^1: such a map is called *uniquely ergodic*.

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

A similar dichotomy occurs for more general orientation-preserving homeomorphisms of $f : S^1 \to S^1$. Each such homeomorphism can be lifted into a homemorphism $F : \mathbb{R} \to \mathbb{R}$, which is of degree 1 (that is, $F(x+1) = F(x)+1$. To such an F is associated a single *rotation number*

$$\rho(F) = \lim \frac{F^n(x) - x}{n}, \quad \text{independent of } x.$$

We call $\rho(f) = \rho(F) \bmod 1$. There is a single lift F s.t. $\rho(F) \in [0,1)$.

Claim 2.4. $\rho(f)$ is rational if and only if f admits a periodic orbit. If $\rho(f) = \frac{p}{q}$, then every periodic point x has minimal period q, and $F^q(x) = x + p$.

2.4. **Expanding maps on the circle.** A smooth noninvertible smooth map on S^1 is the projection on $S^1 = \mathbb{R}/\mathbb{Z}$ of the dilation by an positive integer $\tilde{E}_m(x) = mx$, $m \in \mathbb{N}$:

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

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 $dist(E_m(x), E_m(y)) = m\, dist(x, y)$. Its iterates have the same form: $E_m^n = E_{m^n}$.

Each (small enough) interval 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

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

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. Notice that 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 fine properties, is facilitated when one notices the (semi)conjugacy 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 $x_i \in \{0, 1, \ldots, m-1\}$. 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^+:

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

Equivalently, the following diagram commutes:

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

One then 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 $x_1 x_2 \cdots x_n 1000 \cdots \equiv x_1 x_2 \cdots x_n 0111 \cdots$. This defect of injectivity is not very significant when counting periodic points: σ has exactly m^n points of period n (the shift is due to the two fixed points $00000 \equiv 111111$).

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

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

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 or Lipschitz) map $\Sigma_m^+ \to S^1$. Through this conjugacy, one can easily (re)-construct:

(1) the periodic orbits: an n-periodic point is the image of a n-periodic sequence $\boldsymbol{x} = \overline{x_1 x_2 \cdots x_n}$.
(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.
(4) nontrivial (fractal) invariant measures. Ex: the push-forward of *Bernouilli measures* on Σ_m^+ (see definition).

2.4.2. *Another way to interpret the semiconjugacy. Structural stability of E_m.* For the simple 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 g which of topological degree m and is expanding. The construction proceeds by re-interpreting the semiconjugacy between E_m and Σ_m^+ in terms of a Markov 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 each $\Delta_{\alpha_1\cdots\alpha_n}$ satisfies $E_m^n(\Delta_{\alpha_1\cdots\alpha_n}) = [0, 1]$ in a 1-to-1 correspondence. It is an interval of the form $[\frac{k}{m^n}, \frac{k+1}{m^n}]$, which (here) 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 that each infinite intersection $\Gamma_{\boldsymbol{\alpha}}$ 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 $C_{\alpha_1\cdots\alpha_n}$, mapping 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 easily check that the map h is well-defined, bijective and is bicontinuous on S^1, and that it satisfies

$$E_m \circ h = h \circ g.$$

It thus topologically conjugates E_m and g.

2.4.3. *A variation on the proof of semiconjugacy between E_m and g.* The above conjugacy equation can also be solved by interpreting this equation as the fixed point for a contracting operator (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 $dist(h_1, h_2) = \max_x |h_1(x) - h_2(x)|$. 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:

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

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 above semiconjugacy.

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 $(x_i)_{i \ge 1}$ on m symbols. One can also define the two-sided sequences $(x_i)_{i \in \mathbb{Z}}$, and let the shift σ act on them. The space of bi-infinite sequences is denoted by Σ_m. We call (Σ_m, σ) the full two-sided shift on m symbols. As opposed to the one-sided shift, the two-sided one is an invertible, bicontinuous map. Still, 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 full 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**: they are defined by an $m \times m$ matrix $A = (A_{kl})$ with entries given by 0 or 1, called the *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, because all properties of Σ_A 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.

More generally, a subshift of finite type (of type k) is given by forbidding certain $k+1$-words, that is certain combinations $x_i x_{i+1} \cdots x_{i+k}$.

Exercise 2.5. Each SFT of type k is conjugate to a certain SFT of type 1 (i.e. a topol. Markov chain).

Hint: change the alphabet.

Exercise 2.6. The number of n-words in Σ_A starting with $x_1 = \epsilon_1$ and ending with $x_n = \epsilon_n$ is given by $(A^n)_{\epsilon_1 \epsilon_n}$. The number of n-periodic points of Σ_A is given by $\mathrm{tr} A^n$.

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).

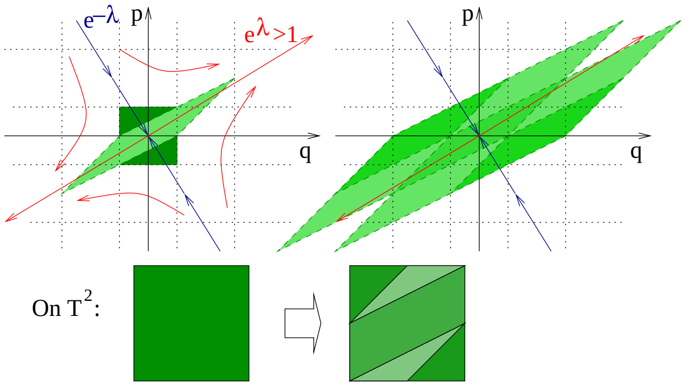

FIGURE 2.1. Arnold's cat map

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 2.7. *[Perron]*

Let A be a primitive *$m \times m$ matrix with nonnegative entries. Then A has a positive eigenvalue λ with the following properties: λ is simple, λ has a positive eigenvector v (that is, all components $v_i > 0$), every other eigenvalue satisfies $|\lambda'| < \lambda$, the only non-negative eigenvectors are multiples of v.*

2.6. **Hyperbolic torus automorphisms ("Arnold's cat map").** We now get back to smooth maps on manifolds. A linear automorphism on $\mathbb{T}^d = \mathbb{R}^d/\mathbb{Z}^d$ is given by projecting on $\mathbb{T}^d$ the action on $\mathbb{R}^d$ of a matrix M 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 obviously 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

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

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 $E_x^{\pm} = E^{\pm}$ are independent of x. They are invariant through the map: $df(x)E_x^{\pm} = E_{f(x)}^{\pm}$. The tangent map $df(x)$ acts on E_x^- (resp. E_x^+) by a contraction (resp. a dilation).

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 $dist(f^n(x), f^n(y)) \xrightarrow{n \to +\infty} 0$. 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 $dist(f^n(x), f^n(y)) \xrightarrow{n \to -\infty} 0$.

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 2.8. 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 2.9. Count the number of n-periodic points for a hyperbolic automorphism A on $\mathbb{T}^2$.

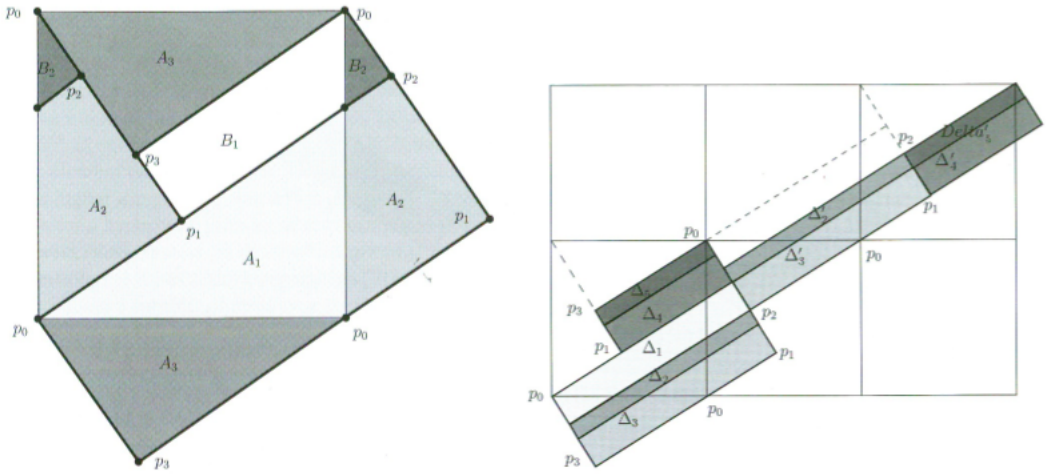

FIGURE 2.2. Adler and Weiss's Markov partition for Arnold's cat map

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 (equivalently, the Lebesgue measure is ergodic).

2.6.1. *Markov partition for Arnol'd cat map.* One first defines two rectangles $A, B \subset \mathbb{T}^2$ 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 $\Delta_1, \ldots, \Delta_5$. 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 below for the 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:

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

From this matrix we construct a (double-sided) the 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 consists in a single point, which we denote by $x(\boldsymbol{\alpha}) = \pi(\boldsymbol{\alpha})$. Hence we have a semiconjugacy between the subshift (Σ_A, σ) and f, similar with the semiconjugacy between (Σ_m^+, σ) and E_m:

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

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

Remark 2.10. One can check that the Perron-Frobenuis eigenvalue of A is exactly given by $\lambda = \frac{3+\sqrt{5}}{2}$. This is coherent 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 linear map M, that is $g = M + \delta g$, $\|\delta g\|_{C^1} \leq \epsilon$. We want to prove the following property.

Theorem 2.11. *The linear hyperbolic automorphism M is C^1-structurally stable.*

Proof. We want to solve the semiconjugacy equation

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

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, so the perturbation δg is $\mathbb{Z}^2$-periodic. Similarly, h must be in the same homotopy class as the identity, so $h = Id + \delta h$, with δh biperiodic. The above equation thus reads:

$$\begin{aligned} &M^{-1} \circ (Id + \delta h) \circ (M + \delta g) = Id + \delta h \\ &\iff M^{-1} \circ \delta g + M^{-1} \circ \delta h \circ g = \delta h. \end{aligned} \tag{2.4}$$

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

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

where $h_\pm$ are continuous biperiodic functions on $\mathbb{R}^2$. 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 (2.4) 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 (2.3).

In order to prove that we have a conjugacy, one can solve the symmetrical equation

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

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

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

These two components can be easily inverted by Neumann series:

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

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 $\mathcal{L}^{-1}$ is not contracting a priori. The equation (2.5) can then be rewritten

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

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. The full operator $\mathcal{L}^{-1}\mathcal{G}$ will also be contracting, and admit a unique fixed point $\tilde{h}_0$. One easily checks that $h_0 \tilde{h}_0$ commutes with M, and is thus equal to the identity. □

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

Theorem 2.12. *Any Anosov diffeomorphism is structurally stable.*

2.7. **Quadratic maps on the interval.** A family of (very) simple polynomial maps on $\mathbb{R}$ has been extensively studied, because in spite if its simplicity it features various interesting dynamical phenomena. These maps depend on a real parameter $\mu > 0$, and are given 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. Its single fixed point (on I) is the origin, which is attracting.
(2) for $\mu > 1$, the origin becomes a repelling fixed point (because $q'_\mu(0) > 1$), but q_μ admits a second fixed point $x_\mu = 1 - 1/\mu$. This latter is attracting for $\mu < 3$ ($-1 < q'_\mu(x_\mu) < 0$). It becomes repulsive for $\mu > 3$.
(3) for $\mu > 3$ the fixed point x_μ becomes repulsive, and an attractive period-2 periodic orbit appears nearby. $\mu = 3$ is 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 set of trapped points when $\mu > 4$.

Proposition 2.13. *For $\mu > 4$ the trapped set Λ_μ is a Cantor set*[2] *in I. The restriction $q_\mu \restriction \Lambda_\mu$ is (topologically) conjugate with the full shift (Σ_2^+, σ).*

[2]Topologically, a Cantor set is a closed set which is perfect (has no isolated points) and is nowhere dense in I.

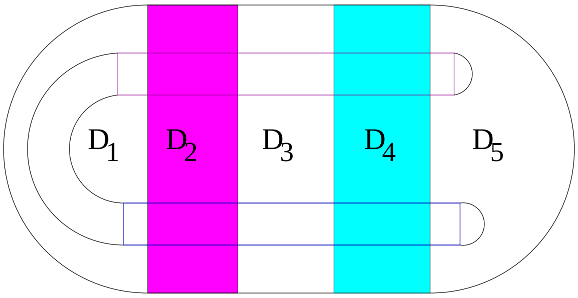

FIGURE 2.3. Smale's horseshoe

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

$$I_{\epsilon_1\epsilon_2\cdots\epsilon_n} = f_{\epsilon_n} \circ f_{\epsilon_{n-1}} \circ \cdots \circ f_{\epsilon_1}(I).$$

Observe that

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

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_n}, then its first iterate $q_\mu(x) \in I_{\epsilon_{n-1}}$, and so on: $q_\mu^j(x) \in I_{\epsilon_{n-j}}$, up to finally $q_\mu^n(x) \in I$.

This property shows that for each $n > 0$ the intervals $\{I_{|\boldsymbol{\epsilon}|}, |\boldsymbol{\epsilon}| = n\}$ 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 $0 \le j \le n$. As a result the trapped set can be defined as the closed set

$$\Lambda_\mu = \bigcap_{n\ge1} I^n.$$

For $\mu > 2 + \sqrt{5}$ the maps f_i are contracting: $|f_i'(x)| \le \lambda_\mu < 1$, $\lambda_\mu = \mu\sqrt{1 - \frac{4}{\mu}}$. As a result, the length of the intervals $I_{\boldsymbol{\epsilon}}$ decreases exponentially as $|I_{\boldsymbol{\epsilon}}| \le \lambda_\mu^{|\boldsymbol{\epsilon}|}$. 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 proof is more tricky in the case $\mu > 4$, but the conclusion is the same. □

2.8. **Smale's horseshoe.** Smale's horseshoe is a 2-dimensional nonlinear map, which could be considered as the hyperbolic analogue of the above quadratic maps. It 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 rectangle R is split between three vertical rectangles D_2, D_3, D_4 of height and width $= 1/3$. The main assumption on f is the following:

(1) $f_{\restriction D_2}$ is a similarity, which stretches D_2 vertically by a factor $\lambda < 1/2$ and expands it horizontally by a factor $\mu > 3$, such that $f(D_2)$ intersects both D_1 and D_5.

(2) similarly, $f_{\restriction D_4}$ is a similarity, with the same contraction and expansion constants.
(3) the map $f_{\restriction D_3}$ is nonlinear, $f(D_3)$ is contained in D_1.
(4) $f(D_1) \subset D_5$, $f(D_5) \subset D_5$.

The map f is obviously not surjective.

The two rectangles $R_0 \stackrel{\text{def}}{=} f(D_2) \cap R$, $R_1 \stackrel{\text{def}}{=} f(D_4) \cap R$ (of height λ and width 1) are disjoint, and $f(R) \cap R = R_0 \cup R_1$.

Similarly, $f(R_0) \cap R$ is the union of two rectangles of width λ^2 and width 1, respectively contained in R_0 and R_1, which we denote by $R_{00} = f(R_0) \cap R_0$ and $R_{01} = f(R_0) \cap R_1$. The same holds for $f(R_1) \cap R = R_{10} \cup R_{11}$. One can proceed iteratively, and define the rectangles

$$R_{\epsilon_0\epsilon_1\cdots\epsilon_{n-1}} = \bigcap_{j=0}^{n-1} f^j(R_{\epsilon_j})$$

of heights λ^n and width 1. As in the previous section, for each $\boldsymbol{\epsilon} \in \Sigma_2^+$, the set $R_{\boldsymbol{\epsilon}} = \bigcap_{j=0}^{\infty} f^j(R_{\epsilon_j})$ is a single horizontal segment (of width 1). By construction, each point $x \in R_{\boldsymbol{\epsilon}}$ has a (unique) infinite backwards trajectory through f. The set

$$H^+ \stackrel{\text{def}}{=} \bigcup_{\boldsymbol{\epsilon}\in\Sigma_2^+} R_{\boldsymbol{\epsilon}} = \bigcap_{j\geq 0} f^j(R)$$

is the product of a vertical Cantor set by a horizontal segment. This set is forward-invariant.

One can proceed similarly in the backwards direction. $f^{-1}(R_0)$ and $f^{-1}(R_1)$ are disjoint vertical segments of width μ^{-1} and height 1, respectively contained in D_2 and D_4. The iterates $f^{-1}(R_{\epsilon_{-1}}) \cap f^{-2}(R_{\epsilon_{-2}}) \cap \cdots \cap f^{-n}(R_{\epsilon_{-n}})$ are vertical rectangles of width μ^{-n}, and for each $\boldsymbol{\alpha} = \cdots \alpha_{-2}\alpha_{-1} \in \Sigma_2^-$ the set

$$R_{\boldsymbol{\alpha}\cdot} = \bigcap_{j\leq -1} f^j(R_{\alpha_j})$$

is a vertical segment. The set $H^- \stackrel{\text{def}}{=} \bigcup_{\boldsymbol{\alpha}\in\Sigma_2^-} R_{\boldsymbol{\alpha}\cdot} = \bigcap_{j\leq -1} f^j(R)$ is the product of a horizontal Cantor set with the vertical interval. It is made of points x whose forward trajectory always remains in R.

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 call $\boldsymbol{\beta} \stackrel{\text{def}}{=} \boldsymbol{\alpha} \cdot \boldsymbol{\epsilon}$ the bi-infinite sequence in $\Sigma_2 = \Sigma_2^- \times \Sigma_2^+$. By construction, the intersection $R_{\boldsymbol{\beta}} = R_{\boldsymbol{\alpha}\cdot} \cap R_{\boldsymbol{\epsilon}}$ 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 tangent to the map is the same hyperbolic matrix $df(x) = \begin{pmatrix} \mu & 0 \\ 0 & \lambda \end{pmatrix}$. 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).

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: the rectangles have small enough diameters; 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 a horizontal segments of length unity (intersecting 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 segments of length μ (intersecting Λ), its image is a vertical segment of length μ^2 intersecting Λ, and fully contained in eigher $\mathcal{R}_0$ or $\mathcal{R}_1$, so the third property is OK.

Lemma 2.14. *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 2.15. Describe the unstable and stable manifolds of $x \in \Lambda$, defined by

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

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 one takes $\omega = \sum_{i=1}^d dp_i \wedge dq_i$.

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 on X, given by

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

Equivalently, the solutions to the flow equation are trajectories $(q(t), p(t))$ satisfying

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

The flow preserves the symplectic form on T^*M:

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

so that the variation of $\sum_i dq_i \wedge dp_i$ gives

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

(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 set. The Liouville measure $d\mu_E = \delta(H(q,p) - E)dvol$ on $H^{-1}(E)$ is naturally invariant. The Hamiltonian flow on $H^{-1}(E)$ can be analyzed with the above tools.

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

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

(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: $ds^2 = \sum_{ij} g_{ij}dx_idx_j$). The dynamics on the unit cotangent bundle $H^{-1}(1/2) = S^*M$ of unit momenta, is equivalent with the *geodesic flow*, which lives on the space SM of unit velocities. Here, the Liouville measure on S^*M projects onto the Lebesgue measure on M.

Depending on the topology of M and the riemannian metric g on it, the dynamical properties of the geodesic flow are very 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.

Euclidean billiards. Another possibility is to restrict the motion of the free particle inside a bounded region of M, with specular reflection at the boundaries. For instance, a bounded connected domain $D \subset \mathbb{R}^2$ is called a billiard. The particle moves at velocity $|\dot{q}| = 1$ along straight lines inside the domain, and is reflected when touching the boundary (which is, say, C^1 so that the tangent 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 (that is, on the boundary ∂D).

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:

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

One can easily show 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 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.

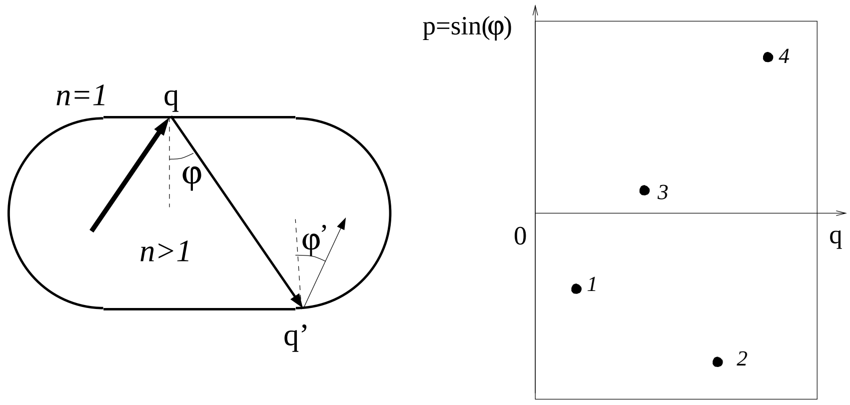

FIGURE 2.4. A Euclidean billiard and its associated billiard map

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[3]. 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 $x \in X$.

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

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 $Y \subset X$.*

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[4] $\{f^n(x_0),\ n \in \mathbb{N}\}$ 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:

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

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.

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

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. 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

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

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*

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

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*[5] 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[6]. 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.

[5] $\operatorname{supp} \mu$ *is the intersection of all closed sets of full measure. Equivalently, its complement is the union of all null open sets.*

[6] More precisely, the elements of L^p are equivalence classes of functions, $f \sim g$ iff $f(x) = g(x)$ almost everywhere.

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*

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

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 $\mu\left(T^{-1}(A)\Delta A\right) = 0$ [7].

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.

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

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.

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 $E \stackrel{\rm def}{=} \{x \in X,\ \rho_1(x) < 1\}$. 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 $F \stackrel{\rm def}{=} \{x \in X,\ \rho_1(x) > 1\}$ 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*

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

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 $\bar f \stackrel{\rm def}{=} \Pi_0 f$. 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: $\|f_n - \bar{f}\|_2 \stackrel{n\to\infty}{\to} 0$) 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 $F_n(x) \stackrel{\text{def}}{=} \max_{k\leq n} k f_k(x)$. For a given $x \in X$, the sequence $(F_n(x))_{n\geq 1}$ 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

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

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 $f_n \leq \frac{F_n}{n}$, so for any $x \notin A_f$ (that is, for μ-a.e. x) one has

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

and hence $\limsup_n \varphi_n(x) \leq \varphi_{\mathcal{I}}(x) + \epsilon$. a.e. Since this holds for any $\epsilon > 0$, we have $\limsup_n \varphi_n(x) \leq \varphi_I(x)$ a.e. Applying the same reasoning to the observable $-\varphi$, we get $\liminf_n \varphi_n \geq \varphi_{\mathcal{I}}$ 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

$$\lim_{n\to\infty} \mu(T^{-n}(A) \cap B) = \mu(A)\,\mu(B). \tag{4.2}$$

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[8] $f^{(K)}(x) = \sum_{|k|\leq K} \hat{f}_k\, e_k(x)$, 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

$$\forall x \in \mathbb{T}^2, \qquad \frac{1}{T}\int_0^T dt\, e_{\mathbf{m}}(T^t_\alpha(x)) = \frac{1}{T}\int_0^T dt\, e^{2i\pi(m_1\alpha+m_2)t}\, e_{\mathbf{m}}(x) = \frac{1}{T}\frac{e_{\mathbf{m}}(\alpha,1) - 1}{m_1\alpha + m_2}\, e_{\mathbf{m}}(x) \overset{T\to\infty}{\to} 0, \tag{4.3}$$

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

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

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

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

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).

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.

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:

$$\forall n \geq 1, \ \forall \epsilon_1 \cdots \epsilon_n, \qquad \nu_{\boldsymbol{p}}(C_{\epsilon_1 \epsilon_2 \cdots \epsilon_n}) = \prod_{i=1}^{n} p_{\epsilon_i}. \tag{4.4}$$

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 $\boldsymbol{p}_{\max} = \{\frac{1}{m}, \ldots, \frac{1}{m}\}$. 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. $\square$

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*[9].

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.

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 $\left\{e_{\mathbf{m}},\ \mathbf{m}\in\mathbb{Z}^d\right\}$ of $L^2(\mathbb{T}^d)$. For any $(\mathbf{m},\mathbf{n}) \neq (0,0)$, we have

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

Since the orbits $\{{}^t\!M^n\mathbf{n},\ n\in\mathbb{Z}\}$ 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 (4.2) 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 (4.3), namely the unique ergodicity of irrational translation flows).

9

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 $\nu(A) = 0$.*

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 (4.4). 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:

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

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. □

References

[1] Michael Brin et Garrett Stuck, Introduction to Dynamical Systems, Cambridge 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`