Discrete dynamical systems (Tsinghua, february–march 2017)

Alain Chenciner

Observatoire de Paris, IMCCE (UMR 8028), ASD

chenciner@imcce.fr

&

Département de mathématique, Université Paris VII

Abstract

Originating in the qualitative studies of differential equations by Henri Poincaré (Poincaré sections), the notion of "discrete dynamical system" is understood to day as the study of the orbits of the action of a discrete group, such as $\mathbb{Z}$ (or a discrete monoid such as $\mathbb{N}$) on a set X. In other words, in the case of an action of $\mathbb{Z}$ we are interested in the properties of the orbits $T^n(x), n \in \mathbb{Z}$ of an invertible map $T : X \to X$, with particular emphasis on their asymptotic properties when $n \to \pm\infty$. The so-called *qualitative theory* concerns the case when X is a topological space (resp a C^k-manifold) and T a continuous map (resp. a C^k map), while the *ergodic theory* concerns the case of a measure space X and a measure preserving map T. In these notes, we address mainly the ergodic side, leaving aside many basic notions like hyperbolicity.

Contents

1 First examples: linear maps, linear automorphisms and rotations on tori, annulus twists

A *discrete topological dynamical system* is a continuous map : $T : X \to X$ from a metric (or at least metrisable) space X to itself. Iterating T defines an action $n \cdot x = T^n(x)$ of the semi-group $\mathbb{N}$, which extends to an action of the group $\mathbb{Z}$ if T is a homeomorphism. In this last case, it can be considered as the discrete version of an autonomous differential equation whose flow is defined for all times (and hence defines an $\mathbb{R}$-action). We start with basic examples given by analytic maps from an analytic manifold to iself. This is the occasion to introduce some important words: the *orbit* of $x \in X$ under the action of $\mathbb{N}$ (resp. $\mathbb{Z}$) is the set

$$\mathcal{O}_T^+(x) = \{T^n(x), n \in \mathbb{N}\}, \quad \text{resp.} \quad \mathcal{O}_T(x) = \{T^n(x), n \in \mathbb{Z}\}.$$

In case T is a homeomorphism, $\mathcal{O}_T^+(x)$ is called the *positive orbit* of x in order to avoid confusion. One says that the orbit of x is *periodic* if there exists an integer $n \geq 1$ such that $T^n(x) = x$. One also says that x is a *periodic point of period* n of T (a *fixed point* if $n = 1$).

Definition 1 *The dynamical system $T : X \to X$ is said to be positively topologically transitive (resp. positively minimal) if there exists $x \in X$ such that (resp. if for every $x \in X$) the closure $\overline{\mathcal{O}}_T^+(x)$ of its positive orbit is equal to X. The homeomorphism T is said to be topologically transitive (resp. minimal) if there exists $x \in X$ such that (resp. if for every $x \in X$) the closure of its orbit $\overline{\mathcal{O}}_T(x)$ is equal to X.*

For another definition of topological transitivity, equivalent if the space X is reasonable, see Lemma 4.

Definition 2 *A subset $A \in X$ is called T-invariant (or invariant by T) if $T^{-1}A = A$ (compare to definition 16).*

Exercise 1 *A homeomorphism $T : X \to X$ is minimal if and only if the only closed invariant subsets are $\emptyset$ and X.*

1.1 Linear maps

Let X be a finite dimensional real or complex vector space and $T : X \to X$ be a linear map. Using the spectral decomposition on $\mathbb{C}$ one can write X as the direct sum of invariant subspaces on each of which T takes a simple form in a well-chosen basis.

Exercise 2 *Describe all the possible behaviours of the orbits of a linear map $T : \mathbb{R}^2 \to \mathbb{R}^2$.*

Exercise 3 *Prove that a linear map $T : E \to E$ of a finite dimensional vector space on $K = \mathbb{R}$ or $\mathbb{C}$ into itself cannot be transitive.*

A direct proof could follow from the classification of such linear maps but a much nicer proof, borrowed from [Fa], can be obtained by showing that if T had a dense orbit in E, its transpose $T^* : E'_{\mathbb{C}} \to E'_{\mathbb{C}}$ (defined by $\varphi \mapsto \varphi \circ T$ on the $\mathbb{C}$-vector space $E'_{\mathbb{C}}$ be formed by the K-linear maps from E to $\mathbb{C}$) could not have any eigenvalue.

Caution ! *this is definitely false in infinite dimension. A surprising example is the derivation $T(\varphi) = \varphi'$ from the Fréchet space $C^\infty([0,1], \mathbb{R})$ to itself: there exists a (see Exercise 7).*

1.2 Linear automorphisms on tori

One defines $\mathbb{T}^r$ as the quotient $\mathbb{T}^r = \mathbb{R}^r/\mathbb{Z}^r$.

Exercise 4 *$\mathbb{T}^r$ inherits from $\mathbb{R}^r$ a quotient topology and a quotient group structure: if $\pi : \mathbb{R}^r \to \mathbb{T}^r$ is the quotient map, $\Omega \subset \mathbb{T}^r$ is open iff $\pi^{-1}(\Omega) \subset \mathbb{R}^r$ is open; the formula $\hat{x} + \hat{y} = \pi(x + y)$ defines the group law independently of the choice of representatives x, y of $\hat{x} = \pi(x)$ and $\hat{y} = \pi(y)$.*

If $f : \mathbb{R}^r \to \mathbb{R}^r$ is such that $x - y \in \mathbb{Z}^r \implies f(x) - f(y) \in \mathbb{Z}^r$, the composition $\pi \circ f$ defines uniquely a map $F : \mathbb{T}^r \to \mathbb{T}^r$ such that $\pi \circ f = F \circ \pi$. In particular, if the linear map $A : \mathbb{R}^r \to \mathbb{R}^r$ is such that $A(\mathbb{Z}^r) \subset \mathbb{Z}^r$, it induces a map $\overline{A} : \mathbb{T}^r \to \mathbb{T}^r$.

Exercise 5 *$\overline{A}$ is surjective if and only if $\det A \neq 0$, it is injective if and only if $\det A = \pm 1$.*

In particular, a linear automorphism of $\mathbb{T}^r$ preserves *Haar measure* m: for any measurable subset $E \subset \mathbb{T}^r$, one has $m(\overline{A}^{-1}(E)) = m(E)$ (see 2.3 and lemma 19for definitions).

Proposition 1 *If $\overline{A} : \mathbb{T}^r \to \mathbb{T}^r$ is linear automorphism, the periodic points are dense in $\mathbb{T}^r$.*

Proof. The subset $\mathbb{Q}^r/\mathbb{Z}^r$ is dense in $\mathbb{T}^r$ and each of its elements is a periodic point of $\overline{A}$. Indeed, for any $q \in \mathbb{N} \setminus \{0\}$, the map $\overline{A}$ sends the finite set $(\frac{1}{q}\mathbb{Z}^r)/\mathbb{Z}^r$ to itself. But a bijection of a finite set is necessarily periodic.

Exercise 6 *A quite famous example in the history of discrete dynamical systems – a parangon of Anosov diffeomorphism – is the linear automorphism $\overline{A}$ of the 2-torus $\mathbb{T}^2$ defined by the matrix $A = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$. Using the irrationality of the eigenvalues of A and lemma 2 below, show that the images $\overline{E}_s, \overline{E}_u$ under the canonical projection $\pi : \mathbb{R}^2 \to \mathbb{T}^2$ of the eigenspaces E_s, E_u of A are dense in $\mathbb{T}^2$. Deduce that $\overline{A}$ is transitive.* **Hint:** *use criterion in lemma 4 below after noticing that given two open subsets U and V in $\mathbb{T}^2$, one can choose $x \in U \cap \overline{E}_s$, $y \in V \cap \overline{E}_u$ and a small enough neighborhood W of 0 such that $x + W \subset U$ and $y + W \subset V$, and that as soon as $n \in \mathbb{N}$ is large enough, $\overline{A}^n(x)$*

and $\overline{A}^{-n}(y)$ *both belong to* W*; conclude by noticing that* $x+\overline{A}^{-n}(y) \in U$ *and that* $\overline{A}^n\left(x+\overline{A}^{-n}(y)\right) = \overline{A}^n(x)+y \in V$ *(see figure 0.1).*

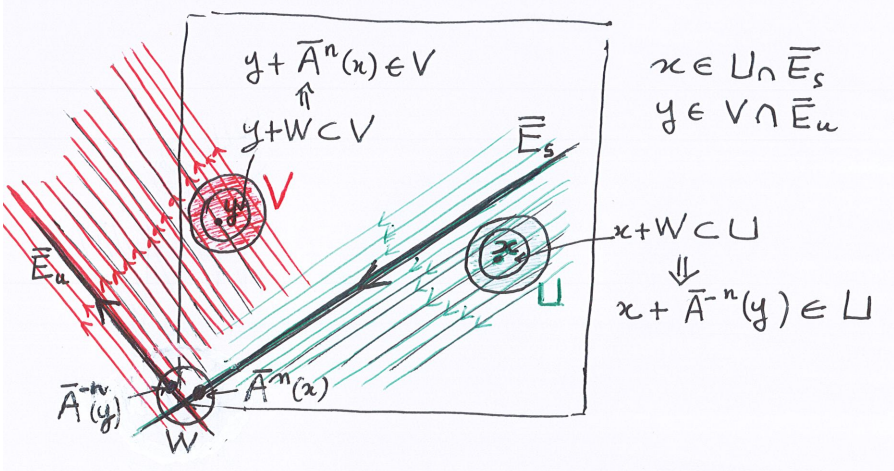


Figure 0.1 Proof of transitivity of $\overline{A}$.

The following figure, quite well known under the name of "Arnold's cat" shows the fate of a 2-dimensional cat lying in $\mathbb{T}^2$ when submitted to iterates of $\overline{A}$.

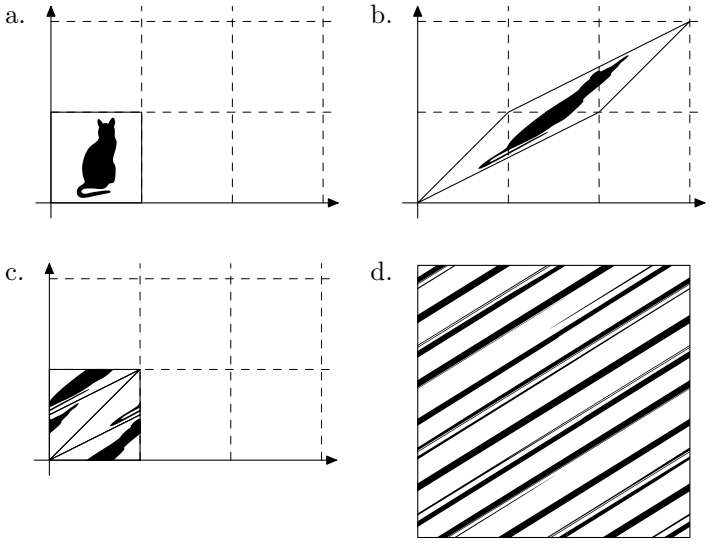


Figure 0.2 Arnold's cat and its image under the third iterate of $\overline{A}$.

Exercise 7 *2) Surprizingly, essentially the same proof shows that the derivation* $T(\varphi)=\varphi'$ *from the Fréchet space* $C^\infty([0,1],\mathbb{R})$ *(endowed with the distance* $d_\infty(\varphi,\psi)=\sum_{n\in\mathbb{N}}\frac{1}{2^n}\inf\left(||\varphi^{(n)}-\psi^{(n)}||_0,1\right)$*) to itself is transitive. Indeed, let*

A and B be the continuous linear maps from $C^\infty([0,1],\mathbb{R})$ to itself defined respectively by

$$A(\varphi)=\varphi', \qquad B(\varphi)(x)=\int_0^x \varphi(t)dt.$$

Check that $AB = Id$ and that, if P is a polynomial, $\lim_{n\to\infty} A^n(P) = 0$ and $\lim_{n\to\infty} B^n(P) = 0$. Admitting the density of polynomials in $C^\infty([0,1],\mathbb{R})$ (Weirstrass theorem), show that the proof of transitivity of $\overline{A}$ in Exercise 6 becomes a proof of transitivity of A when replacing $\overline{E}_s$ and $\overline{E}_u$ by the polynomials, $\overline{A}$ by A and $\overline{A}^{-1}$ by B.

1.3 Rotations on tori

The *translation*

$$R_\alpha = R_{\alpha_1,\alpha_2,\cdots,\alpha_r}(x_1,x_2,\cdots,x_r) = (x_1+\alpha_1, x_2+\alpha_2,\cdots,x_r+\alpha_r)$$

induces the *rotation* $R_\alpha = R_{\alpha_1,\alpha_2,\cdots,\alpha_r} : \mathbb{T}^r \to \mathbb{T}^r$ *(as there is no ambiguity, we shall keep the same notation for the translation of $\mathbb{R}^r$ and the corresponding rotation of $\mathbb{T}^r$).*

Lemma 2 *The real number α is irrational if and only if the rotation $R_\alpha : \mathbb{T}^1 \to \mathbb{T}^1$ is minimal.*

Proof. As the orbit of any point is just translated (rotated) from the orbit of 0, it is enough to show that the orbit of 0 is dense, that is to show that the additive subroup generated by α is dense. If α is rational, this subroup is finite. If α is irrational, the points in the orbit are all distinct; as the circle $\mathbb{T}^1$ is compact, there is an accumulation point, that is there exist integers $i \neq j$ and p such that $|i\alpha - j\alpha - p| = |(i-j)\alpha - p| < \epsilon$. The end of the proof is an easy exercise. Notice that taking positive (or negative) orbits would be sufficient. The analogous result on $\mathbb{T}^r, r \geq 1$ is

Theorem 3 (Kronecker) *Let $\alpha = (\alpha_1,\alpha_2,\cdots,\alpha_r) \in \mathbb{R}^r$. The real numbers $1,\alpha_1,\cdots,\alpha_r$ are rationally independent, i.e. there is no $(r+1)$-tuple $(k_0,k_1,\cdots,k_r) \in \mathbb{Z}^{r+1} \setminus \{0\}$ such that $k_0 + k_1\alpha_1 + \cdots + k_r\alpha_r = 0$, if and only if the rotation $R_\alpha : \mathbb{T}^r \to \mathbb{T}^r$ is minimal.*

Proof. Of course, here also it is enough to show that the orbit of 0 is dense, that is to show that the subgroup of $\mathbb{T}^r$ generated by (the class of) α is dense. There exist proofs based on the geometry of numbers (for ex. by Mahler) but here is a proof closer to the spirit of ergodic theory; it is based on the following

Lemma 4 (criterion of transitivity) *Let $T : X \to X$ be a homeomorphism of a complete metric space X with a countable basis of open sets and no isolated point. T is topologically transitive (resp. positively topologically transitive) If and only if for any two nonempty open sets $U, V \subset X$, there exists $n \in \mathbb{Z}$ (resp. $n \in \mathbb{N}$) such that $T^{-n}(U) \cap V$ is nonempty.*

This criterion is often taken as a definition of topological transitivity.

Proof of the "if" part. Let $(U_i)_{i\in\mathbb{N}}$ be a countable basis of (nonempty) open sets. By hypothesis, $\forall i \in \mathbb{N}$, $\cup_{n\in\mathbb{Z}} T^{-n}U_i$ intersects all the U_j, and hence is dense in X. Hence,

$$\{x \in X,\ \overline{\mathcal{O}}_T(x) = X\} = \cap_{i\geq 0}\big(\cup_{n\in\mathbb{Z}} T^{-n}U_i\big)$$

is a countable intersection of open and dense sets. One concludes by Baire's theorem (recalled below) that the set of x whose orbit is dense is dense in X, hence non empty. For the positively transitive part, just replace $\mathbb{Z}$ by $\mathbb{N}$.

Theorem 5 (Baire) *In a complete metric space (X, d), a countable intersection $\cap_{n\in\mathbb{N}} V_n$ of open dense subsets V_n is dense.*

Proof. Given U open, one constructs by induction a sequence of balls $B(x_n, r_n)$ such that

$$\overline{B}(x_n, r_n) \subset V_n \cap U \cap \mathring{B}(x_{n-1}, r_{n-1}) \quad \text{and} \quad 0 < r_n \leq 1/n.$$

One concludes that the closed balls $\overline{B}(x_n, r_n)$ are nested and that their centers converge to a point $x \in U \cap (\cap_{n\in\mathbb{N}} V_n)$.

Remark. Baire's theorem is the topological counterpart of the notion of full measure. It is the basis of the topological notion of *genericity*. For a beautiful parallel discussion of the two notions, see [O]. For an example, see Exercise 26

Proof of the "only if" part. If the orbit of x is dense, it intersects every non empty open set. Hence, given U and V, there exists $k, l \in \mathbb{Z}$ such that $T^k(x) \in U$ and $T^l(x) \in V$, hence $T^{(l-k)}(U) \cap V \neq \emptyset$. It is for the positively transitive case that the condition that X has no isolated point is needed. Indeed, this condition implies that an open set is necessarily infinite; now, if $T^p(x) \in V$ with $p \in \mathbb{N}$, $U' = U \setminus \{x, T(x), \cdots, T^p(x)\}$ is still open hence there exists m, necessarily strictly greater that p, such that $T^m(x) \in U'$, which implies that $T^p(x) \in T^{-n}(U) \cap V$ with $n = m - p > 0$.

End of the proof of theorem 3. To show that a homeomorphism T is not topologically transitive, it is enough to find a continuous non constant T-invariant function $f : X \to \mathbb{R}$; but, If $k_0 + k_1\alpha_1 \cdots k_r\alpha_r = 0$, $x \mapsto \sin\big(2\pi \sum_{i=1}^r k_i x_i\big)$ is a R_α-invariant function, hence R_α is not topologically transitive.

Under the hypotheses of lemma 4, to prove that a homeomorphism T is topologically transitive, it is enough to show that there are no two disjoint open non empty invariant sets $\tilde{U}, \tilde{V}$. Indeed, given arbitrary open subsets U, V of X, $\tilde{U} = \cup_{n\in\mathbb{Z}} T^{-n}U$ and $\tilde{V} = \cup_{n\in\mathbb{Z}} T^{-n}V$ are invariant. If they cannot be disjoint, there exists m, n such that $T^m U \cap T^n V \neq 0$, hence $T^{m-n}U \cap V \neq \emptyset$.

Now, let us suppose that U is a R_α-invariant open set and let $\mathcal{X}$ be its characteristic function. If the Fourier expansion of $\mathcal{X}$ is

$$\mathcal{X}(x_1, \cdots, x_r) = \sum_{(k_1,\cdots,k_r)\in\mathbb{Z}^r} \mathcal{X}_{k_1\cdots k_r} e^{2\pi i \sum_{j=1}^r k_j x_j},$$

the R_α-invariance implies

$$\mathcal{X}_{k_1 \cdots k_r}(1 - e^{2\pi i \sum_{j=1}^r k_j \alpha_j}) = 0.$$

If $1, \alpha_1, \cdots, \alpha_r$ are rationally independent, these identities imply that the only non-zero coefficient is $\mathcal{X}_{0\cdots 0}$, which means that $\mathcal{X}$ is constant outside a set of zero Lebesgue measure. One concludes that the complement of U has empty interior (any open set has positive Lebesgue measure), hence a fortiori, U cannot be disjoint from another invariant open set which proves that R_α is transitive.

We shall prove now a stronger version of Kronecker's theorem 3 (for another proof, see Proposition 37).

Theorem 6 *In Theorem 3, one can replace "minimal" by "positively minimal".*

As it suffices to deal with topological transitivity, this theorem is an immediate corollary of the following proposition:

Proposition 7 *Let X be a metric space without isolated point. For a continuous homeomorphism $T : X \to X$, topological transitivity and positive topological transitivity in the sense of criterion 4 are equivalent.*

A counter example is the map $x \mapsto x + 1$ from $\mathbb{Z}$ to itself.

Proof. 1) One starts proving the following key property:

> *if for every pair U, V of non empty open subsets of X there exists $k \in \mathbb{Z}$ such that $U \cap T^k V$ is non empty, then for any non empty U and any integer m, there exists $n \geq m$ such that $U \cap T^n U$ is not empty.*

By contradiction, suppose that there exists U and $n_0 \geq 0$ such that, $\forall n \geq n_0$, $U \cap T^n U$ is empty. Then, $\forall x \in U,\ \forall j \geq 0,\ T^i x \neq x$ (it cannot come back because U does not) and the same will be true of a small enough open neighborhood V (we are in a metric space): $\forall i \geq 1,\ V \cap T^i V = \emptyset$. As X has no isolated points, one can find two disjoint open sets $V_1, V_2 \subset V$. Then $\forall i, V_1 \cap T^i V_2 = \emptyset$ and $V_2 \cap T^i V_1 = \emptyset$. This implies $\forall k \in \mathbb{Z},\ V_1 \cap T^k V_2 = \emptyset$, which contradicts the hypothesis.

To conclude the proof of proposition 7, one notices that given U, V and $k \in \mathbb{Z}$ such that $U \cap T^k V$ is not empty, the key property implies that there exist $l > -k$ such that $(U \cap T^k V) \cap T^l (U \cap T^k V)$ is not empty. But this implies that $U \cap T^n V$ is not empty, with $n = k + l > 0$.

Exercise 8 *Give examples of rotations of $\mathbb{T}^r$ whose orbits are dense in a subtorus of $\mathbb{T}^r$. What condition insures that all the orbits are periodic ?*

Exercise 9 (Doubling the angle) *Show that the periodic points of the map $F : \mathbb{T}^1 \ni \theta \mapsto 2\theta \in \mathbb{T}^1$ are dense in $\mathbb{T}^1$ (a periodic point of period n is a point θ such that $F^n(\theta) = \theta$). Compare with the informations given by exercise 18.*

Remark: why are rotations on tori important ? They arise of course as invariant subsystems of linear maps having part of their spectrum on the unit

circle and hence are closely related to problems of stability and bifurcations. More significantly, linear flows on tori are the building blocks of *completely integrable Hamiltonian systems*, and their *Poincaré return maps* (see section 1.4) are rotations in well chosen coordinates. The perturbation theory of such completely integrable systems, known under the acronym of KAM (= Kolmogorov, Arnold, Moser) theory is one of the landmarks of the theory of dynamical systems in the 20th century. The existence of a Cantor set of invariant closed curves in Figure 2 is a consequence of this theory (see [C0]).

1.4 A glimpse to Poincaré return maps

To a C^∞ function $H : \mathbb{R}^{2n} \to \mathbb{R}$ (the *Hamiltonian*), one associates the *Hamiltonian vector-field*[1] X_H defined by

$$\frac{dx_i}{dt} = -\frac{\partial H(x,y)}{\partial y_i}, \ \frac{dy_i}{dt} = \frac{\partial H(x,y)}{\partial x_i}, \quad i = 1, \cdots, n$$

One checks immediately that the function H remains constant along any integral curve of X_H: this is the *conservation of energy*. Let $\omega = \sum_{i=1}^n dx_i \wedge dy_i$ be the *canonical symplectic form* on $\mathbb{R}^{2n} \equiv T^*\mathbb{R}^n$. Let $\Sigma_h = H^{-1}(h) \subset \mathbb{R}^{2n}$ be a regular energy manifold, and let ω_h be the restriction of ω to Σ_h. The restriction $X_{H,h}$ to Σ_h of the vector-field X_H can be shown to generate the kernel of the differential 2-form ω_h, which means that it is well defined up to multiplication by a scalar by the condition that for any $(x,y) \in \Sigma_h$ and any tangent vector $X \in T_{(x,y)}\Sigma_h$, one has $\omega_h(X_{H,h}, X) = 0$.

Definition 3 *A hypersurface $\mathcal{S} \subset \Sigma_h$ is called a local hypersurface of section (or a Poincaré section) of the vector-field $X_{H,h}$ if it cuts transversally its integral curves and if a first return map P (also called the Poincaré return map) can be defined, at least locally, by associating to a point of some domain $\mathcal{S}' \subset \mathcal{S}$ the first point on its positive integral curve which returns to $\mathcal{S}$.*

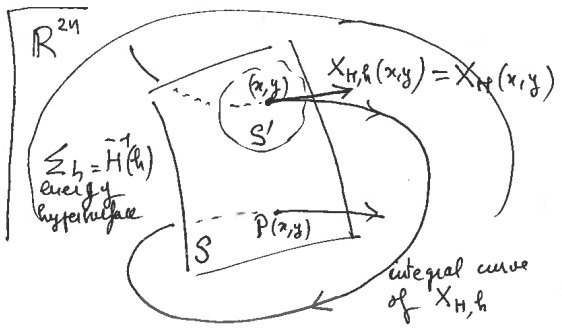


Figure 1. Poincaré section and return map

[1] The $2n$ coordinates in $\mathbb{R}^{2n}$ are written $x_1, \cdots, x_n, y_1, \cdots, y_n$ because, more conceptually, one should think of $\mathbb{R}^{2n}$ as the cotangent bundle $T^*\mathbb{R}^n$ of $\mathbb{R}^n$.

Iterating the first return map P gives informations on the flow of $X_{H,h}$ which do not depend on the law of time (i.e. one does not distinguish between $X_{H,h}(x,y)$ and its reparametrized versions $\lambda(x,y)X_{H,h}(x,y)$, where $\lambda : \Sigma_h \to \mathbb{R}$ is a smooth function). A fixed point of P corresponds to a periodic solution of $X_{H,h}$, an invariant curve of P corresponds to an invariant 2-torus of $X_{H,h}$, ... The above assertion that the restriction $X_{H,h}$ to Σ_h of the vector-field X_H generates the kernel of the differential 2-form ω_h implies that the restriction of this 2-form to a hypersurface of section $\mathcal{S}$ is non-degenerate, hence that its $(n-1)$th wedge product $\omega_h^{\wedge(n-1)}$ is a volume form on $\mathcal{S}$. In particular, 2 degrees of freedom autonomous Hamiltonians (i.e. when $n = 2$) give rise to diffeomorphisms of surfaces which preserve a smooth area form. Poincaré introduced this notion when studying the so-called *planar circular restricted 3-body problem* (see [C1, C2]). In the version given by Birkhoff, the surface of section is an annulus and the return map is an area preserving monotone twist map of the 2-dimensional annulus. The dynamics of the first return map is very complicated, in particular it contains as subdynamics both minimal rotations on $\mathbb{T}^1$ and Bernoulli shifts.

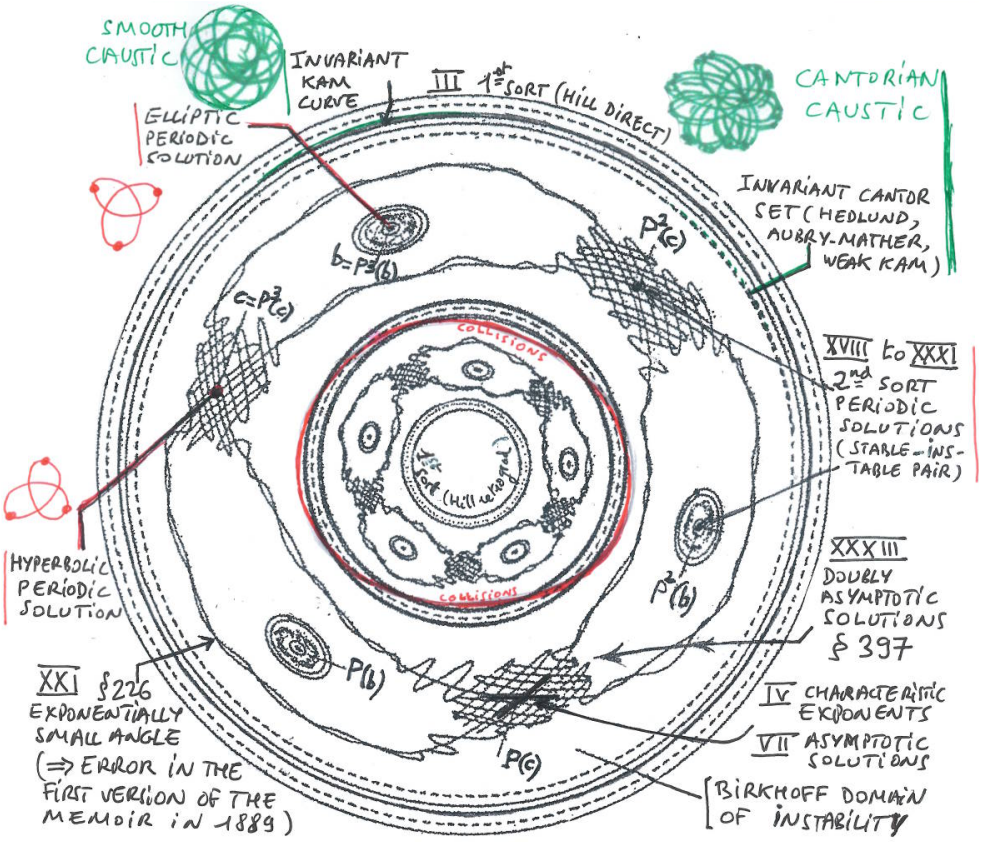


Figure 2. The return map of the restricted 3-body problem at high Jacobi constant.

Billiard maps Given a convex billiard table, i.e. a compact convex domain of $\mathbb{R}^2$ with smooth boundary Γ , a billiard trajectory is made of straight segments which reflect on Γ by changing the sign of the angle with the normal to Γ at the contact point. Such a trajectory is naturally associated to a map $T : A \to A$ of the annulus $A = S^1 \times [0, \pi]$ in the following way: let $\gamma : [0, 2\pi] \to \mathbb{R}^2$ be a parametrization of Γ by arclength t. To a couple (t, α) of a reflection point $\gamma(t)$ and the reflection angle α, the map T associates the couple (t_1, α_1) corresponding to the next reflection.

Lemma 8 *Let $l(t, t_1)$ be the length of the Euclidean chord between the points $\gamma(t)$ and $\gamma(t_1)$. One has*

$$\frac{\partial l}{\partial t}(t, t_1) = -\cos\alpha, \quad \frac{\partial l}{\partial t_1}(t, t_1) = \cos\alpha_1$$

Proof. See figure 3 for a proof in the spirit of Newton.

Exercise 10 *Compute the billiard map T for a billiard in a round disc.*

For more on billiards, see [Ta].

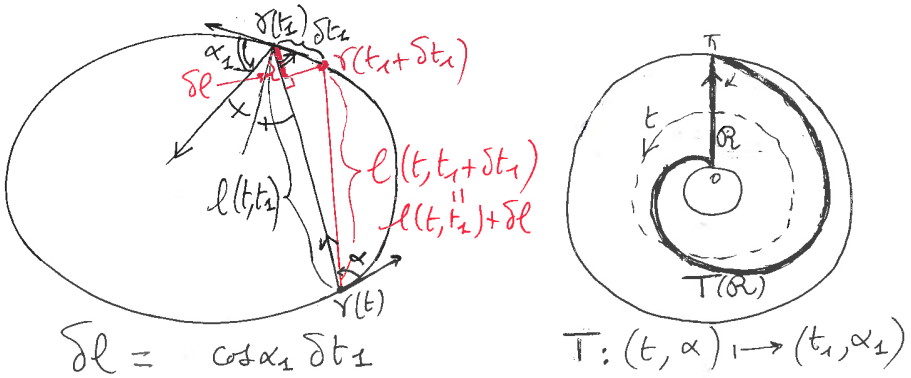


Figure 3. A billiard and the associated billiard map.

1.5 Some topological notions of recurrence

This section is very sketchy; for more see [LC].

The simplest examples of recurrence are the *fixed points* ($T(x) = x$) or the *periodic points* ($\exists n \in \mathbb{N}, T^n(x) = x$).

Definition 4 *Given a continuous map $T : X \to X$ and $x \in X$, its ω-limit set is the set of accumulation points of the sequence $(T^n(x)_{n\geq 0})$:*

$$\omega(x) = \{y,\ \exists n_i \to +\infty,\ \lim_{i\to+\infty} T^{n_i}(x) = y\}, \quad \textit{i.e.} \quad \omega(x) = \bigcap_n \overline{\{T^i(y),\ i \geq n\}}.$$

If T is a homeomorphism, one defines the α-limit set of x as the ω-limit set of T^{-1}, that is the set of accumulation points of the sequence $(T^{-n}(x)_{n\geq 0})$:

$$\alpha(x) = \{y,\ \exists n_i \to +\infty,\ \lim_{i\to+\infty} T^{-n_i}(x) = y\}, \quad \textit{i.e.} \quad \alpha(x) = \bigcap_n \overline{\{T^{-i}(y),\ i \geq n\}}.$$

If $x \in \omega(x)$ (resp. $x \in \alpha(x)$, one says that x is positively (resp. negatively) recurrent.

Note that $\omega(x)$ is closed and positively invariant, even invariant if T is a homeomorphism but that, in general, the set of recurrent points is not closed: an example is the "time τ map" of the pendulum equation (exercise 12).

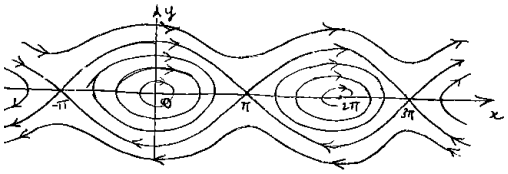
Definition 5 *The point x is said to be wandering if there exists a neighborhood U of x such that $\forall n \geq 1$, $T^{-n}(U) \cap U = \emptyset$. Otherwise x is said to be non wandering. One usually calls $\Omega(T)$ the set of non wandering points of T. An open set U such that $\forall n \geq 1$, $T^{-n}(U) \cap U = \emptyset$ is called a wandering domain.*

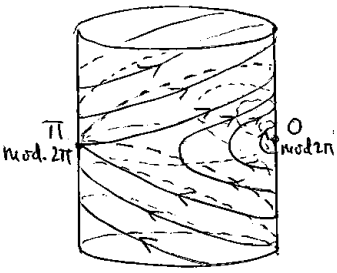
Exercise 11 *Show that the set $\Omega(T)$ of non wandering points is closed and positively invariant (and invariant if T is a homeomorphism). Show that a positively recurrent point is non wandering and that the same is true of a negatively recurrent point if T is a homeomorphism.*

Exercise 12 *The pendulum differential equation*

$$\frac{d^2x}{dt^2} + \omega^2 \sin x = 0, \quad \textit{that is} \quad \frac{dx}{dt} = y, \quad \frac{dy}{dt} = -\omega^2 \sin x,$$

defines a flow (a global $\mathbb{R}$-action) that is a 1-parameter group of diffeomorphisms of the plane $\varphi_t : \mathbb{R}^2 \to \mathbb{R}^2$, $t \in \mathbb{R}$ which factorizes through a 1-parameter group of diffeomorphisms of the cylinder $\hat{\varphi}_t : \mathbb{R}/2\pi\mathbb{R} \times \mathbb{R} \to \mathbb{R}/2\pi\mathbb{R} \times \mathbb{R}$, $t \in \mathbb{R}$.





Let $T = \varphi_\tau$ or $T = \hat{\varphi}_\tau$, where τ is a real number. Study in both cases the various invariant sets

$$Fix(T) \subset Per(T) \subset Rec(T) \subset \Omega(T)$$

(fixed points, periodic points, recurrent points, non wandering points).

2 From coin tosses to Bernoulli shifts: an introduction to measured spaces and measured dynamical systems

2.1 Algebras and σ-algebras

Let X be the set $\{0,1\}^n$ of sequences $\omega = (a_1, a_2, \cdots, a_n)$, where each a_i is 0 or 1. One can think of X as a sequence of n coin tosses, where 0 represents "tail" and 1 represents "head". An element of X (resp. a subset C of X) will be called an *elementary event* (resp. an *event*). The interpretation of such a subset is "it happens any one of the elements of C". Suppose now that the *probability* to get 0 is $P(0) = p$ and the probability to get 1 is $P(1) = q = 1 - p$; mathematically, this means endowing the finite set $\{0,1\}$ with a *probability measure*. To suppose the coin tosses are *independent* (this is the basic notion which makes probability theory distinct from measure theory) means that the elementary event $\omega = (a_1, \cdots, a_n)$ has probability $\mu(\omega) = p^{n_0} q^{n_1}$, where n_0 is the number of i such that $a_i = 0$ and $n_1 = n - n_0$ is the number of j such that $a_j = 1$. As there are exactly $\binom{n}{n_0} = \binom{n}{n_1}$ elementary events having the same probability as ω, we get that

$$\sum_{\omega \in X} \mu(\omega) = \sum_{n_0=0}^{n} \binom{n}{n_0} p^{n_0} q^{n-n_0} = (p+q)^n = 1.$$

This endows X with a probability measure μ, the probability $\mu(C)$ of an event C being defined as the sum of the probabilites of its elements.

The set $\mathcal{P}$ of all events, which is nothing but the set of all subsets of X is an *algebra* in the following sense :

Definition 6 *Let X be a set. A collection $\mathcal{G}$ of subsets of X is called an algebra if it contains X itself and if it is closed under the operations of formation of complements and finite unions.*

It follows from the definition that an algebra is also closed under finite intersections. In our case, as X is a finite set, $\mathcal{G}$ is also finite, and the following theorem asserts that it is equivalent to a *partition* of X :

Theorem 9 *If the algebra $\mathcal{G}$ is finite, there exists pairwise disjoint subsets $A_1, \ldots, A_r$ of X, called the "atoms", whose union is X, such that $\mathcal{G}$ coincides with the set of unions of atoms. One says that the atoms A_i "generate" $\mathcal{G}$. Conversely every partition $\mathcal{P}$ "generates" the finite algebra formed by the unions of its atoms.*

Sketch of proof (see [Si] page 4): consider the intersections $G_{\pm 1} \cap G_{\pm 2} \cap \ldots \cap G_{\pm r}$, where the G_i, $i = 1, \ldots, r$ are an enumeration of the elements of $\mathcal{G}$ and $G_{-i} = G_i^c$ is the complement of G_i.
In the case of the algebra $\mathcal{P}$ of all the subsets of the finite set X, the A_i are the elements $\{\omega\}$ of X, that is the elementary events.

Why introduce this notion of algebra ? Unavoidable when the cardinal of X is infinite, this notion is already encountered very naturally in the finite case : suppose, for example, that we are only able to count the number $f(\omega) = \sum_{i=1}^{n} a_i$ of 1's in $\omega = a_1 a_2 \dots a_n \in \{0,1\}^n$. The map $f : X \to \{0,1,2,\dots,n\}$ is an example of a *random variable* with finite values (called a *simple random variable* by Billingsley [B1]). Its levels $f^{-1}(k)$ form a partition of X and hence generate an algebra different from $\mathcal{P}$. The only probability we can measure is the one of elements of this algebra: more precise events are not perceived by our means of observation. If instead of considering f we consider the random variable with 2 values, which associates to $\omega \in X$ the parity $g(\omega)$ of the number of 1's, it means that we have only access to the coarser algebra generated by the partition into 2 parts, the "even" elementary events and the "odd" ones.

Exercise 13 *Compute the probabilities of the atoms A_i in the above examples.*

The infinite case: When the cardinal of X is infinite, the definition of an event is subtler. The formalization of the calculus of probabilities given in the 30's by Kolmogorov (see [Ko]) is anchored in Lebesgue measure theory and appeals to the fundamental notion of *σ-algebra*, or *tribe*, due to Emile Borel. The events are the subsets of X which can be "measured", that is the ones to which one can assign a probability.

Definition 7 *Let X be a set. An algebra $\mathcal{X}$ of subsets of X is called a σ-algebra (or a tribe) if it is closed under the operations of countable union. The pair $(X, \mathcal{X})$ is called a "measure space".*

Hence, a σ-algebra is closed under formation of complements and countable unions and intersections.

Lemma 10 *A family $\mathcal{G}$ of subsets of X (for example an algebra) being given, there exists a smallest σ-algebra $\mathcal{X}$ containing it, defined as the intersection of all the σ-algebras containing it. One says that $\mathcal{X} = \sigma(\mathcal{G})$ is the σ-algebra "generated" by $\mathcal{G}$.*

Definition 8 *A probability measure on the measure space $(X, \mathcal{X})$ is a function $\mu : \mathcal{X} \to [0,1]$ such that $\mu(X) = 1$ and $\mu(\cup C_i) = \sum \mu(C_i)$ if the C_i are a (at most) countable family of elements of $\mathcal{X}$ which, as subsets of X, are pairwise disjoint. One says that μ possesses the property of countable additivity (or σ-additivity). If $C \in \mathcal{X}, \mu(C)$ is the "probability" of the event C and the triple $(X, \mathcal{X}, \mu)$ is called a probability space.*

The restriction to countable families is justified by the fact that the non-zero elements of a summable family of real numbers form at most a countable subset. Kolmogorov had chosen as an alternative to the countable additivity axiom the *continuity axiom*, which is a kind of monotone convergence: if a decreasing sequence $A_1 \supset A_2 \supset \dots \supset A_n \supset \dots$ of elements of $\mathcal{X}$ has an empty intersection, one has $\lim_{n\to\infty} \mu(A_n) = 0$. One can show that the countable additivity axiom

is equivalent to the continuity axiom and also to each of the following properties: ainsi qu'à chacune des propriétés suivantes :

1) For any increasing sequence A_i of elements of $\mathcal{X}$, $\mu(\cup A_i) = \lim \mu(A_i)$.

2) For any decreasing sequence A_i of elements of $\mathcal{X}$, $\mu(\cap A_i) = \lim \mu(A_i)$.

Remark. A convenient notation used by Kolmogorov is $\sum C_i$ for disjoint unions.

The basic theorem is the following (see [B1]):

Theorem 11 *A function μ defined on an algebra $\mathcal{G}$ of subsets of a set X, which is σ-additive and has total mass 1, admits a unique extension to a probability measure defined on the σ-algebra $\mathcal{X}$ generated by $\mathcal{G}$.*

A supersonic overflight of the proof (see [B1] section 3): one starts by defining an *outer measure* μ^* which, to each subset A of X associates

$$\mu^*(A) = \inf \sum_n \mu(A_n),$$

where the inf is taken on the set of sequences $A_1, A_2, \ldots$ of elements of $\mathcal{G}$ such that $A \subset \cup_n A_n$. One gets an *inner measure* μ_* by going to the complement :

$$\mu_*(A) = 1 - \mu^*(A^c),$$

where $A^c = X \setminus A$ is the complement of A. One then shows that if $A \in \mathcal{X}$, one has for any subset E of X the equality

$$\mu^*(A \cap E) + \mu^*(A^c \cap E) = \mu^*(E).$$

One then deduces that μ^* is a probability measure on $\mathcal{X}$ and one checks easily that $\mu^*(A) = \mu(A)$ if $A \in \mathcal{G}$. Unicity is a consequence (with some work) of the minimality of $\mathcal{X}$.

CAUTION (see [B1] section 2) : a transfinite induction is necessary to pass from an algebra to the σ-algebra it generates. In particular, a tribe is either finite or it has at least the cardinal of the continuum. This suggest the difficulty inherent in the explicitation of general Borelian sets.

Lebesgue measure on the interval $[0,1]$. The basic example of a probability measure is the interval $[0,1]$ endowed with the *Borelian tribe* and the *Lebesgue measure*. It is convenient to define the Borelian tribe $\mathcal{B}$ as the σ-algebra generated by the algebra $\mathcal{I}$ formed by the finite unions of disjoint intervals $[x_i, y_i[\subset [0,1]$. The Lebesgue measure λ is the unique extension of the length of the intervals: $\lambda([x,y[) = |y-x|$. From the proof of theorem 11, one sees that the measure of a Borelian B can be defined as the inf of the sum $\sum_n |y_n - x_n|$ of the lengths of a sequence of intervals $[x_n, y_n[$ whose union contains B. One can extend λ to a larger tribe but one can show that, if one accepts the axiom of choice and the continuum hypothesis, one cannot define on the tribe of all subsets of $[0,1]$ a probability measure μ such that each singleton $\{x\}$ has probability $\mu(\{x\}) = 0$. Nevertheless, one extension is compulsory,

the extension of λ to all *negligible* sets, that is to all parts of $[0,1]$ which may be covered by a sequence of intervals whose sum of the lengths is arbitraryly small: one decides to give them the measure 0 even if they do not belong to the Borelian tribe. The countable subsets are negligible but there are many others .....

Remark. We shall see in the sequel (Proposition 14) that the probability space $([0,1],\mathcal{B},\lambda)$ is in a precise sense equivalent to a fair game of heads and tails with an infinite number of coin tosses.

Remark. To check directly a property on an arbitrary element of a tribe $\mathcal{X}$ may be difficult while checking it on a set of generators (for example the elements of an algebra $\mathcal{G}$ such that $\mathcal{X}=\sigma(\mathcal{G})$) may reveal much simpler. The following lemma allows often to restrict to such a check (the role of *symmetric difference* $A\Delta A_0=(A\cap A_0^c)\cup(A^c\cap A_0)$ in the notion of almost everywhere invariance is exposed in section 5.2):

Lemma 12 *Let $(X,\mathcal{X},\mu)$ be a probability space and let $\mathcal{G}$ be an algebra generating $\mathcal{X}$ (i.e. $\mathcal{X}=\sigma(\mathcal{G})$). For any $A\in\mathcal{X}$ and $\epsilon>0$, there exists $A_0\in\mathcal{G}$ such that $\mu(A\Delta A_0)<\epsilon$.*

In other words, any element of the tribe is arbitrarily well approximated by an element of the algebra.

Sketch of proof. It is a direct consequence of the construction of the extension of μ from $\mathcal{G}$ to $\sigma(\mathcal{X})$: by definition of $\mu^*(A)$ as the inf of the sums $\sum_{n=1}^{\infty}\mu(A_n)$ on the set of sequences of elements of $\mathcal{G}$ whose union contains A (see the sketch of proof of theorem 11), there exists, for all $\epsilon>0$, a sequence C_n of elements of $\mathcal{G}$ such that $A\subset\cup_{i=n}^{\infty}C_n$ and $\sum_{i=n}^{\infty}\mu(C_n)<\mu(A)+\frac{1}{2}\epsilon$. As the sequence $\mu(C_n)$ converges, there exists an integer N such that $\sum_{n=N+1}^{\infty}C_n<\frac{1}{2}\epsilon$. One can then show (exercise) that $A_0=\sum_{i=1}^{N}C_n$ is a suitable choice.

2.2 The game of "heads or tails" as a stochastic process

Let $X=\{0,1\}^{\mathbb{N}^*}$ be the set of *infinite* sequences

$$\omega=a_1a_2\dots$$

of 0's and 1's. As above, each such sequence can be thought of as an *infinite* sequence of "independent" coin tosses in a "heads or tails" game. It is the realization of a *stationary stochastic process* without memory: "stationary" means that the probability p that $a_i=0$ and the probability $q=1-p$ that $a_i=1$ are independent of the "time" i of the coin toss ; the independence (or absence of memory) means that the probability of a *cylinder*

$$A_{i_1i_2\dots i_k}^{j_1j_2\dots j_k}=\left\{\omega\in X;a_{i_1}=j_1,a_{i_2}=j_2,\dots,a_{i_k}=j_k\right\},i_1,\dots\in\mathbb{N}^*,\ j_1,\dots\in\{0,1\},$$

is

$$\mu(A_{i_1i_2\dots i_k}^{j_1j_2\dots j_k})=\mu(A_{i_1}^{j_1})\mu(A_{i_2}^{j_2})\dots\mu(A_{i_k}^{j_k}),$$

that is $p^{k_0}q^{k_1}$ if the sequence $j_1j_2\dots j_k$ contains k_0 terms equal to 0 and $k_1=k-k_0$ terms equal to 1.

Exercise 14 *1) A finite intersersection of cylinders is still a cylinder;*

2) the complement of a cylinder is a disjoint union of a finite number of cylinders ;

3) a finite union of cylinders may also be written as a finite union union of disjoint cylinders ;

4) deduce from 1),2),3) that the finite unions of disjoint cylinders form an algebra $\mathcal{G}$ of subsets of X (compare to the algebra of finite unions of disjoint intervals $[a_i, b_i[$ of $[0,1]$).

It is natural to define the tribe $\mathcal{X}$ as the one generated by the algebra $\mathcal{G}$ of finite unions of cylinders. One says that the probability measure $\mu = \mu_{p,q}$ whose value on the cylinders was just given is the *product* of an infinity of copies of the measure (p, q) on $\{0, 1\}$.

Apart from the countable unions of disjoint cylinders, to produce non trivial elements of $\mathcal{X}$ is not so easy. In fact, we shall show that the problem is the same as the one of producing a non trivial *Borelian* of the interval $[0, 1] \subset \mathbb{R}$. The tribe $\mathcal{X}$ is indeed the *Borelian tribe* for the topology on X generated by the cylinders, that is the *infinite product* topology (see exercise 15).

Finally, the probability of an element of $\mathcal{X}$ is defined as the unique extension of the probability we have defined for cylinders, in exactly the same way as the mesure of Borelians of $[0, 1]$ is deduced from the measure (length) of intervals.

Exercise 15 (the topological space $\{0,1\}^{\mathbb{N}^*}$ as a Cantor set). *One endows $\{0,1\}^{\mathbb{N}^*}$ with the product topology: a basis of open sets is formed by the cylinders. In other words, an open set is an arbitrary union of cylinders. Another definition is via the introduction of the distance $d(a_1a_2\ldots, b_1b_2\ldots) = \sum_{k=1}^{\infty} \frac{|a_k - b_k|}{2^k}$. Show that the map*

$$f_3 : \{0,1\}^{\mathbb{N}^*} \to [0,1], \quad f_3(a_1a_2\ldots a_n\ldots) = \sum_{k=1}^{\infty} \frac{2a_k}{3^k}$$

is a homeomorphism from $\{0,1\}^{\mathbb{N}^*}$ to the standard triadic Cantor set K. Show that K is of zero Lebesgue measure.

From $\{0,1\}^{\mathbb{N}^*}$ to the interval $[0,1]$: Let us now consider the map

$$f_2 : \{0,1\}^{\mathbb{N}^*} \to [0,1], \quad f_2(a_1a_2\ldots a_n\ldots) = \sum_{k=1}^{\infty} \frac{a_k}{2^k}.$$

As any element of $[0, 1]$ possesses a *dyadic expansion* , this map is surjective. It is not injective: the inverse image of $\frac{1}{2}$ consists in $1000\ldots$ and $0111\ldots$, and same non-unicity phenomenon of the dyadic expansion occurs on the countable dense set of *dyadic numbers*, of the form $\frac{m}{2^k}$ where m and k are integers.

But, surprisingly, in the case of equiprobability ($p = q = 1/2$) i.e. fair coin toss, f_2 is as good as a bijection from the measure point of view. Making precise this assertion requires the introduction of some definitions.

2.3 Measurability and preservation of measure

Measurable maps play the part of arrows in the *category* whose objects are measure spaces:

Definition 9 *A map $f : (X, \mathcal{X}) \to (Y, \mathcal{Y})$ from a measure space to another one is said to be* measurable *if the inverse image $f^{-1}(B)$ of an element of $\mathcal{Y}$ is an element of $\mathcal{X}$.*

Notice that *the smaller the left tribe $\mathcal{X}$, the more demanding a property is mesurability.* If for example $\mathcal{X}$ is defined by a finite partition and if $(Y, \mathcal{Y})$ is $\mathbb{R}$ with its Borelian tribe, the only measurable maps are the ones which are constant on each piece of the partition.

Definition 10 *Given a measurable map $f : (X, \mathcal{X}) \to (Y, \mathcal{Y})$ and a (probability) measure μ on $(X, \mathcal{X})$, the direct image $f_*\mu$ of μ is the (probability) measure on $(Y, \mathcal{Y})$ defined by $f_*\mu(B) = \mu\big(f^{-1}(B)\big)$.*

In practice, $f_*\mu$ is characterized by the equality $\int_Y \varphi d(f_*\mu) = \int_X (\varphi \circ f) d\mu$.

Definition 11 *A measurable map $f : (X, \mathcal{X}, \mu) \to (Y, \mathcal{Y}, \nu)$ from a probability space to another one is said to be measure preserving if $f_*\mu = \nu$*

Definition 12 *A measured dynamical system is a quadruple $(X, \mathcal{X}, \mu, T)$, where $(X, \mathcal{X}, \mu)$ is a probability space and $T : (X, \mathcal{X}, \mu) \to (X, \mathcal{X}, \mu)$ is measurable and measure preserving.*

From now on, when we shall speak of a measure preserving map, we shall imply that it is measurable. In practice, the following lemma is used:

Lemma 13 *If the tribe $\mathcal{Y}$ is generated by the algebra $\mathcal{G}$, in order that the map $f : (X, \mathcal{X}) \to (Y, \mathcal{Y})$ be measurable, it is enough that the inverse images $f^{-1}(G)$ of the elements G of $\mathcal{G}$ belong to $\mathcal{X}$. In the same way, in order that f be measure preserving, it is enough that $\mu(f^{-1}(G)) = \nu(G)$ holds for any element G of the algebra $\mathcal{G}$.*

As a hint of the proof, the reader should meditate on the identity

$$f^{-1}(\sigma(\mathcal{G})) = \sigma(f^{-1}(\mathcal{G})).$$

Exercise 16 *Show that a measurable map $T : (X, \mathcal{X}, \mu) \to (X, \mathcal{X}, \mu)$ from a probability space to itself is measure preserving if and only if it satisfies any one of the following properties (the same notation is used for a function and its class in $L^p(X, \mathcal{X}\mu)$):*

1) for any measurable $f : X \to [0, +\infty[$, $\int_X f \circ T d\mu = \int_X f d\mu$,
2) for any $f \in L^1(\mathcal{X}, \mu)$, $f \circ T \in L^1(\mathcal{X}, \mu)$ and $\int_X f \circ T d\mu = \int_X f d\mu$,
3) for any $f \in L^p(\mathcal{X}, \mu), 1 \leq p < +\infty$, $f \circ T \in L^p(\mathcal{X}, \mu)$ and $||f \circ T||_p = ||f||_p$,
4) for any non-negative $f \in L^\infty(\mathcal{X}, \mu)$, $\int_X f \circ T d\mu = \int_X f d\mu$.

Definition 13 *A map $f : (X, \mathcal{X}, \mu) \to (Y, \mathcal{Y}, \nu)$ from a probability space to another one is an isomorphism of probability spaces if*

1) it is a bijection modulo zero-measure sets (i.e. there exists a μ-negligible $A \subset X$ and a ν-negligible $B \subset Y$ such that f define a bijection from $X \setminus A$ to $Y \setminus B$) (on speaks of a bijection "mod 0")

2) f et f^{-1} are measurable and measure preserving.

We can now make precise the relation between Lebesgue measure on the interval and a fair game of heads and tails alluded to at the end of section 2.1 (such a relation was formulated for the first time by Steinhaus in [St]):

Proposition 14 *The map*

$$f_2 : \left(\{0,1\}^{N^*}, \mathcal{X}, \mu_{\frac{1}{2},\frac{1}{2}}\right) \to \left([0,1], \mathcal{B}, \lambda\right)$$

is an isomorphism of probability spaces: $(f_2)_* \mu_{\frac{1}{2},\frac{1}{2}} = \lambda$ and $(f_2^{-1})_* \lambda = \mu_{\frac{1}{2},\frac{1}{2}}$. In other words, the space of infinite sequences of 0's and 1's endowed with its Borelian tribe and the probability measure corresponding to independent and unbiased coin tosses is, from the point of view of measure theory, equivalent to the interval $[0,1]$ endowed with its Borelian tribe and the Lebesgue measure.

Sketch of proof: we deduce from lemma 13 that it is enough to check measurability and measure preservation by f_2 on intervals and even on intervals of the form $]\frac{p}{2^k}, \frac{p+1}{2^k}]$ which are easily seen to generate the Borelian tribe. But, if $x = \sum_{i=1}^{k} \frac{a_i}{2^i}$ et $y = x + \frac{1}{2^k}$, one checks immediately that $f_2^{-1}[x,y] = A_{12\ldots k}^{a_1 a_2 \ldots a_k}$ and hence that $\mu_{\frac{1}{2},\frac{1}{2}}\left(f_2^{-1}[x,y]\right) = \frac{1}{2^k} = |y - x| = \lambda([x,y])$.

On the other hand, non-injectivity of f_2 holds on a negligible set: indeed, let $\mathcal{D}$ be the subset of $\{0,1\}^{N^*}$ formed by the sequences which after some rank consist only of 1's; $\mathcal{D}$ is contained in a countable union of subsets, each of them contained in a finite union of cylinders whose sum of probabilities may be chosen arbitrarily small (exercise). Its complement $\{0,1\}^{N^*} \setminus \mathcal{D}$ is in bijection with the interval $[0,1[$ obtained by removing a unique point.

Exercise 17 (sequel to exercise 15) *Show that the map $\delta = f_2 \circ f_3^{-1}$ is continuous and surjective from the standard triadic Cantor set K onto the interval $[0,1]$. Show that it takes the same value at the extremities of any interval in $[0,1] \setminus K$. Draw the graph of the unique continuous map from $[0,1]$ onto itself obtained by extending δ by a constant on each connected component of $[0,1] \setminus K$. Check that this graph deserves the name of "devil's stair" given to it by the dynamicists: it is a nice example of function which has bounded variation but is not absolutely continuous (i.e. different from the integral of its derivative, which exists and is Lebesgue-almost everywhere equal to zero).*

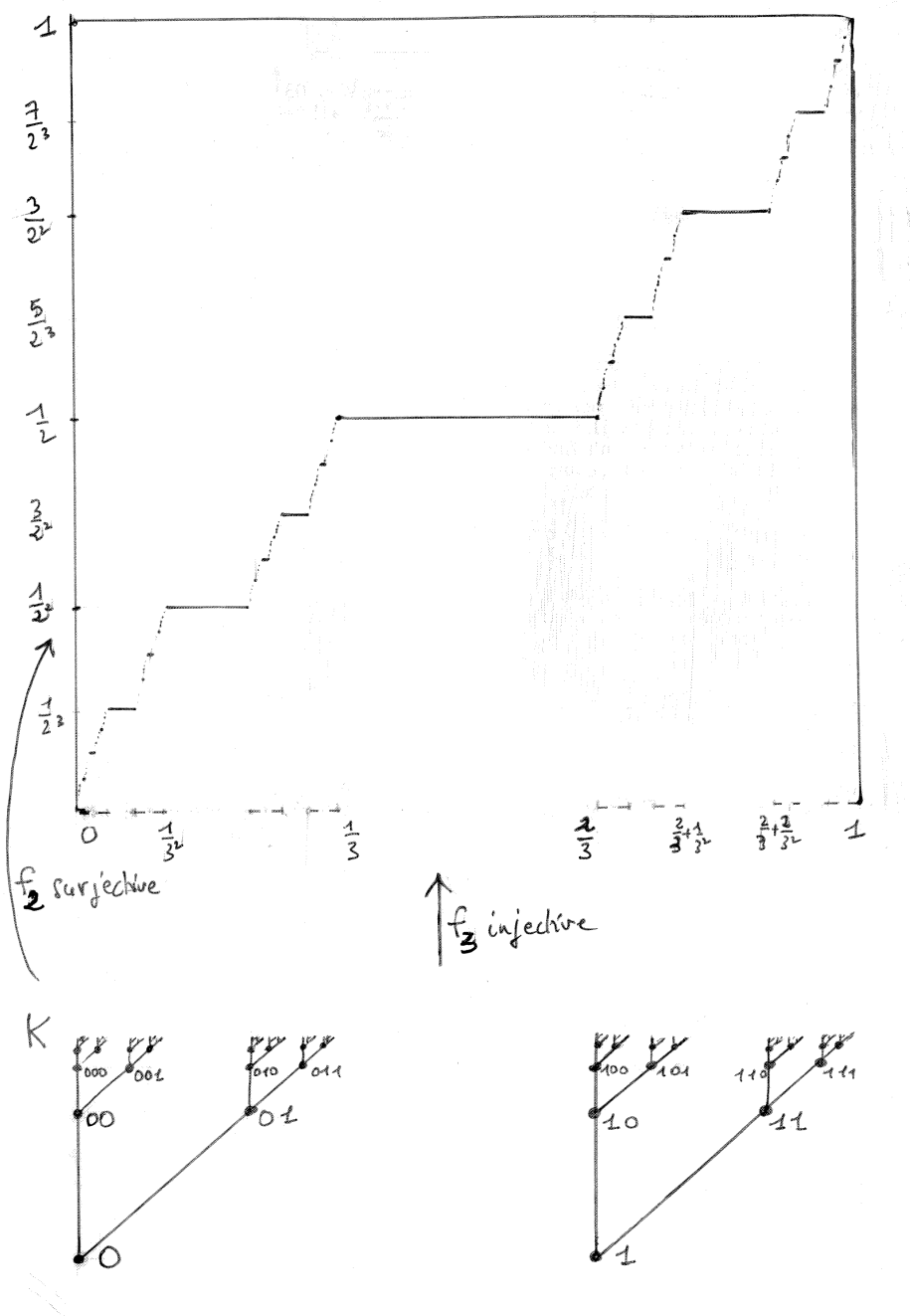


Figure 4. The devil's staircase.

2.4 The game of "heads or tails" as a dynamical system: Bernoulli shifts

The stochastic properties of an infinite sequence of independent coin tosses (with probabilities p and q of 0 and 1 respectively) are nicely reflected in the dynamical properties of a dynamical system, the so-called *Bernoulli shift*, which is the map

$$T : (\{0,1\}^{\mathbb{N}^*}, \mathcal{X}, \mu_{p,q}) \to (\{0,1\}^{\mathbb{N}^*}, \mathcal{X}, \mu_{p,q}), \quad T(a_1a_2a_3\ldots) = (a_2a_3a_4\ldots).$$

As it "forgets" a_1, this map is surjective but not injective: the inverse image of any element contains two elements. The map T is measurable and its fundamental property is the preservation of all the probability measures $\mu = \mu_{p,q}$ on the Borelian tribe $\mathcal{X}$ of $\{0,1\}^{\mathbb{N}^*}$ which are naturally associated to the probabilities (p,q) on $\{0,1\}$. Indeed, the inverse image $T^{-1}(A)$ of the cylinder $A = A^{j_1 j_2 \cdots j_k}_{i_1 i_2 \ldots i_k}$ is the cylinder $A^{j_1 j_2 \cdots j_k}_{i'_1 i'_2 \ldots i'_k}$, where $i'_n = i_n + 1$; hence, the process being stationnary, it has the same probability $p^{k_0}q^{k_1}$ as A. One concludes by applying lemma 13.

Orbits and dynamics. An *orbit* $\{\omega, T\omega, T^2\omega, \ldots, T^n\omega, \ldots\}$ of T is a dynamical description of the sequence ω of coin tosses; the language and the method of the theory of dynamical systems, which consist in treating such an orbit as the discrete version of an integral curve of a differential equation, prove remarkably appropriate and effective in the description of stationary processes of this type.

Exercise 18 *When $p = q = \frac{1}{2}$, the translation of T into the world of the interval $[0,1]$ is the map $x \mapsto 2x(mod.1) = 2x - [2x]$ which is easily seen to preserve Lebesgue measure ($[x]$ is the integer part of x).*

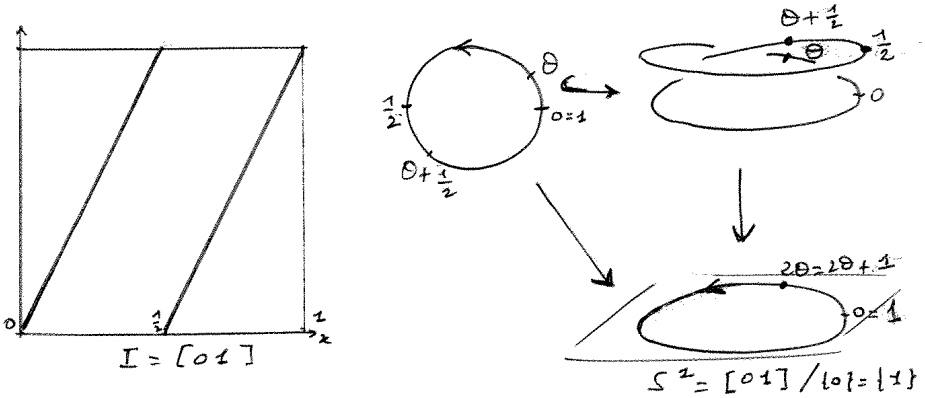


Figure 5 : The map $x \mapsto 2x$ on the interval and on the circle.

From simply infinite to doubly infinite sequences:
It is often more pleasant to work with invertible transformations ; in the case of Bernoulli shifts, this is accomplished by the consideration of an infinite sequence of coin tosses in the past as well as in the future: one defines a bimeasurable bijection $T : \{0,1\}^{\mathbb{Z}} \to \{0,1\}^{\mathbb{Z}}$ which preserves all the probability measures $\mu = \mu_{p,q}$ by the formula

$$T(\ldots a_{-2}a_{-1}a_0a_1a_2\ldots) = (\ldots b_{-2}b_{-1}b_0b_1b_2\ldots), \quad \text{where} \quad b_i = a_{i+1}.$$

Exercise 19 *One supposes that $p = q = \frac{1}{2}$. Show that if $g_2 : \{0,1\}^{\mathbb{Z}} \to [0,1]^2$ is defined by*

$$g_2(\ldots a_{-2}a_{-1}a_0a_1a_2\ldots) = \left(\sum_{k=0}^{-\infty} \frac{a_k}{2^{1-k}}, \sum_{k=1}^{\infty} \frac{a_k}{2^k}\right),$$

the direct image of $\mu_{\frac{1}{2},\frac{1}{2}}$ by g_2 is the Lebesgue measure on $[0,1]^2$ and that g_2 conjugates (mesurably) $T : \{0,1\}^{\mathbb{Z}} \to \{0,1\}^{\mathbb{Z}}$ to the map $\tau : [0,1]^2 \to [0,1]^2$ defined by

$$\tau(x,y) = \big(\frac{1}{2}(x + [2y]), 2y - [2y]\big).$$

Explain why τ is called the "baker transformation".

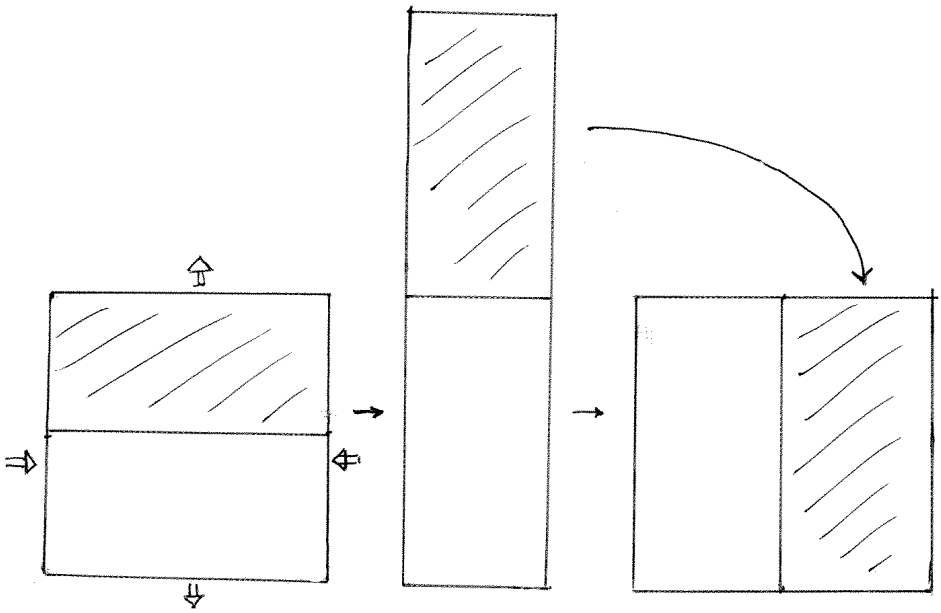


Figure 6 : The baker transformation.

Remark. The existence of the continuous surjective map from the Cantor set $\{0,1\}^{\mathbb{N}^*}$ to the interval $[0,1]$ and the fact that a Cantor set is homeomorphic to any of its finite cartesian powers, imply the existence of generalized Peano curves, that is of continuous surjective maps from $[0,1]$ to $[0,1]^n$ for any $n \in \mathbb{N}$.

3 Existence of invariant probability measures

We are interested in the existence of Borel probability measures which are invariant by a continuous map $T : X \to X$ of a topological space X into itself. Being Borel means that the measure is defined on the *Borelian tribe* of X, that is on the tribe generated by the open subsets of X.

If X is not compact, such a measure may well not exist at all. Think of a translation from $\mathbb{R}$ to $\mathbb{R}$.

When X is compact, we shall use the interpretation of measures on X as *Radon measures* which is given by the *Riesz representation theorem*. Let $C^0(X, \mathbb{C})$ be the space of continuous maps $f : X \to \mathbb{C}$ endowed with the topology of uniform convergence (i.e. the norm $||f|| = \sup_{x \in X} |f(x)|$) and let $C^0(X, \mathbb{C})^*$ be its *topological dual*, that is the set of continuous linear maps $L : C^0(X, \mathbb{C}) \to \mathbb{C}$.

Theorem 15 (Riesz representation theorem) *If X is a compact topological space, the map $\mu \mapsto \left[f \mapsto \int_X f d\mu\right]$ is a bijection from the set of Borel probability measures on X to the set of elements $L \in C^0(X, \mathbb{C})^*$ which are positive (i.e. take positive values on real positive functions) and satisfy $L(1) = 1$.*

We shall endow $C^0(X, \mathbb{C})^*$ with the *weak*topology*, defined as follows: a sequence $(L_n)_{n \geq 0}$ converges to L if $\lim_{n \to \infty} L_n(f) = L(f)$ for every $f \in C^0(X, \mathbb{C})$. This choice is dictated by the following fundamental property: let $||L||$ be the norm on $C^0(X, \mathbb{C})^*$ which defines its strong topology:

$$||L|| = \sup_{f \in C^0(X, \mathbb{C}), ||f|| \leq 1} |L(f)|.$$

Then, the unit ball $\{L \in C^0(X, \mathbb{C})^*, ||L|| \leq 1\}$ is compact for the weak* topology (this is essentially a consequence of Tychonov's theorem which asserts that an arbitrary product of compact sets is compact, see for example Brezis' book on functional analysis). As $\left|\int f d\mu\right| \leq \int |f| d\mu \leq ||f||$, the subset $\mathcal{M}(X)$ of Borel probability measures is contained in this unit ball. As it is closed, it is compact. Moreover, it is convex.

3.1 The theorem of Krylov and Bogoliubov

Theorem 16 *Every continuous map $T : X \to X$ of a compact space X admits an invariant Borel probability measure. The set $\mathcal{M}_T(X)$ of such measures is a compact convex subset of $\mathcal{M}(X)$.*

The compacity hypothesis is necessary, as shown by the example of a translation on the real line.

Proof. We are looking for a fixed point of the map $\mu \mapsto T_* \mu$ from $\mathcal{M}(X)$ to itself. This map is continuous for the weak* topology: if $\lim_{n \to \infty} \mu_n = \mu$ and $f \in C^0(X, \mathbb{C})$,

$$\lim_{n \to \infty} \int f dT_*(\mu_n) = \lim_{n \to \infty} \int f \circ T d\mu_n = \int f \circ T d\mu = \int f dT_*(\mu).$$

Given a probability measure μ, let

$$\mu_n = \frac{1}{n}\sum_{i=0}^{n-1}(T_*)^i\mu,$$

and let μ' be the weak* limit of a subsequence $(\mu_{n_k})_{k\in\mathbb{N}}$. One has

$$||T_*(\mu_{n_k}) - \mu_{n_k}|| = ||\frac{1}{n_k}\big((T_*)^{n_k}\mu - \mu\big)|| \leq \frac{2}{n_k}.$$

On the one hand, the sequence $T_*(\mu_{n_k}) - \mu_{n_k}$ converges weakly to $T_*\mu' - \mu'$, on the other hand it converges strongly, and hence also weakly, to 0. Finally, as T_* is continuous in the weak* topology, $\mathcal{M}_T(X)$ is closed; as T_* is affine (that is $T_*(t\mu + (1-t)\nu) = tT_*(\mu) + (1-t)T_*\nu$ for all μ, ν and all $t \in [0,1]$), $\mathcal{M}_T(X)$ is convex.

Corollary 17 *If $T_1, T_2, \cdots, T_n, \cdots$ are pairwise commuting continuous maps of a compact space X into itself, they possess a common invariant probability measure.*

Proof. The subset $C_1 = \mathcal{M}_{T_1} \subset \mathcal{M}(X)$ of fixed points of $(T_1)_*$ is compact, convex and it is invariant under $T_2, T_3, \cdots$ Applying the same reasoning to $T_2, T_3, \cdots$, leads to a nested sequence of non empty compact convex subsets of $\mathcal{M}(X)$ whose intersection is also non empty, compact and convex.
Notice that in these proofs we have used on the one hand the fact that T_* is an affine map from the compact convex subspace $\mathcal{M}(X) \subset C^0(X,\mathbb{C})^*$ to itself, on the other hand the norm on $C^0(X,\mathbb{C})$. In this abstract setting, the Krylov-Bogoliubov theorem and its corollary appear as a special case of an abstract theorem due to Markov and Kakutani

Theorem 18 (Markov-Kakutani) *Let C be a compact convex subset of a locally convex topological vector space. If $\mathcal{T}_1, \mathcal{T}_2, \cdots$ are pairwise commuting continuous affine maps from C to itself they have a common fixed point.*

The proof is the same except that the use of the norm to prove that the weak limit of the μ_{n_k} is indeed a fixed point is replaced by the use of the Hahn-Banach theorem.

3.2 Back to the examples of section 1

LINEAR MAPS

Exercise 20 *Find the probability measures invariant by a linear map of a finite dimensional vector space. It will be useful to make use of lemma 19:*

MINIMAL ROTATIONS OF TORI

Lemma 19 *If $\alpha = (\alpha_1, \cdots, \alpha_r)$ with $1, \alpha_1, \cdots, \alpha_r$ rationally independent, the Haar measure is the only probability measure which is invariant by $R_{\alpha_1,\cdots,\alpha_r} : \mathbb{T}^r \to \mathbb{T}^r$.*

Proof. The *Haar measure* μ on $\mathbb{T}^r$ is the direct image $\pi_*\lambda$ of the Lebesgue measure on the unit cube $[0,1]^r$. The Riesz representation theorem insures that it is well defined by its values on continuous functions $\varphi : \mathbb{T}^r \to \mathbb{C}$:

$$\int_{\mathbb{T}^r} \varphi d\mu = \int_{[0,1]^r} \varphi \circ \pi \, d\lambda,$$

where $\pi : [0,1]^r \to \mathbb{T}^r = [0,1]^r/\mathbb{Z}^r$ is the quotient projection.

1) As the Lebesgue measure is invariant under any translation of $\mathbb{R}^r$, this measure is invariant under any rotation of $\mathbb{T}^r$ and it is the only one with this property. The proof is an easy exercise which consists in considering finer and finer partitions of $\mathbb{T}^r$ into equal pieces which are obtained from one of them by rotations, for example the half-open cubes

$$C^m_{k_1 \cdot k_r} = \pi\left(\left[\frac{k_1}{2^m}, \frac{k_1+1}{2^m}\right) \times \cdots \times \left[\frac{k_r}{2^m}, \frac{k_r+1}{2^m}\right)\right),$$

and approach uniformly any continuous function $\varphi : \mathbb{T}^r \to \mathbb{C}$ by a family of functions φ_m constant on the pieces of such partitions. One concludes because two rotation-invariant probability measures take the same value on the φ_m.

2) To prove that the Haar measure is the only probability measure which is invariant under the single "irrational" rotation R_α, we shall use theorem 3: let $\varphi : \mathbb{T}^t \to \mathbb{C}$ be a continuous function and let μ be a R_α-invariant measure. For any $\theta \in \mathbb{T}^r$, there exists a sequence $(n_k)_{k\in\mathbb{N}}$, which tends to $+\infty$ such that $\lim_{k\to\infty} R^{n_k}_\alpha = R_\theta$. As φ is uniformly continous, one deduces that

$$\int_{\mathbb{T}^r} \varphi(R_\theta x) d\mu(x) = \lim_{k\to\infty} \int_{\mathbb{T}^r} \varphi(R^{n_k}_\alpha x) d\mu(x)) = \int_{\mathbb{T}^r} \varphi(x) d\mu(x).$$

Hence μ is invariant under the full group of rotations of $\mathbb{T}^r$ and one concludes by 1). We shall come back in section 5.4 to this property, called *unique ergodicity.*

POINCARÉ RETURN MAPS

The notations are those of section 1.4.

Exercise 21 *Show that the map $P : \mathcal{S}' \to \mathcal{S}$ preserves the restriction ϖ to $\mathcal{S}$ of the differential 2-form ω, hence in particular the volume form ϖ^{n-1}.*

BILLIARD MAPS

Exercise 22 *Show that the map $T : S^1 \times [0,\pi]$ defined in section 1.4 preserves the 2-form $\sin\alpha \, d\alpha \wedge dt$.*

Indication. Write $d^2 l(t, t_1) = 0$.

4 Homeomorphisms of the circle

Orientation preserving circle homeomorphisms (and diffeomorphisms) play a central role in the theory of dynamical systems. They appear naturally as return maps on a curve of section for differential equations without singular points on the 2-dimensional torus $\mathbb{T}^2$. Poincaré emphasizes that the problem posed by their study is simpler but reminiscent of problems which arise in Celestial Mechanics.

4.1 Lifting a homeomorphism of the circle to the real line

It is technically more convenient to deal with homeomorphisms of the real line, which means working in the universal cover $D^0(\mathbb{T}^1)$ of the group $Homeo_+(\mathbb{T}^1)$ of orientation preserving homeomorphisms of the circle:

Lemma 20 *Each orientation preserving homeomorphism $F : \mathbb{T}^1 \to \mathbb{T}^1$ lifts to a homeomorphism $f = Id_{\mathbb{R}} + \varphi : \mathbb{R} \to \mathbb{R}$ which is the sum of the Identity and a continuous 1-periodic function $\varphi : \mathbb{R} \to \mathbb{R}$ (which one identifies with a continuous function $\varphi : \mathbb{T}^1 \to \mathbb{R}$). This means that $\pi \circ f = F \circ \pi$, where $\pi : \mathbb{R} \to \mathbb{T}^1$ is the canonical projection. Two such lifts f_1 and f_2 differ by an element $z \in \mathbb{Z}$, that is: $f_2 = f_1 + z$.*

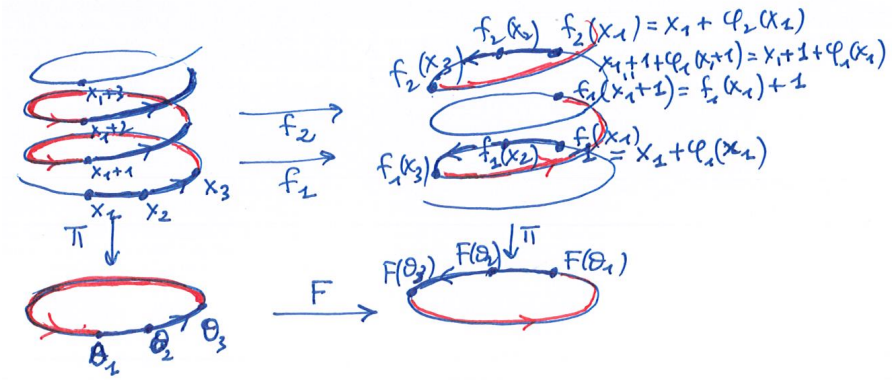


Figure 6bis : Lifting a homeomorphism of the circle.

Proof. The existence of $f : \mathbb{R} \to \mathbb{R}$ such that $F \circ \pi = \pi \circ f$ follows formally from elementary homotopy theory because $\mathbb{R}$ is contractible but it is essentially obvious on the figure. If f is a lift, so is $x \mapsto f_z(x) := f(x + z)$ whatever be $z \in \mathbb{Z}$. Indeed, $\pi \circ f_z(x) = \pi \circ f(x + z) = F \circ \pi(x + z) = F \circ \pi(x)$. Moreover, being in the kernel of π, the difference $f_z - f$ has values in $\mathbb{Z}$ and hence is constant; as F is bijective and orientation preserving, this implies that $f_1(x) = f(x + 1) = f(x) + 1$. One concludes easily.

It follows from this lemma that the universal cover $D^0(\mathbb{T}^1)$ of $Homeo_+(\mathbb{T}^1)$ is

$$D^0(\mathbb{T}^1) = \{f : \mathbb{R} \to \mathbb{R},\ f \text{ increasing homeomorphism},\ f(x + 1) = f(x) + 1\}.$$

We endow $D^0(\mathbb{T}^1)$ with the distance

$$d(f,g) = \max\left(\max_{x\in\mathbb{R}} |f(x)-g(x)|, \max_{x\in\mathbb{R}} |f^{-1}(x)-g^{-1}(x)|\right).$$

Exercise 23 *Show that $D^0(\mathbb{T}^1)$ is a topological group (i.e. that the maps $(f,g)\mapsto f\circ g$ and $f\mapsto f^{-1}$ are continuous) and that it is complete.*

Remark. This lemma generalizes to arbitrary continuous maps and to higher dimensions in the following way:

Proposition 21 *Any continous map $F:\mathbb{T}^r\to\mathbb{T}^r$ admits a lift $f:\mathbb{R}^r\to\mathbb{R}^r$ such that $\pi\circ f = F\circ\pi$. Moreover, there exists a unique linear map $\ell:\mathbb{R}^r\to\mathbb{R}^r$, depending only on F and which is itself the lift of a map $L:\mathbb{T}^r\to\mathbb{T}^r$, such that any lift f of F is of the form $f=\ell+\varphi$, where for any $z\in\mathbb{Z}^r$, $\varphi(x+z)=\varphi(x)$ which means that φ may be identified with a map from $\mathbb{T}^r$ to $\mathbb{R}^r$. In particular, F and L ar homotopic.*

4.2 Poincaré's rotation number

The rotation number of a homeomorphism of the circle was defined in 1885 by Poincaré in the third part of his series of papers on the curves defined by differential equations.

We are interested in the behaviour of $f\in D^0(\mathbb{T}^1)$ under iteration. One sees by induction that the k-th iterate f^k of f can be expressed as a so-called *Birkhoff sum* (see 6)

$$f^n = Id+\varphi_n = Id+\sum_{i=0}^{n-1}\varphi\circ f^i.$$

Theorem 22 *For all $f\in D^0(\mathbb{T}^1)$, the sequence of periodic functions $\frac{1}{n}(f^n-Id)$ converges uniformly when $n\to\infty$ to a real number $\rho(f)\in\mathbb{R}$ which is called the rotation number of f. It follows that $\frac{1}{n}f^n$ converges to $\rho(f)$ uniformly on compact subsets.*

The key to the proof is the following lemma, typical of dimension 1:

Lemma 23 *Let $f=Id+\varphi\in D^0(\mathbb{T}^1)$. Let $m=\min_{x\in\mathbb{R}}\varphi(x)$, $M=\max_{x\in\mathbb{R}}\varphi(x)$. One has*

$$0\le M-m<1.$$

Proof. As φ is defined and continuous on the circle, there exist real numbers x_m and x_M such that

$$\varphi(x_m)=m,\ \varphi(x_M)=M,\quad 0\le x_M-x_m<1.$$

Because f is a homeomorphism which sends an interval of length 1 onto an interval of length 1, one has $f(x_M)-f(x_m)<1$, i.e. $M-m<1-(x_M-x_m)<1$.

Corollary 24 *For all $x \in \mathbb{R}$, the sequence*

$$u_n = f^n(x) - x + 1, \quad n \geq 1,$$

is subadditive, that is: $u_{n+m} \leq u_n + u_m$.

Proof. Applying the lemma to $f^n = Id + \varphi_n \in D^0(\mathbb{T}^1)$ and using the fact that f^n is increasing, we get

$$\forall x, y \in \mathbb{R}, \quad y - x - 1 \leq f^n(y) - f^n(x) \leq y - x + 1.$$

Taking $y = f^m(x)$ we get

$$u_n + u_m - 2 \leq u_{n+m} \leq u_n + u_m.$$

The following lemma is classical and it will be used again in section 6.3.

Lemma 25 *If $(u_n)_{n\geq 1}$ is a subadditive sequence in $\mathbb{R} \cup \{-\infty\}$, the sequence $\left(\frac{u_n}{n}\right)_{n\geq 1}$ converges in $\mathbb{R} \cup \{-\infty\}$ and*

$$\lim_{n\to\infty} \frac{u_n}{n} = \inf_{n\geq 1} \frac{u_n}{n}.$$

Proof. Fix $p \geq 1$. For every $n \geq p$, write $n = kp + r$, $r < p$ and observe that

$$\frac{u_n}{n} \leq \frac{u_{kp} + u_r}{kp + r} \leq \frac{u_{kp}}{kp} + \frac{u_r}{kp + r} \leq \frac{u_p}{p} + \frac{u_r}{n}.$$

This implies that

$$\limsup_{n\to\infty} \frac{u_n}{n} \leq \frac{u_p}{p} \quad \text{and therefore} \quad \limsup_{n\to\infty} \frac{u_n}{n} \leq \inf_{p\geq 1} \frac{u_p}{p} \leq \liminf_{n\to\infty} \frac{u_n}{n}$$

.

Proof of theorem 22. From Corollary 24 and Lemma 25, one deduces that for all $x \in \mathbb{R}$, the sequence $\frac{1}{n}\big(f^n(x) - x\big)$ converges in $\mathbb{R} \cup \{-\infty\}$. Moreover, the inequality used in the proof of the Corollary shows that the limit $\rho(f)$ is independent of x and that the convergence is uniform. Also, from the inequalities $u_n + u_m - 2 \leq u_{n+m} \leq u_n + u_m$, one gets $u_{n-1} + u_1 - 2 \leq u_n \leq u_{n-1} + u_1$ and by induction $nu_1 - 2n \leq u_n \leq nu_1$. It follows that $u_1 - 2 \leq \lim_{n\to\infty} \frac{u_n}{n} \leq u_1$, that is

$$\forall x \in \mathbb{R}, \;\; f(x) - x - 1 \leq \rho(f)) \leq f(x) - x + 1.$$

4.3 Rotation number and invariant measures

Let μ be a probability measure on $\mathbb{T}^1$ which is invariant under the homeomorphism $\overline{f} : \mathbb{T}^1 \to \mathbb{T}^1$ and let $f = Id + \varphi \in D^0(\mathbb{T}^1)$ be a lift of $\overline{f}$. We have, for any $n \in \mathbb{N}$,

$$\mu(f^n - Id - n\mu(\varphi)) = \mu\big(\sum_{i=0}^{n-1} \varphi \circ f^i - n\mu(\varphi)\big) = \mu\big(\sum_{i=0}^{n-1} \varphi \circ \overline{f}^i\big) - n\mu(\varphi) = 0.$$

It follows that the function $f^n - Id - n\mu(\varphi)$ must vanish somewhere. Applying lemma 23 to $f^n \in D^0\mathbb{T}^1$ one gets that

$$\max\left(f^n - Id - n\mu(\varphi)\right) - \min\left(f^n - Id - n\mu(\varphi)\right) < 1,$$

and hence

$$|f^n - Id - n\mu(\varphi)|_{C^0} < 1.$$

This gives another proof of the uniform convergence of the sequence $\frac{1}{n}\big(f^n - Id\big)$ to a constant. Summarizing, we have proved

Proposition 26 *Let $\overline{f} \in Homeo^+(\mathbb{T}^1)$ and let $f = Id + \varphi \in D^0\mathbb{T}^1$ be a lift of $\overline{f}$. The rotation number $\rho(f)$ of f satisfies $\rho(f) = \mu(\varphi)$ for any $\overline{f}$-invariant probability measure on $\mathbb{T}^1$. One has*

$$\begin{cases} \forall n,\, \exists x_n \in \mathbb{R} \quad \text{such that} \quad f^n(x_n) - x_n - n\rho(f) = 0, \\ \forall x \in \mathbb{R},\, \forall n, \;\; -1 < f^n(x) - x - n\rho(f) < 1. \end{cases}$$

Moreover, changing the lift f of $\overline{f}$ does not change the class of $\rho(f)$ in $\mathbb{T}^1 = \mathbb{R}/\mathbb{Z}$. This class is called the rotation number $\rho(\overline{f})$ of $\overline{f}$.

Only the last part concerning the behaviour of $\rho(f)$ under a change of the lift f remains to be proved: for this one notices that, as $f^n(x+k) = f^n(x) + k$, if $g = f + k$, one has $g^n(x) = f^n(x) + nk$ and hence $\rho(g) = \rho(f) + k$. Alternatively, one notices that, as the total mas of μ is 1, $\mu(\varphi + k) = \mu(\varphi) + k$.

Corollary 27 *If $p \in \mathbb{Z}$ and $q \in \mathbb{N}, q \geq 1$,*

$$\begin{cases} \rho(f) = p/q \Longleftrightarrow \exists x_q,\; f^q(x_q) = x_q + p, \\ \rho(f) > p/q \Longleftrightarrow \forall x \in \mathbb{R},\; f^q(x) > x + p, \\ \rho(f) < p/q \Longleftrightarrow \forall x \in \mathbb{R},\; f^q(x) < x + p, \end{cases}$$

Proof. If $f^q(x_q) = x_q + p$, one has also $f^{kq}(x_q) = x_q + kp$ hence $\rho(f) = \lim_{k\to\infty} \frac{1}{kq}(f^{kq}(x_q) - x_q) = p/q$. If $\rho(f) > p/q$, one deduces from proposition 26 that $\forall x \in \mathbb{R},\; f^q(x) > x + q\rho(f) - 1 > x + p - 1$. If for some $x \in \mathbb{R}$ we have $x + p - 1 < f^q(x) < x + p$, the interval $[x, x+1]$ is sent homeomorphically by f^q onto the interval $[f^q(x), f^q(x) + 1]$; hence by the intermediate value theorem, there is some $x_q \in [x, x+1]$ such that $f^q(x_q) = x_q + p$ which implies that $\rho(f) = p/q$, a contradiction.

J.C. Yoccoz commented in a lecture that this Corollary gives a definition of the rotation number in the spirit of the definition of real numbers by Dedekind, while the definition by a limit is more in the spirit of Cauchy.

Lemma 28 *A homeomorphism $f \in D^0(\mathbb{T}^1)$ with rotation number $\rho(f) = p/q$ is conjugate to the translation $R_{p/q}$ if and only if $f^q = R_p$.*

Proof. If $f = h^{-1} \circ R_{p/q} \circ h$ with $h \in D^0(\mathbb{T}^1)$, then $f^q = h^{-1} \circ R_p \circ h = R_p$. Conversely, if $f^q = R_p$, one checks that $h = \frac{1}{q}\sum_{i=0}^{q-1}\left(f^i - i\frac{p}{q}\right)$ belongs to $D^0(\mathbb{T}^1)$ and conjugates f to $R_{p/q}$.

Proposition 29 (Structure implied by a rational rotation number) *Let $\overline{f} \in Homeo_+(\mathbb{T}^1)$ be such that $\rho(\overline{f}) = p/q \in \mathbb{Q}/\mathbb{Z}$ (irreducible). 1)$\overline{f}$ has periodic points of period q and every periodic point of $\overline{f}$ has minimal period q. 2) The limit sets $\alpha(x)$ and $\omega(x)$ of any element $x \in \mathbb{T}^1$ are periodic orbits.*

Proof. Let $f \in D^0(\mathbb{T}^1)$ be the lift of $\overline{f}$ whose rotation number is $p/q \in \mathbb{R}$. It follows from Corollary 27 that $f^q - p$ has a fixed point x_q, and hence that $\overline{f}^q$ has a fixed point. If $\overline{x}_{q'} \in \mathbb{T}^1$ is a periodic point of $\overline{f}$ of period q', it lifts to $x_{q'} \in \mathbb{R}$ such that $f^{q'}(x_{q'}) = x_{q'} + p'$; this implies that $\rho(f) = p'/q' = p/q$, hence that $q' = kq$ which shows that $x_{q'}$ is a periodic point of $g = f^q - p$ of period k. Now, the structure of elements $g \in D^0(\mathbb{T}^1)$ whose rotation number is 0 is easily understood: the set $Fix(g)$ of fixed points is closed and invariant under integer translations. If $]a, b[$ is a connected component of $\mathbb{R} \setminus Fix(g)$, one deduces from the fact that g is increasing that if $x \in]a, b[$, $\alpha(x) = a$ and $\omega(x) = b$ (resp. $\alpha(x) = b$ and $\omega(x) = a$) if $g - Id$ is positive (resp. negative) in the interval. This implies that the only periodic points are fixed points and proves also the last part of the proposition.

Proposition 30 (Invariance under semi-conjugation) *Let $f, g \in D^0(\mathbb{T}^1)$ such that there exists a continuous map $h = Id + \varphi \in C^0(\mathbb{T}^1, \mathbb{R})$ (i.e. φ continuous and 1-periodic) satisfying $h \circ f = g \circ h$ (one says that f and g are semi-conjugated), then $\rho(f) = \rho(g)$.*

Proof. For any n, $h \circ f^n = g^n \circ h$, hence $f^n + \varphi \circ f^n = g^n \circ h$ and

$$\frac{1}{n}\big(f^n - Id\big) + \frac{\varphi \circ f^n}{n} = \frac{1}{n}(g^n - Id) \circ h + \frac{\varphi}{n},$$

hence the conclusion because φ is bounded. In particular, if f and g are *conjugated* by $h \in D^0(\mathbb{T}^1)$, i.e. if $g = h \circ f \circ h^{-1}$, they have the same rotation number $\rho(f) = \rho(g)$.

Exercise 24 *If $\overline{f}$ and $\overline{g}$ are semi-conjugated in $Homeo^+(\mathbb{T}^1)$, that is if $\overline{g} = \overline{h} \circ \overline{f} \circ \overline{h}^{-1}$, there exist lifts f, g, h to $D^0(\mathbb{T}^1)$ such that $h \circ f = g \circ h$.*

Proposition 31 *If $f; g \in D^0(\mathbb{T}^1)$ commute, then $\rho(f \circ g) = \rho(f) + \rho(g)$.*

Proof. It follows from theorem 18 that there exists a probability measure on $\mathbb{T}^1$ which is invariant by both $\overline{f}$ and $\overline{g}$. If $f = Id + \varphi$ and $g = Id + \psi$, one has $f \circ g = Id + \psi + \varphi \circ g$, hence $\rho(f \circ g) = \mu(\psi + \varphi \circ g) = \mu(\psi) + \mu(\varphi) = \rho(f) + \rho(g)$.

Exercise 25 *Show that if $f, g \in D^0(\mathbb{T}^1)$ are lifts of two commuting elements $\overline{f}, \overline{g} \in Homeo^+(\mathbb{T}^1)$, they also commute.*
Hint: use that, if μ is a probability measure leaving both $\overline{f}$ and $\overline{g}$ invariant, $\mu(f \circ g - Id) = \mu(g \circ f - Id)$.

Proposition 32 (Structure implied by an irrational rotation number) *Let $\overline{f} \in Homeo_+(\mathbb{T}^1)$ be such that $\rho(\overline{f}) \in (\mathbb{R}\setminus\mathbb{Q})/\mathbb{Z}$. 1) There exists a surjective continuous map $\overline{h} : \mathbb{T}^1 \to \mathbb{T}^1$ such that $\overline{h}\circ\overline{f} = R_{\rho(\overline{f})}\circ\overline{h}$, i.e. $\overline{f}$ is semi-conjugated to the corresponding rotation. 2) If $\overline{f}$ is not actually conjugated to the corresponding rotation, there exists an invariant Cantor set $X \subset \mathbb{T}^1$ which is the unique closed invariant minimal (for the inclusion) set. 3) Moreover X is at the same time the set $\Omega(\overline{f})$ of non wandering points and the α-limit set $\alpha(x)$ and the ω-limit set $\omega(x)$ of every $x \in \mathbb{T}^1$.*

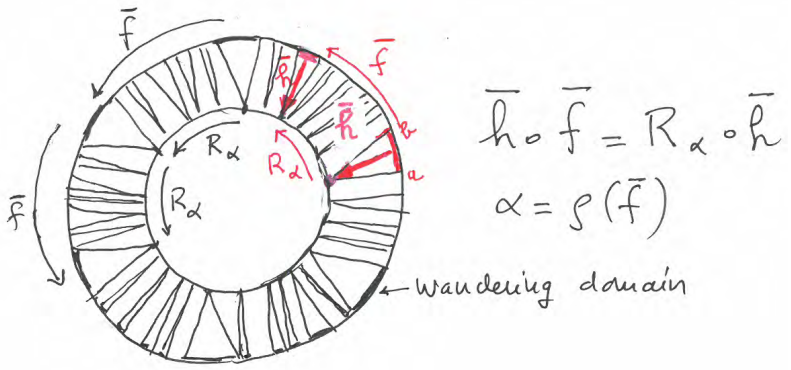


Figure 7 : Typical behavior with irrational rotation number.

Proof. 1) Let μ be a $\overline{f}$-invariant probability measure on $\mathbb{T}^1$. We shall still use the notation μ for its lift to a positive Borel measure on $\mathbb{R}$ invariant under integer translations (exercise: construct the lift); let $f \in D^0(\mathbb{T}^1)$ be a lift of $\overline{f}$. Let $h(x) = \mu([0, x[)$. As $\overline{f}$ has no periodic point, μ has no atomic mass hence $h : \mathbb{R} \to \mathbb{R}$ is a continuous non decreasing function such that $h(x+n) = h(x)+n$ for any $n \in \mathbb{Z}$. Hence it defines a continuous surjective map $\overline{h} : \mathbb{T}^1 \to \mathbb{T}^1$. The $\overline{f}$-invariance of μ implies that $h(f(x)) - h(f(0)) = h(x) - h(0)$, that is $h \circ f = R_{h(f(0))-h(0)} \circ h$. Finally, proposition 30 insures that $h(f(0)) - h(0) = \rho(f)$.

2) The map h is a homeomorphism if it is strictly increasing, that is if the support X of μ is the whole circle $\mathbb{T}^1$. If not, X is a closed invariant set without isolated point (because μ has no atomic mass) whose image by $\overline{h}$ is $\mathbb{T}^1$; moreover, the restriction $\overline{h}|_X$ is injective except on the countable subset $D \subset X$ formed by the extremities of the connected components of the complement $\mathbb{T}^1 \subset X$. Let $M \subset \mathbb{T}^1$ be a non empty closed $\overline{f}$-invariant set. Because $\overline{h}$ is a semi-conjugation of $\overline{f}$ to the rotation $R_{\rho(\overline{f})}$, $\overline{h}(M)$ is invariant under this rotation, hence $\overline{h}(M) = \mathbb{T}^1$; as $\overline{h}$ is injective on $X \setminus D$ and non decreasing, this implies that M must contain $X \setminus D$. As X has no isolated point, the closure of $X \setminus D$ is X, hence $M \supset X$ which proves that X is the unique $\overline{f}$-invariant minimal closed set. It follows that X has no interior (otherwise its boundary would be invariant, contradicting minimality of X), hence X is Cantor set.

3) Let I_0 be a connected component of $\mathbb{T}^1 \setminus X$. Its iterates $\overline{f}^n(I_0)$ are also connected components of $\mathbb{T}^1 \setminus X$ and hence are two by two disjoint because $\overline{f}$ has no periodic point. This means that any $x \in \mathbb{T}^1 \setminus X$ is wandering, hence that

$\Omega(\overline{f})$ is contained in X and hence equal to it because X is minimal. Finally, being minimal, $\Omega(\overline{f})$ coincides with $\omega(x)$ and $\alpha(x)$ for any $x \in \mathbb{T}^1$.

The behavior depicted in Proposition 32, is typical for homeomorphisms and can occur even for some C^1 diffeomorphisms (Denjoy examples) but it cannot occur as soon as a litte more regularity of the map is granted. More precisely:

Theorem 33 (Denjoy) *A C^1 diffeomorphism of the circle $\mathbb{R}/\mathbb{Z}$ whose derivative has bounded variation, and whose rotation number is irrational, is topologically conjugate to the corresponding rotation. In particular, it has no wandering domains.*

A fundamental strengthening of this theorem has been given in Herman's thesis [He1] with further development by Yoccoz: under an appropriate diophantine hypothesis on the rotation number, the conjugacy is C^{r-2} (resp. C^∞, resp. analytic) if f is $C^r, r \geq 3$, resp. C^∞, resp. analytic). A local version, for f close to the corresponding rotation had been given first by Arnold in [A].

Again the devil staircase: 1-parameter families:

Proposition 34 *The map $f \mapsto \rho(f)$ is non decreasing and continuous.*

Proof. Both properties are easy consequences of the equality $\rho(Id + \varphi) = \mu(\varphi)$.

Exercise 26 (Arnold's family, see [A]) *For $\alpha \in [0, 1/2\pi]$, let*

$$f_t(x) = x + a\sin(2\pi x) + t.$$

Show that the graph of the map $t \mapsto \rho(f_t)$ is a devil staircase (see figure 4). This example is a good illustration of the types of dynamics encountered in a "generic" family of analytic diffeomorphisms. For a fixed value of a, the set of t for which the rotation number of f_t is rational is big in the sense of topology, namely it is open and dense, but its complement is big in the sense of measure, namely, its measure tends to 1 when $a \to 0$.

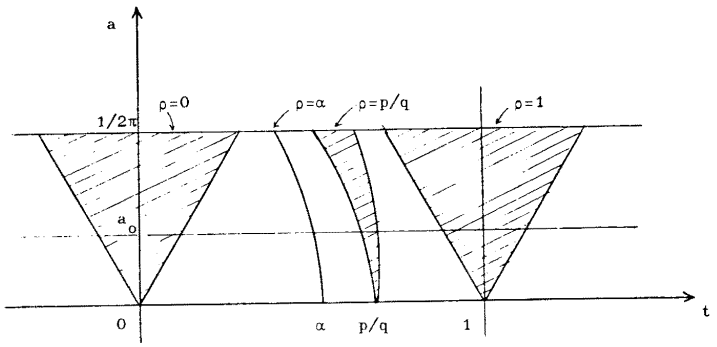


Figure 8 : Typical behavior of a family of diffeomorphisms.

5 Introduction to ergodic theory

Ergodic theory originates in the works of Boltzmann on statistical mechanics. Its mathematical form was shaped in the years 1930, with Von Neumann's and Birkhoff 's theorems, which are strong dynamical versions of the law of large numbers, but one can trace it back to Poincaré's *recurrence theorem* [Po] where the constraints imposed on dynamics by the preservation of a measure whose total mass is finite were already exploited in a subtle way.

5.1 Poincaré recurrence theorem

This theorem, which was proved by Poincaré in his book "The New Methods of Celestial Mechanics", may be considered as the birth of ergodic theory, that is the probabilistic side of the theory of dynamical systems. After the discovery of an important mistake in the first version of his Memoir on the Three Body problem (the one which won the prize of the King of Sweden), Poincaré put a great emphasis on this theorem because it gave him an ersatz, which he calls ""stability according to Poisson", of the stability result he thought he had proved: the result was the fact that the semi-major axes of the approximate Keplerian ellipses described by the planets come back indefinitely arbitrarily close to their initial values.

Theorem 35 *Let $(X, \mathcal{X}, \mu, T)$ be a measured dynamical system. For $A \in \mathcal{X}$, let $\tilde{A} \subset A$ be the subset formed by the elements $x \in A$ whose orbit $\{T^n(x)\}_{n\in\mathbb{N}}$ comes back to A an infinite number of times. Then $\tilde{A} \in \mathcal{X}$ and $\mu(\tilde{A}) = \mu(A)$.*

Proof. Let $A_n = A \setminus \cup_{i\geq n} T^{-i}(A)\big)$ be the set of $x \in A$ whose iterates of rank higher than n never come back to A. As $A_n \subset A$, we have $T^{-i}(A_n) \subset T^{-i}(A)$, hence $T^{-i}(A_n) \cap A_n = \emptyset$ for all $i \geq n$. This implies that for any $k > l \geq 0$, we have $\emptyset = T^{-nl}\big(A_n \cap T^{-n(k-l)}(A_n)\big) = T^{-nk}(A_n) \cap T^{-nl}(A_n)$. In other words, the $T^{-nk}(A_n), k \in \mathbb{N}$, are pairwise disjoint. As they all have the same measure $\mu(A_n)$, we conclude that $\mu(A_n) = 0$. As $A \setminus \tilde{A} = \cup_{n\geq 1} A_n.$, this ends the proof.

Remarks. 1) The recurrence times (i.e. number of iterations) are very far from being uniform. For an illustration, see theorem 44.

2) These recurrence times may be extremely long and this already cuts short many philosophical discussions about the relation of this theorem to the second principle of thermodynamics.

5.2 Invariant sets, invariant functions

In the category of probability spaces and measure preserving maps, the pertinent notion of invariance is *almost everywhere invariance*:

Definition 14 (symmetric difference) *Given two subsets U, V of a set X, their symmetric difference $U\Delta V$ is*

$$U\Delta V = (U \cap V^c) \cup (V \cap U^c).$$

Exercise 27 *1) Show that for any three subsets U, V, W of X, one has $U\Delta V \subset (U\Delta W) \cup (W\Delta V)$ and hence $\mu(U\Delta V) \leq \mu(U\Delta W) + \mu(W\Delta V)$.*

2) Show that if $\mu(A\Delta A_0) < \epsilon$ and $\mu(B\Delta B_0) < \epsilon$, one has

$$\mu\big((A \cap B)\Delta(A_0 \cap B_0)\big) < 2\epsilon \quad \textit{and} \quad \mu\big((A\Delta B)\Delta(A_0\Delta B_0)\big) < 2\epsilon.$$

Definition 15 *Given a measure space $(X, \mathcal{A}, \mu)$, two elements $U, V \in \mathcal{A}$ are said to be almost everywhere equal (a.e.=) if $\mu(U\Delta V) = 0$.*

Definition 16 *Let $(X, \mathcal{X}, \mu, T)$ be a measured dynamical system (recall definition 12). One says that $A \in \mathcal{X}$ is T-invariant a.e. (or invariant by T a.e.) if $T^{-1}(A)$ is almost everywhere equal to A. In case there is no ambiguity, one simply says that A is T-invariant (or invariant by T).*

One can show that if A is invariant in this sense, there exists B invariant in the strict sense, i.e. $\mu(A\Delta B) = 0$ and $T^{-1}(B) = B$, which coincides almost everywhere with A.

Definition 17 (The (almost) invariant tribe) *Let $(X, \mathcal{X}, \mu, T)$ be a measured dynamical system. The set of $A \in \mathcal{X}$ which are a.e. invariant by T is a tribe (in the sense of definition 7) $\mathcal{I} = \mathcal{I}(T)$, called the (almost) invariant tribe of the dynamical system.*

The subset A is a.e. invariant if and only if the characteristic function $\mathcal{X}_A$ of A is almost everywhere equal to the characteristic function $\mathcal{X}_{T^{-1}A}$ of $T^{-1}A$. More generally

Exercise 28 *Let $(X, \mathcal{X}, \mu, T)$ be a measured dynamical system. If the function $f : X \to \mathbb{C}$ is $\mathcal{X}$ measurable, the following properties are equivalent:*

1) $f \circ T = f$ a.e. , i.e. $\mu\big(\{x \in X,\ f \circ T(x) \neq f(x)\}\big) = 0$;

2) f is $\mathcal{I}(T)$-measurable.

Hint. 1) If B is a borelian, show that

$$T^{-1} \circ f^{-1}(B) \subset f^{-1}(B) \cup \{y \in X,\ f \circ T(y) \neq f(y)\};$$

2) Recall that a measurable function is always a simple limit of functions which are finite sums of characteristic functions of measurable sets.

Notice that one can replace f by a function which is defined only on a subset of X whose complement has measure 0. In fact, as in the case of subsets, one can show that if a $\mathcal{A}$-measurable function f satisfies $f \circ T = f$ a.e., there exists a $\mathcal{X}$-measurable function g such that $g \circ T = g$ and $g = f$ a.e.

5.3 Ergodicity

Independence of the coin tosses in a "heads or tails" game implies "forgetting of the initial condition": each toss ignores the result of all the former tosses; to this corresponds a very strong property of the Bernoulli shifts, called *ergodicity*:

Definition 18 (Ergodicity) *Let $(X, \mathcal{X}, \mu, T)$ be a measured dynamical system. One says that T (or that the dynamical system) is ergodic if every set $A \in \mathcal{X}$ which is invariant by T satisfies $\mu(A) = 0$ or $\mu(A) = 1$. When T is given, one says also that the invariant measure μ is ergodic.*

Exercise 29 *Show that the map T is ergodic if and only if any one of the following properties is satisfied:*

1) every measurable T-invariant function $f : X \to \mathbb{C}$ is a.e. constant;

2) There exists $p \geq 1$, such that every T-invariant function $f \in L^p(X, \mathbb{C})$ is a.e. constant.

Remark: an abstract characterization in the context of theorem 16. If T is a continuous map from a compact space into intself, one shows that the ergodic measures are precisely the *extremal points* of the compact convex set $\mathcal{M}_T(X)$ of T-invariant probability measures. In particular, such measures always exist.

The analogue of theorem 3 in the world of measured dynamical systems is

Theorem 36 *Let $\alpha = (\alpha_1, \alpha_2, \cdots, \alpha_r) \in \mathbb{R}^r$. The real numbers $1, \alpha_1, \cdots, \alpha_r$ are rationally independent, i.e. there is no $(r+1)$-tuple $(k_0, k_1, \cdots, k_r) \in \mathbb{Z}^{r+1} \setminus \{0\}$ such that $k_0 + k_1\alpha_1 + \cdots + k_r\alpha_r = 0$, if and only if the rotation $R_\alpha : \mathbb{T}^r \to \mathbb{T}^r$ is ergodic.*

Proof. A nice proof based on density of the orbits of such a rotation can be found in Billingsley's book. Here is a simpler but maybe less transparent proof, in fact exactly the same as the one given of topological transitivity (Theorem 3): by exercise 29, it is enough to prove that, under the hypothesis on α, every R_α-invariant function in $L^2(\mathbb{T}^r, \mathcal{B}, Haar)$ is a.e. constant. Such a function, considered as $\mathbb{Z}^r$-periodic function on $\mathbb{R}^r$, admits a Fourier expansion

$$f(x) = \sum_{k \in \mathbb{Z}^r} c_k e^{2i\pi < k, x >}$$

and computing $f(R_\alpha x) = \sum_{k \in \mathbb{Z}^r} c_k e^{2i\pi < k, \alpha >} e^{2i\pi < k, x >}$, one gets the equations

$$\forall k \in \mathbb{Z}^r, \quad c_k e^{2i\pi < k, \alpha >} = c_k.$$

The hypothesis on α being equivalent to $e^{2i\pi < k, \alpha >} \neq 1$ if $k \neq 0$, all the c_k except c_0 are equal to 0.

We end this section with a theorem which makes a link between ergodicity and topological transitivity. As a corollary, we recover Theorem 6.

Proposition 37 *Let $(X, \mathcal{B}, \mu, T)$ be a measured dynamical system such that*

1) X is a metric space with a countable basis of open sets and $T : X \to X$ is continuous ;

2) $\mathcal{B}$ is the Borel tribe and if U is a non empty open set, $\mu(U) > 0$;

3) μ is ergodic.

Then, T is positively topologically transitive; more precisely, for μ almost every $x \in X$, the closure $\overline{\mathcal{O}_T^+}$ of the positive orbit of x coincides with X.

Proof. Given two non empty open subsets U, V of X we need show that $V \cap O \neq \emptyset$, where $O = (\cup_{n\in\mathbb{N}} T^{-n}U)$ and for this it is enough to show that $\mu(O) = 1$. This is obvious because, on the one hand, as $U \subset O$, $\mu(O) \geq \mu(U) > 0$, on the other hand $T^{-1}O \subset O$ hence, as both sets have the same measure, O is a.e. invariant; as T is ergodic, $\mu(O) = 1$.
Now, if $(U_i)_{i\in\mathbb{N}}$ is a countable basis of open sets and $O_i = \cup_n T^{-n}U_i$, the set of points whose positive orbit is dense is precisely $\cap_{i\in\mathbb{N}} O_i$ (exercise). As a countable intersection of subsets of measure 1, this set has measure 1.

5.4 Unique ergodicity

Let $T : X \to X$ be a continuous map of a compact topological space into itself.

Proposition 38 *The following conditions are equivalent:*

i) There is a unique probability measure on X which is invariant under T.

ii) For any continuous function $f : X \to \mathbb{C}$, the sequence $\frac{1}{n}\sum_{i=0}^{n-1} f \circ T^i$ converges uniformly to a constant function.

Proof. Suppose there is a unique T-invariant probability measure μ and suppose by contradiction that there is a real number $\epsilon > 0$, a sequence $\left(n_k\right)_{k\geq 0}$ of integers tending to $+\infty$ and a sequence $\left(x_k\right)_{k\geq 0}$ of points of X such that for all k,

$$\left| \frac{1}{n_k} \sum_{i=0}^{n_k-1} f \circ T^i(x_k) - \int f \, d\mu \right| > \epsilon.$$

Let $\mu_k = \frac{1}{n_k}\sum_{i=0}^{n_k-1} T_*^i \delta_{x_k}$, where δ_{x_k} is the Dirac measure at x_k. By compacity of the space $\mathcal{M}(X)$ of Borel probability measures on X in the *weak** topology (see section 3), one can suppose that the sequence $\left(\mu_k\right)_{k\geq 0}$ converges weakly to a probability measure μ' which is T-invariant because

$$||T_*\mu_k - \mu_k|| = ||\frac{1}{n_k}\left(T_*^{n_k}\delta_{x_k} - \delta_{x_k}\right)|| \leq \frac{2}{n_k}.$$

This contradicts unicity because $\int f \, d\mu' = \lim_{k\to\infty} \int f \, d\mu_k \neq \int f \, d\mu$. For the converse, if $L(f)$ is the uniform limit of the sequence $\frac{1}{n}\sum_{i=0}^{n-1} f \circ T$, each T-invariant probability measure μ satisfies

$$L(f) = \int L(f) d\mu = \lim_{n\to\infty} \int \frac{1}{n} \sum_{i=0}^{n-1} f \circ T d\mu = \frac{1}{n} \sum_{i=0}^{n-1} \int f \, d\mu = \int f \, d\mu.$$

Definition 19 (Unique ergodicity) *A mapping T satisfying the above equivalent properties is said to be "uniquely ergodic".*

Proposition 39 (Carleman, Denjoy, Furstenberg) *If $\overline{f} \in Homeo_+(\mathbb{T}^1)$ has an irrational rotation number, it is uniquely ergodic.*

Proof. By lemma 19, a rotation with irrational rotation number is uniquely ergodic and by proposition 32, there exists a semi-conjugation $\overline{h}$ of $\overline{f}$ to the rotation $R_\rho(\overline{f})$. Let $S \subset \mathbb{T}^1$ be the set of points x such that $\overline{h}^{-1}(x)$ is an interval. S is countable hence of Haar measure 0. On the other hand, any $\overline{f}$-invariant probability measure μ satisfies $\mu(\overline{h}^{-1}(S)) = 0$ because the wandering open intervals are disjoint, hence they have measure 0, and their boundaries are countable, hence also of measure zero because the absence of periodic points implies that μ has no atoms. Finally, $\overline{h} : \mathbb{T}^1 \setminus \overline{h}^{-1}(S) \to \mathbb{T}^1 \setminus S$ is a bimeasurable bijection, which proves that $\overline{h} : (T^1, \mu) \to \mathbb{T}^1, Haar)$ is a an isomorphism of measured space, which defines uniquely the measure μ.

Corollary 40 *Let f be a C^1 diffeomorphism of $\mathbb{T}^1$ with an irrational rotation number and let μ be its unique invariant probability measure. Then*

$$\lim_{n\to+\infty} \frac{1}{n} \log Df^n = \int_{\mathbb{T}^1} \log(Df) d\mu = 0.$$

In words, the derivative of the iterates of f has at most a subexponential growth.

Proof. As $\frac{1}{n} \log Df^n$ coincides with the Birkhoff sum $\frac{1}{n} \sum_{k=0}^{n-1} \log Df \circ f^k$, it converges uniformly to $\int_{\mathbb{T}^1} \log(Df) d\mu$. This implies the result because, f^n being a diffeomorphism of $\mathbb{T}^1$, its average $\int_{\mathbb{T}^1} Df^n = +1$ (if f is orientation preserving) but if $\int_{\mathbb{T}^1} \log(Df) d\mu$ was strictly positive (resp. strictly negative), Df^n would converge uniformly to $+\infty$ (resp. to 0), which would be a contradiction. This corollary plays a role in the proof of Denjoy theorem 33 .

5.5 Mixing

In order to prove ergodicity of the Bernoulli shifts, we shall prove that they are *mixing*, a strictly stronger property:

Definition 20 (Mixing) *Let $(X, \mathcal{X}, \mu, T)$ be a measured dynamical system. One says that T is mixing if for any $A, B \in \mathcal{X}$,*

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

Definition 21 (Weak mixing) *Let $(X, \mathcal{X}, \mu, T)$ be a measured dynamical system. One says that T is weak mixing if for any $A, B \in \mathcal{X}$, $\mu(A \cap T^{-n}(B))$ converges to $\mu(A)\mu(B)$ in the sense of Cesaro, that is if*

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

Exercise 30 *Mixing implies weak mixing and weak mixing implies ergodicity.*

Exercise 31 *T is mixing if and only if for every $f, g \in L^2(X, \mathcal{X}, \mu)$,*

$$\lim_{n\to\infty} \int_X f \cdot (g \circ T^n) d\mu = \left(\int_X f d\mu\right)\left(\int_X g d\mu\right);$$

it is weak mixing if and only if for every $f, g \in L^2(X, \mathcal{X}, \mu)$,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \left| \int_X f \cdot (g \circ T^k) d\mu - \left(\int_X f d\mu\right)\left(\int_X g d\mu\right)\right| = 0.$$

The following lemma shows that weak mixing is strictly stronger than ergodicity:

Lemma 41 *Let $\alpha = (\alpha_1, \cdots, \alpha_r) \in \mathbb{R}^r/\mathbb{Z}^r$ be such that $1, \alpha_1, \cdots, \alpha_r$ are rationally independent. The rotation R_α of the r-torus $\mathbb{T}^r = \mathbb{R}^r/\mathbb{Z}^r$ is ergodic for the Haar measure but it is never weakly mixing.*

Proof. To prove that the system is not weakly mixing, it is enough to choose the functions $f(x) = e^{2i\pi x_1}$ and $g(x) = e^{-2i\pi x_1}$.

Theorem 42 *Bernoulli shifts T on $\{0,1\}^{\mathbb{N}^*}$ or $\{0,1\}^{\mathbb{Z}}$ are mixing (and hence ergodic) for any one of the product probability measures $\mu = \mu_{p,q}$.*

Proof. One deduces from lemma 12 and exercise 14 that it is enough to check the defining property on the algebra $\mathcal{G}$ of finite unions of cylinders. Indeed, if $A, B \in \mathcal{X}$ and $A_0, B_0 \in \mathcal{G}$ satisfy $\mu(A\Delta A_0) < \epsilon$ and $\mu(B\Delta B_0) < \epsilon$, one has, for all n, $\mu(T^{-n}(A)\Delta T^{-n}(A_0)) = \mu(A\Delta A_0) < \epsilon$ by preservation of the measure and hence $\mu\big(Tf^{-n}(A) \cap B)\Delta(T^{-n}(A_0) \cap B_0)\big) < 2\epsilon$. One deduces that $|\mu(T^{-n}(A) \cap B) - \mu(T^{-n}(A_0) \cap B_0)| < 2\epsilon$ and hence that lim sup and lim inf of $\mu(T^{-n}(A) \cap B)$ differ from those of $\mu(A_0 \cap T^{-n}(B_0))$ at most by 2ϵ.
But, given two finite unions of cylinders A_0 and B_0, the set of indices associated to A and $T^{-n}(B_0)$ are disjoint as soon as n is big enough, and this implies that $\mu(T^{-n}(A_0) \cap B_0) = \mu(A_0)\mu(B_0)$. The end of the proof is left to the reader.

Corollary 43 *The map $x \mapsto 2x(mod.1) : [0,1] \to [0,1]$ and the baker map $\tau : [0,1]^2 \to [0,1]^2$ are mixing, and hence ergodic, for the Lebesgue measure.*

In [T] (page 142), Tao gives the following variant of theorem 35:

Theorem 44 *Let $(X, \mathcal{X}, \mu, T)$ be a measured dynamical system and let $A \in \mathcal{X}$ be a set of positive measure. Then*

$$\limsup_{n\to\infty} \mu(A \cap T^{-n}A) \geq \mu(A)^2.$$

Proof. Applying Cauchy-Schwarz inequality to the sum of the characteristic functions of the sets $T^{-i}(A)$ and the constant function equal to 1, one gets

$$\int_X \left(\sum_{i=1}^n 1_{T^{-i}A}\right)^2 d\mu \geq n^2\mu(A)^2.$$

The left hand side equals

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\mu(T^{-i}A\cap T^{-j}A)=\sum_{i=1}^{n}\sum_{j=1}^{n}\mu(A\cap T^{i-j}A),$$

hence it admits a bound of the form

$$n^2\left(\limsup_{i\to\infty}\mu(A\cap T^iA)+o(1)\right)\geq\int_X\left(\sum_{i=1}^{n}1_{T^{-i}A}\right)^2d\mu,$$

where the term $o(1)\to 0$ when $n\to\infty$, and one concludes by combining both inequalities and letting n tend to infinity.

Remark. Recall that given a family $(C_n)_{n\in\mathbb{N}}$ of subsets of a set X, one defines $\limsup A_n=\cap_n\cup_{k\geq n}A_k$. Fatou's inequality

$$\mu(\limsup(A\cap T^{-i}A))\geq\limsup\mu(A\cap T^{-i}(A))$$

is in general strict: by theorem 35, $\mu(\limsup(A\cap T^{-i}A))=\mu(A)$. Comparing the two theorems, one gets an idea of the non uniformity of return times in A, the extreme cases being a periodic system, where $\limsup_{n\to\infty}\mu(A\cap T^{-n}A)=\mu(A)$ and a mixing system, where $\limsup_{n\to\infty}\mu(A\cap T^{-n}A)=\mu(A)^2$.

6 The main ergodic theorems

6.1 The operator point of view: Von Neuman's ergodic theorem

Let $H=L^2(X,\mathcal{X},\mu)$ be the (separable) Hilbert space of square μ-integrable complex functions on X endowed with the scalar product

$$\langle f,g\rangle=\int_X f(x)\bar{g}(x)d\mu(x),$$

and let $U_T:H\to H$ be the operator defined by $U_Tf=f\circ T$.

Lemma 45 *If T is measure preserving, U_T is an isometry of H; if moreover it is invertible, U_T is unitary.*

Proof. As the L^2 norm of f is the square root of the L^1 norm of f^2, it is enough to notice that $U=U_T$ is an isometry of the normed space $L^1(X,\mathcal{X},\mu)$. If f is the characteristic function of a measurable set $A\in\mathcal{A}$, the L_1-norm of Uf is $||Uf||_1=\mu(T^{-1}A)=\mu(A)=||f||_1$. By linearity, this is true for a "simple function", that is linear combinations of such charateristic functions; as U preserves the order, the theorem of monotone convergence, implies that this is true of simple limits of increasing sequences of positive simple functions. This proves the assertion for real positive functions, hence for all functions (replace f by $|f|$). Being an isometry is equivalent to $U^*U=Id$; if T is invertible, so is U and $U^*=U^{-1}$, hence $UU^*=Id$.

Lemma 46 *If $(X_1, \mathcal{X}_1, T_1)$ and $(X_2, \mathcal{X}_2, T_2)$ are measurably isomorphic, that is if there exists an isomorphism of measurable spaces $\varphi : (X_1, \mathcal{X}_1) \to (X_2, \mathcal{X}_2)$ such that $\varphi \circ T_1 = T_2 \circ \varphi$, the corresponding operators U_{T_1} and U_{T_2} are conjugate: $\Phi \circ U_{T_2} = U_{T_1} \circ \Phi$, où $\Phi f = f \circ \varphi$. One says that T_1 and T_2 have the same spectral type.*

Exercise 32 *1) Show that $(X, \mathcal{X}, T)$ is ergodic if and only if 1 is a simple eigenvalue of the operator U_T*

Caution. Same spectral type does not imply isomorphism !

Theorem 47 (Von Neuman's ergodic theorem) *Let $(X, \mathcal{X}, \mu, T)$ be a measured dynamical system. For any $f \in L^2(X, \mathcal{X}, \mu)$, the means*

$$S_n(f) := \frac{1}{n} \sum_{i=0}^{n-1} f \circ T^i$$

converge in L^2 to the orthogonal projection of f on the closed subspace formed by the invariant functions (i.e. the ones such that $f \circ T = f$ a.e.).

Remark. The measure μ needs not be of finite volume and the transformation T needs not be invertible.

This theorem follows from the following, purely geometric, property of isometries of a Hilbert space:

Theorem 48 *Let $(H, <, >)$ and $U : H \to H$ be respectively a (separable) Hilbert space and an isometry. For any $f \in H$, the sequence*

$$S_n U(f) := \frac{1}{n} \sum_{i=0}^{n-1} U^i(f)$$

converges to the image $\pi(f)$ of f by the orthogonal projector π on the closed subspace $H^U := \ker(Id - U)$ of invariant elements.

Proof. 1) If f belongs to H^U, the $S_n U(f)$ are all equal to f;

2) if $f = g - Ug$ belongs to the (not necessarily closed) subspace

$$W := Im(Id - U),$$

one has $||S_n U f|| \leq \frac{2}{n}||g||$, hence $\lim_{n\to\infty} S_n U f = 0$. This is also true if $f \in \overline{W}$, thanks to the uniform (in n) bound $||S_n U|| \leq 1$: indeed, if $(f_k)_{k\in\mathbb{N}}$ is a sequence of elements of W converging to $f \in \overline{W}$, one has for any n, k,

$$||S_n U f|| \leq ||S_n U(f - f_k)|| + ||S_n U f_k|| \leq ||(f - f_k)|| + ||S_n U f_k||.$$

Given $\epsilon > 0$, choose k such that $||f - f_k|| \leq \frac{\epsilon}{2}$, then n such that $||S_n U f_k|| \leq \frac{\epsilon}{2}$;

3) the orthogonal complement $W^\perp$ of W is H^{U^*}: indeed, let h be an element of $W^\perp$; this is equivalent to $\langle h, g - Ug\rangle = 0$, that is $\langle h, g\rangle = \langle h, Ug\rangle = \langle U^*h, g\rangle$ or $\langle h - U^*h, g\rangle = 0$, for all $g \in H$, which is equivalent to $h = U^*h$;

4) U being an isometry, $H^{U^*} = H^U$. In one direction, if $Uf = f$, then $U^*Uf = U^*f$ hence $U^*f = f$ because U is an isometry if and only if $U^*U = Id$. Conversely, if $U^*f = f$, one computes

$$||Uf-f||^2 = ||Uf||^2+||f||^2-\langle Uf, f\rangle-\langle f, Uf\rangle = ||Uf||^2+||f||^2-\langle f, U^*f\rangle-\langle U^*f, f\rangle\,,$$

that is

$$||Uf - f|| = ||Uf||^2 - ||f||^2 = 0.$$

5) it follows from 2) and 3) that $(H^U)^\perp = (W^\perp)^\perp = \overline{W}$ and hence that $H = H^U \oplus \overline{W}$ from which one concludes using 1) and 2).

6.2 Birkhoff's ergodic theorem

Theorem 49 *Let $(X, \mathcal{X}, \mu, T)$ be a measured dynamical system. For every function $f \in L^1(X, \mathcal{X}, \mu)$, the limit of "Birkhoff sums"*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f\big(T^k(x)\big) := f^*(x)$$

exists for μ-almost every $x \in X$ and it defines a function $f^* \in L^1(X, \mathcal{X}, \mu)$ satisfying $f^* \circ T = f^*$ (μ-a.e.) and $\int_X f(x)\, d\mu(x) = \int_X f^*(x)\, d\mu(x)$. If T is invertible, the functions f^* and $\bar{f}^*$ respectively defined by T and T^{-1} coincide almost everywhere.

Corollary 59 will give a probabilistic interpretation of the limit.

Corollary 50 *Under the same hypotheses, if moreover T is ergodic, f^* is a constant, equal to $\int_X f d\mu$.*

In words, this means that if T is ergodic, the *time average*, that is the limit of the *Birkhoff sums* exists almost everywhere and is equal to the integral, that is to the *spatial average*. If for example f is the characteristic function $\mathcal{X}_A$ of a measurable subset $A \in \mathcal{X}$, the corollary asserts that, for almost every x, the proportion of "time" the orbit of x spends in A coincides with the measure (the probability) of A ($n \in \mathbb{N}^*$ or $n \in \mathbb{Z}$ should indeed be interpreted as a discrete time, the unit of time corresponding to one iteration of T).

The proof of Birkhoff's theorem will follow from the first part of the proof of Kingman's theorem. It remains to prove the last assertion, that is $f^* = \overline{f}^*$ if T is invertible. Suppose that the a.e. T-invariant set $Y = \{x \in X,\ f^* > \overline{f}^*\}$ has positive measure. Applying Birkhoff's theorem to the restrictions of T and T^{-1} to Y one gets

$$\int_Y f^* d\mu = \int_Y f d\mu = \int_Y \overline{f}^* d\mu\,,$$

hence $\int_Y (f^* - \overline{f}^*) d\mu = 0$. But this is a contradiction because $f^* - \overline{f}^*$ is strictly positive on Y.

6.3 Kingman's subadditive ergodic theorem

This is a generalization of Birkhoff's theorem with important applications to cocycles and their Lyapunov exponents. Many proofs where given of this theorem since Kingman's original one. I follow closely the proof in [AB].

Sequences of Birkhoff sums $f_n = \sum_{i=0}^{n-1} f \circ T^i$, which are *$T$-additive*, that is such that $f_{m+n} = f_m + f_n \circ T^m$, are replaced by sequences of functions which are *T-subadditive*, that is such that

$$f_{m+n} \leq f_m + f_n \circ T^m.$$

Example: coycles. If $F : X \times E \to X \times E$ is a *cocycle* above T, that is a map of the form

$$F(x, v) = \big(T(x), A(x)v\big),$$

where $x \mapsto A(x)$ is a family of linear endomorphisms of the vector space E, its n-th iterate

$$F^n(x, v) = \big(T^n(x), A_n(x)v\big), \quad \text{where} \quad A_n(x) = A(T^{n-1}(x)) \cdots A(T(x))A(x),$$

satisfies

$$A_{m+n}(x) = A_n(T^m(x))A_m(x).$$

It follows that the family of functions $f_n(x) = \log ||A_n(x)||$ is subadditive.

Theorem 51 *Let $(X, \mathcal{X}, \mu, T)$ and let $f_n : X \to \overline{\mathbb{R}}$ be respectively a measured dynamical system and a T-subadditive sequence of measurable functions such that*

$$\sup(f_1, 0) := f_1^+ \in L^1 = L^1(X, \mathcal{X}, \mu).$$

Then the sequence $\frac{1}{n} f_n$ converges μ-almost everywhere to a function $f : X \to \overline{\mathbb{R}}$ such that

$$f^+ \in L^1 \quad \textit{and} \quad \int_X f d\mu = \lim_{n\to\infty} \frac{1}{n} \int_X f_n d\mu = \inf_n \frac{1}{n} \int_X f_n d\mu \in [-\infty, \infty(\ .$$

Proof. By induction one sees that, for any $n \geq 1$, $f_n^+ \in L^1$. The sequence $u_n = \int_X f_n d\mu \in \mathbb{R} \cup \{-\infty\}$ is subadditive, that is $u_{n+m} \leq u_n + u_m$; hence lemma 25 insures that

$$\lim_{n\to\infty} \frac{1}{n} \int_X f_n d\mu = L := \inf_n \int_X f_n d\mu \in [-\infty, \infty(\ .$$

Let

$$f_\flat = \liminf_{n\to\infty} \frac{1}{n} f_n, \qquad f_\sharp = \limsup_{n\to\infty} \frac{1}{n} f_n.$$

Lemma 52 *$f_\flat$ and $f_\sharp$ are measurable almost everywhere T-invariant functions.*

Proof.

$$f_\flat(x) \leq \liminf_{n\to\infty} \frac{f_1(x) + f_{n-1}(Tx)}{n} = f_\flat(Tx).$$

Hence

$$\forall a \in \overline{\mathbb{R}},\ f_\flat(Tx) \leq a \implies f_\flat(x) \leq a,\ \text{i.e. } T^{-1}(\{x, f_\flat(x) \leq a\}) \subset \{x, f_\flat(x) \leq a\}.$$

As T preserves the finite measure μ, this implies that

$$\forall a \in \overline{\mathbb{R}},\ T^{-1}(\{x, f_\flat(x) \leq a\}) = \{x, f_\flat(x) \leq a\},$$

the equality meaning always "almost everywhere equal". The reasoning for $f_\sharp$ is analogous.
The fundamental lemma is the following:

Lemma 53

$$\int_X f_\flat d\mu = L.$$

Proof. Following an idea introduced by Katznelson and Weiss whose proof was itself inspired by Kamae's proof using non standard analysis, let us define, for $C \in \mathbb{R}$,

$$f_n^{(C)} = \sup(f_n, -Cn).$$

The sequence $f_n^{(C)}$ is subadditive (exercise) and

$$f_\flat^{(C)} := \liminf_{n\to\infty} \frac{1}{n} f_n^{(C)} = \sup(f_\flat, -C), \quad f_\sharp^{(C)} := \limsup_{n\to\infty} \frac{1}{n} f_n^{(C)} = \sup(f_\sharp, -C).$$

By the monotone convergence theorem applied to the decreasing sequences, when $C \to \infty$,

$$f_\flat^{(C)} \to f_\flat = \inf_C f_\flat^{(C)} \quad \text{and} \quad f_n^{(C)} \to f_n = \inf_C f_n^{(C)},$$

one gets

$$\int_X f_\flat d\mu = \inf_C \int_X f_\flat^{(C)} d\mu \quad \text{and} \quad \int_X f_n d\mu = \inf_C \int_X f_n^{(C)} d\mu,$$

hence

$$L := \inf_n \frac{1}{n} \int_X f_n d\mu = \inf_n \inf_C \frac{1}{n} \int_X f_n^{(C)} d\mu = \inf_C \inf_n \frac{1}{n} \int_X f_n^{(C)} d\mu.$$

Hence, it is enough to prove that for C fixed, $\int_X f_\flat^{(C)} d\mu = L^{(C)} := \inf_n \frac{1}{n} \int_X f_n^{(C)} d\mu$. In other words, one can suppose that there exists a constant $C \in \mathbb{R}$ such that

$$\forall n, \frac{1}{n} f_n \geq -C.$$

Now let us recall the

Lemma 54 (Fatou) *Let $(X, \mathcal{X}, \mu)$ be a measure space and let $(F_n)_{n\in\mathbb{N}}$ be a sequence of measurable non negative functions on X. Then*

$$\int_X \liminf_{n\to\infty} F_n d\mu \leq \liminf_{n\to\infty} \int_X F_n d\mu.$$

The proof of this lemma consists in applying the monotone convergence theorem to the increasing sequence $G_p = \inf_{n\geq p} f_n$. Setting $F_n = \frac{1}{n} f_n + C \geq 0$, we get $\int_X f_\flat d\mu + C \leq L + C$, hence $\int_X f_\flat d\mu \leq L$.
We now prove the converse inequality: fixing $\epsilon > 0$, we define for $k \geq 1$,

$$E_k = \big\{x \in X,\ \exists j \in \{1, 2, \cdots, k\}, \quad \frac{1}{j} f_j < f_\flat(x) + \epsilon\big\}.$$

This is an increasing sequence of sets such that $\cup_k E_k = X$. Now define

$$\psi_k(x) = \begin{cases} f_\flat(x) + \epsilon & \text{if} \quad x \in E_k, \\ f_1(x) & \text{if} \quad x \notin E_k. \end{cases}$$

Notice that, if $x \notin E_k,\ f_1(x) \geq f_\flat(x) + \epsilon$, hence $\psi_k \geq f_\flat + \epsilon$.

Lemma 55 *For all $n \geq k$ and almost all $x \in X$, the following key inequality is satisfied:*

$$f_n(x) \leq \sum_{i=0}^{n-k-1} \psi_k(T^i x) + \sum_{i=n-k}^{n-1} \sup(\psi_k, f_1)(T^i x).$$

Proof of lemma 55. Let $x \in X$ be such that $\forall i,\ f_\flat(T^i x) = f_\flat(x)$ (by lemma 52, almost all x are such).
Now, let $m_0 = 0$ and let n_1 be the least integer greater or equal to m_0 such that $T^{n_1} x \in E_k$. By definition, there exists $m_1 \in \{n_1 + 1, \cdots, n_1 + k\}$ such that

$$\frac{1}{m_1 - n_1} f_{m_1 - n_1}(T^{n_1} x) < f_\flat(T^{n_1} x) + \epsilon = f_\flat(x) + \epsilon.$$

Then let n_2 be the least integer greater or equal to m_1 such that $T^{n_2} x \in E_k$. By definition, there exists $m_2 \in \{n_2 + 1, \cdots, n_2 + k\}$, ... In this way, one defines inductively

$$m_0 = 0 \leq n_1 < m_1 \leq n_2 < m_2 \leq \cdots$$

where n_j is the smallest integer greater or equal to m_{j-1} such that $T^{n_j} \in E_k$ and m_j is such that

$$1 \leq m_j - n_j \leq k \quad \text{and} \quad f_{m_j - n_j}(T^{n_j} x) < (m_j - n_j)(f_\flat(x) + \epsilon).$$

Given $n \geq k$, let l be the biggest integer such that $m_l \leq n$. Subadditivity of the sequence f_n implies that

$$f_n(x) \leq f_{n-1}(x) + f_1(T^{n-1} x) \leq \cdots \leq f_{m_l}(x) + f_1(T^{m_l} x) + \cdots + f_1(T^{n-1} x)$$

and also that

$f_{m_l}(x) \leq f_{n_l}(x)+f_{m_l-n_l}(T^{n_l}x)$ and $f_{n_l}(x) \leq f_{m_{l-1}}(x)+f_1(T^{m_{l-1}}x)+\cdots+f_1(T^{n_l-1}x)$.

Hence, adding such inequalities, we get

$$f_n(x) \leq \sum_{i\in A} f_1(T^i x) + \sum_{j=1}^{l} f_{m_j-n_j}(T^{n_j}x), \text{ where } A = \cup_{j=0}^{l-1}[m_j, n_{j+1}[\cup[m_l, n[.$$

Now, all the $T^i x$ such that $i \in A$, except possibly if $i \in [n_{l+1}, n[$ in case n is such that $n_{l+1} < n-1$, are in the complement of E_k where $\psi_k = f_1$. Hence

$$\sum_{i\in A} f_1(T^i x) = \sum_{j=0}^{l-1} \sum_{i\in[m_j,n_{j+1}[} \psi_k(T^i x)+ \sum_{i\in[m_l,\inf(n_{l+1},n-1)]} \psi_k(T^i x)+ \sum_{i=n_{l+1}}^{n-1} f_1(T^i x),$$

where the last term is present only in case $n_{l+1} < n-1$.

On the other hand, from the invariance of $f_\flat$ along the orbit of x and the definition of ψ_k one obtains

$$f_{m_j-n_j}(T^{n_j}(x)) \leq (m_j-n_j)(f_\flat(T^{n_j}x)+\epsilon) = \sum_{i\in[n_j,m_j[} (f_\flat(T^i x)+\epsilon) \leq \sum_{i\in[n_j,m_j[} \psi_k(T^i x).$$

Adding the estimates for $\sum_{i\in A} f_1(T^i x)$ and $f_{m_j-n_j}(T^{n_j}(x))$, we get

$$f_n(x) \leq \sum_{i=0}^{\inf(n_{l+1},n-1)} \psi_k(T^i x) + \sum_{i=n_{l+1}}^{n-1} f_1(T^i x),$$

where the last term is present only in case $n_{l+1} < n-1$.

Finally, as $m_{l+1} > n$ and $m_{l+1} - n_{l+1} \leq k$, one has always $n_{l+1} > n-k$, hence the inequality stated in the key lemma 55.

Integrating lemma 55 and dividing by n gives

$$\frac{1}{n}\int_X f_n d\mu \leq \frac{n-k}{n}\int_X \psi_k d\mu + \frac{k}{n}\int_X \sup(\psi_k, f_1) d\mu.$$

The function $\sup(\psi_k, f_1)$ is integrable because, on the one hand $f_1 \leq f_1^+$ which is supposed to be in L^1, on the other hand, $\psi_k \leq f_\flat + \epsilon$ whose integral was proved to be less or equal to $L+\epsilon$. Letting n tend to $+\infty$ we get that for all k, $\quad L \leq \int_X \psi_k d\mu$. Letting k tend to $+\infty$, as ψ_k converges simply to $f_\flat + \epsilon$ and is dominated by the integrable function $\sup(f_\flat + \epsilon, f_1^+)$ we get $L \leq \int_X f_\flat d\mu + \epsilon$. Letting ϵ tend to 0 we conclude that $L \leq \int_X f_\flat d\mu$. This proves lemma 53.

Proof of Birkhoff theorem. It follows directly from lemma 53: indeed, replacing f_1 by $-f_1$ in the inequality $L \leq \int_X f_\flat d\mu$, we get $\int_X f_\sharp d\mu \leq L$ which implies

$f_\flat = f_\sharp$ almost everywhere. The following immediate corollary of Birkhoff's theorem will be used in the end of the proof of Kingman's theorem:

$$\text{If } f \in L^1(X, \mathcal{X}, \mu), \quad \text{then} \quad \lim_{n\to\infty} \frac{1}{n} g \circ T^n = 0 \quad \text{a.e.}$$

End of the proof of Kingman's theorem. We want to prove that $f_\flat = f_\sharp$. We shall prove that $\int_X f_\sharp \leq L$, which allows to conclude if $L > -\infty$, hence in the general case by a truncation argument.

Lemma 56

$$\forall k \geq 1,\ \limsup_{n\to\infty} \frac{1}{n} f_{kn} = k f_\sharp \quad a.e.$$

Proof. The inequality $\limsup_{n\to\infty} \frac{1}{kn} f_{kn} \leq f_\sharp$ a.e. is obvious because the $(f_{kn})|_{n\in\mathbb{N}}$, form a subsequence of $(f_n)|_{n\in\mathbb{N}}$.

To prove the inequality in the other direction, let us write (Euclidean division)

$$n = km_n + r_n,\ 0 \leq r_n < k,$$

By subdditivity,

$$f_n \leq f_{km_n} + f_{r_n} \circ T^{km_n} \leq f_{km_n} + g \circ T^{km_n},\ \text{where } g = \sup(f_1^+, f_2^+, \cdots, f_{k-1}^+).$$

It follows that

$$k f_\sharp \leq \frac{km_n}{n} \limsup \frac{1}{m_n} f_{km_n} + k \frac{km_n}{n} \limsup \frac{1}{km_n} g \circ T^{km_n}.$$

The first term trends to $\limsup \frac{1}{km_n} f_{km_n} \leq \limsup \frac{1}{n} f_{kn}$, while the second one tends to $k \limsup \frac{1}{km_n} g \circ T^{km_n} \leq k \limsup \frac{1}{n} g \circ T^n$, which tends to 0 by the corollary of Birkhoff's theorem that we just mentioned because we have noticed at the beginning that, as $f_1^+ \in L^1$, so are all f_n^+, hence $g \in L^1$.

To finish the proof that $\int_X f_\sharp d\mu \leq L$, we may first suppose as above that there exists $C \in \mathbb{R}$ such that $\frac{1}{n} f_n \geq -C$ for all $n.$, then use the argument used at the beginning of the proof of lemma 53. Fixing the integer k, we consider the Birkhoff sum

$$F_n = \sum_{i=0}^{n-1} \left(-f_k \circ T^{jk}\right).$$

This is an additive sequence with respect to T^k, which moreover is such that $F_1 = -f_k \leq Ck$, hence satisfies $F_1^+ \in L^1$ (remember $\mu(X) = 1$). Now, let $F_\flat = \limsup_{n\to\infty} \frac{1}{n} F_n$. From lemma 53 and the T^k invariance of μ one deduces

$$\int_X F_\flat d\mu = \lim_{n\to\infty} \frac{1}{n} \int_X F_n d\mu = \inf_n \frac{1}{n} \int_X F_n = -\int_X f_k d\mu.$$

On the other hand, by definition,

$$-F_\flat = -\liminf_{n\to\infty} \frac{1}{n} F_n = \limsup_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} f_k \circ T^{jk}.$$

Exchanging the roles of n and k, we deduce from the subadditivity that

$$f_{nk} \leq f_k + f_k \circ T^k + \cdot + f_k \circ T^{(n-1)k},$$

hence, using lemma 56,

$$\limsup_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} f_k \circ T^{jk} \geq \limsup_{n\to\infty} \frac{1}{n} f_{kn} = k f_\sharp.$$

Finally, $\forall k,\ -F_\flat \geq k f_\sharp$ that is

$$\int_X f_\sharp d\mu \leq -\frac{1}{k} \int_X F_\flat d\mu = \frac{1}{k} \int_X f_k d\mu, \quad \text{hence} \quad \int_X f_\sharp d\mu \leq \inf_k \frac{1}{k} \int_X f_k d\mu := L.$$

This ends the proof of Kingman's subadditive ergodic theorem.

Applied to the example of a cocycle above an ergodic transformation, this theorem implies the existence of the *largest Lyapunov exponent*.

6.4 Conditional expectation and the ergodic theorems

In this section we interpret the limit of Birkhoff's sum given by the Birkhoff theorem in terms of probabilistic notions.

Let $(X, \mathcal{X}, \mu)$ be a probability space and $f \in L^1(X, \mathcal{X}, \mu)$ be a *random variable* (this is just a fancy name used in general for measurable maps from a probability space to $\mathbb{R}$ endowed with its Borelian tribe $\mathcal{B}$). The *expectation* of f is by definition its integeral $\mathbb{E}(f) = \int_X f d\mu$. Given a sub-tribe $\mathcal{Y}$ of $\mathcal{X}$, one can define the *conditional expectation of f given $\mathcal{Y}$*. The definition is particularly transparent in case the sub-tribe $\mathcal{Y}$ is generated by a partition because it is then directly related to the notion of *conditional probability*:

Definition 22 (conditional expectation given a partition) *The conditional expectation $\mathbb{E}(f||\mathcal{Y})$ of the random variable $f : X \to \mathbb{R}$ given the tribe $\mathcal{Y}$ generated by a partition $X = B_1 + B_2 + \ldots + B_n$ is the random variable $\mathbb{E}(f||\mathcal{Y}) : X \to \mathbb{R}$ which takes the constant value*

$$\mathbb{E}(f||\mathcal{Y})(x) = \frac{1}{\mu(B_i)} \int_{B_i} f d\mu \quad \textit{if} \quad x \in B_i.$$

on each of the pieces A_i of the partition.

If $f = \mathcal{X}_A$ is the characteristic function of some event $A \in \mathcal{X}$, the value of $\mathbb{E}(\mathcal{X}_A||\mathcal{Y})$ on B_i is just the *conditional probability* of A given B_i:

$$\mu(A||B_i) = \frac{1}{\mu(B_i)} \mu(A \cap B_i).$$

In the general case, the definition relies on the Radon-Nikodym theorem:

Theorem 57 (Radon-Nikodym) *Let μ, ν be two positive and finite measures defined on the same tribe $\mathcal{X}$ of a set X. If ν is absolutely continuous with respect to μ, i.e. if every subset of μ measure 0 is also of ν measure 0, there exists a $\mathcal{X}$-measurable function $\varphi : X \to [0, +\infty[$ (the density of ν with respect to μ) such that*

$$\forall A \in \mathcal{X}, \quad \nu(A) = \int_A \varphi d\mu.$$

Moreover, two such densities are equal almost everywhere.

Corollary 58 (conditional expectation given a general sub-tribe) *Let $(X, \mathcal{X}, \mu)$ be a probability space and $\mathcal{Y}$ be a sub-tribe of $\mathcal{X}$. There exists a well defined linear projector $f \mapsto \mathbb{E}(f||\mathcal{Y})$ from $L^1(X, \mathcal{X}, \mu)$ to $L^1(X, \mathcal{Y}, \mu)$ such that*

$$\forall B \in \mathcal{Y}, \quad \int_B \mathbb{E}(f||\mathcal{Y}) d\mu = \int_B f d\mu.$$

Proof. A function in $L^1(X, \mathcal{X}, \mu)$ being a linear combination of positive functions, one can assume that $f \geq 0$. Then the map $B \mapsto \int_B f d\mu$ defines a positive measure on $\mathcal{Y}$ which is absolutely continuous with respect to the restriction of μ to $\mathcal{Y}$, hence the Radon-Nykodim theorem insures the existence of the $\mathcal{Y}$ measurable function $\mathbb{E}(f||\mathcal{Y})$. In case $f = \mathcal{X}_B$ is the caracteristic function of $B \in \mathcal{Y}$, the conditional expectation is still related to the conditional probability by the formula $\mu(B||\mathcal{Y}) = \mathbb{E}(\mathcal{X}_B||\mathcal{Y})$,

Corollary 59 *The limit f^* of a Birkhoff sum whose existence is asserted by Theorem 49 is the conditional expectation of f with respect to the invariant tribe $\mathcal{I}(T)$:*

$$f^*(x) = \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f\big(T^k(x)\big) = \mathbb{E}\big(f||\mathcal{I}(T)\big).$$

Proof. f^* is a.e. T-invariant, hence $\mathcal{I}(T)$ measurable (Exercise 28) and its integral is equal to the integral of f.

6.5 Applications: law of large numbers, entropy

6.5.1 Strong law of large numbers

Applied to the Bernoulli shifts, corollary 50 says that the statistical structure of almost all sequences is the same, which is a strong form of the so-called *strong law of large numbers*. In what follows, we consider only the case of random variables with finite values.

To a random variable $f : (X, \mathcal{X}, \mu) \to (\mathbb{R}, \mathcal{B})$, one associates the direct image (see Definition 10) $f_*\mu$ of the probability measure, which is often the only thing we have access to. If f takes only a finite set $\{\alpha_1, \cdots, \alpha_n\}$ of values, it is associated to the finite partition $X = A_1 + \cdots + A_n$ of X, where $A_i = f^{-1}\alpha_i$, and the image probability $f_*\mu$ is defined by $\{p_1, \cdots, p_n\}$, where $p_i = \mu(A_i)$:

the measure of an interval $I \in \mathbb{R}$ (or more generally of a Borelian) is the sum $\sum_{i,\alpha_i \in I} p_i$. The expectation of f is $\mathbb{E}(f) = \int_X f d\mu = \sum_i p_i \alpha_i$.

Remark on the notations. In measure theory one uses in general the notations $f_*\mu(I)$ or $\int_I d(f_*\mu)$ while in probability theory one rather writes $\mu\{f \in I\}$ or $Pr\{f \in I\}$.

Definition 23 *Two random variables f, g with the same image measures are said to be identically distributed.*

Definition 24 *Random variables $f_1, \cdots, f_n$ with finite values on the probability space $(X, \mathcal{X}, \mu)$ are said to be independent if the corresponding partitions of X,*

$$X = A_1^{(i)} + \cdots + A_{r_i}^{(i)}, \quad i = 1, \cdots, n,$$

are independent, that is if they satisfy

$$\forall k_i \in \{1, \cdots r_i\},\ i = 1, \cdots, n, \qquad \mu\big(A_{k_1}^{(1)} \cap \cdots \cap A_{k_n}^{(n)}\big) = \mu\big(A_{k_1}^{(1)}\big) \cdots \mu\big(A_{k_n}^{(n)}\big).$$

Warning. Three partitions can be pairwise independent without being independent ! (Exercise, find an exemple).

A typical, and indeed universal, example of a family of independent identically distributed (i.i.d.) random variables is

$$f_i : (\{0,1\}^{\mathbb{N}^*}, \mathcal{B}, \mu_{p,q}) \to \mathbb{R}, \quad f_i(a_1 a_2 \ldots) = a_i, \quad i = 1, 2, \cdots$$

More generally, one can replace the alphabet with two letters 0,1 by an alphabet $\{\alpha_1, \cdots, \alpha_r\}$ with an arbitrary number r of letters, for example $r = 26$ as in the french alphabet. I leave to the reader the task of defining the probability laws $\mu_{p_1,\cdots,p_r}$ and the Borelian tribe $\mathcal{B}$ on $\{\alpha_1, \cdots, \alpha_r\}^{\mathbb{N}^*}$ and show that they are invariant by the shift and ergodic.

Theorem 60 (Stong law of large numbers in the independent case) *If $f_1, \cdots, f_n, \cdots : (X, \mathcal{X}, \mu) \to \mathbb{R}$ are independent and identically distributed random variables whose values are $\alpha_1, \cdots, \alpha_r$ with probabilities $p_1, \cdots, p_r$, one has*

$$Pr\left\{ \lim_{n\to\infty} \frac{1}{n}(f_1 + \cdots + f_n) = \sum_{i=1}^{r} p_i \alpha_i \right\} = 1.$$

Proof. We apply corollary 50 to the generalized shift)

$$T : (\{\alpha_1, \cdots, \alpha_r\}^{\mathbb{N}^*}, \mathcal{B}, \mu_{p_1,\cdots,p_r}) \to: (\{\alpha_1, \cdots, \alpha_r\}^{\mathbb{N}^*}, \mathcal{B}, \mu_{p_1,\cdots,p_r})$$

and to the functions $f_i(a_1 \cdots a_n \cdots) = a_i = f_1(T^{i-1}(a_1, \cdots, a_n, \cdots))$, which are a universal model of i.i.d. random variables. The conclusion follows because, on the one hand $f_1(x) + \cdots + f_n(x) = \sum_{k=0}^{n-1} f_1(T^k(x))$, on the other hand the integral of f_1 on $\{\alpha_1, \cdots, \alpha_r\}^{\mathbb{N}^*}$ is equal to $\sum_{i=1}^{r} p_i \alpha_i$.

Remark. In the same way, Von Neumann's ergodic theorem 47 is related to the so-called *weak law of large numbers.*

An example of more precise results In $(\{0,1\}^{\mathbb{N}^*}, \mathcal{B}, \mu_{p,q})$ consider the cylinder A defined by $a_1 = a_2 = \ldots = a_{1000} = 0$. The Birkhoff sum $\frac{1}{n}\sum_{k=0}^{n-1} \mathcal{X}_A\big(T^k(x)\big)$, where T is the shift, represents the frequency with which one has $a_{k+1} = a_{k+2} = \ldots = a_{k+1000} = 0$ when k varies from 0 to n. The theorem affirms that for almost every sequence, this frequency tends to a limit equal to p^{1000}, when n tends to $+\infty$. The same is true for every cylinder, that is for any finite configuration of 0's and 1's but this does not exhaust the richness of the theorem as the function f may depend of an arbitrary number of coordinates.

Exercise 33 *Apply the ergodic theorem to the same example under the disguise to the map $T(x = 2x \ (mod\ 1)$ of the interval $[0,1]$ endowed with its Borelian tribe and the Lebesgue measure in itself. Deduce the proof that the set of normal numbers in the sense of Borel has measure 1.*

Notice that, from the definition of ergodicity, one could only conclude that the measure of normal numbers was 0 or 1.

6.5.2 Shannon's entropy (see [C3, CT])

Let us apply the strong law of large numbers to the independent identically distributed random variables

$$f_i : \{\alpha_1, \cdots, \alpha_r\}^{\mathbb{N}^*}, \mathcal{B}, \mu_{p_1,\ldots p_r}) \to \mathbb{R}, \quad f_i(a_1 a_2 \cdots) = \log \frac{1}{p(a_i)},$$

where the log is taken in the basis r and $p(a_i) = p_k$ if $a_i = \alpha_k$. As the probability of the cylinder $A_{1\cdots n}^{a_1\cdots a_n}$ is $p(a_1 \cdots a_n) = p(a_1) \cdots p(a_n)$, one gets the

Theorem 61 (AEP) *If $f_1, \cdots, f_n, \cdots : (X, \mathcal{X}, \mu) \to \mathbb{R}$ are i.i.d. random variables with values $\{A_1, \cdots, A_r\}$ and probabilities $(p_1, \cdots, p_r)$, one has*

$$Pr\left\{ \lim_{n\to\infty} \frac{1}{n} \log \frac{1}{p(a_1 \cdots a_n)} = \sum_{i=1}^{r} p_i \log \frac{1}{p_i} \right\} = 1.$$

The real number $h = \sum_{i=1}^{r} p_i \log \frac{1}{p_i}$ is *Shannon's entropy*. The interpretation of this theorem is that, if n is large enough, the probability to encounter sequences (=messages) $a_1 \cdots a_n$ whose probability is close to r^{-nh} is very high, hence the name *Asymptotic Equipartition Property* (AEP). Note that there are only about r^{nh} such very probable sequences among the r^n possible sequences of length n. If $h = 1/2$, this represents 100 sequences among 10000! Figure 9, where the size of elements indicates their probability, illustrates this assertion.

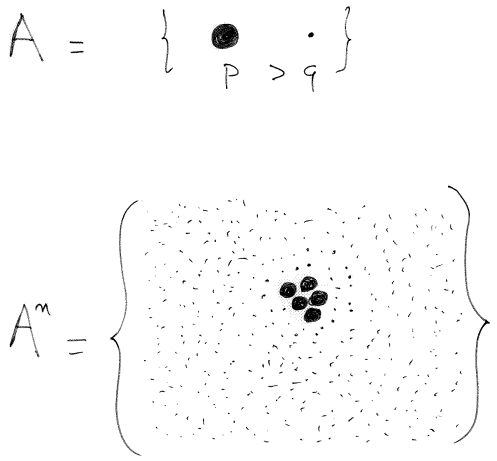


Figure 9 : Shannon's first theorem.

7 A glimpse into dynamical entropies

The aim of this section is simply to give the reader the desire to understand this subtle invariant of dynamical systems. For a complete historical survey, see [K].

7.1 The entropy of a finite probability space

Depending only on the image measure $\xi_*\mu = (p_1, \cdots, p_r)$ on the finite set of values $A = \{A_1, \cdots, A_r\}$, the quantity $\sum_{i=1}^r p_i \log \frac{1}{p_i}$ introduced in the last section as *Shannon's entropy*, is naturally attached to a class of identically distributed (see definition 23) random variables $\xi : \Omega \to A$. In other words, it can be thought of as attached to the finite set $A = \{A_1, \cdots, A_r\}$ endowed with the probability measure $\{p_1, \cdots, p_r\}$. One interprets it is a measure of the information gained from an experiment yielding a value of ξ or, this is equivalent, from picking at random one element of the finite set A : it is maximal if all the p_i are equal (maximal uncertainty before picking), it is minimal equal to 0 in case the probability is concentrated on a single element (the result of the picking is known in advance). In the same way, suppose we make an experiment yielding successively the values of n not necessarily independant random variables $\xi_i : \Omega \to A$, $i = 1, \cdots, n$, that is picking successively at random a sequence of n successive symbols $a_1, a_1, \cdots a_n$. A measure of the information gained from such an experiment can be defined by a similar formula as soon as the set A^n is endowed with a probability measure (the image of μ by the random variable $(\xi_1, \xi_2, \cdots, \xi_n) : \Omega \to A^n$). In case, as in theorem 61, the n random variables are independent and identically distributed (iid), A^n is endowed with the product probability measure defined by giving to the sequence

$A_{j_1}, A_{j_2}, \cdots, A_{j_n}$ the probability $p_{j_1} \cdots p_{j_n}$, which gives for the entropy the value $\sum p_{j_1} \cdots p_{j_n} \log \frac{1}{p_{j_1} \cdots p_{j_n}} = n \sum_i p_i \log \frac{1}{p_i}$.

7.2 The entropy of a discrete source

Allowing n to be infinite, we are led to the

Definition 25 (Discrete source) *A discrete source consists in the data of a finite alphabet $A = \{A_1, \dots A_r\}$ and a probability measure μ on $\Omega = A^{\mathbb{Z}}$ (or $A^{\mathbb{N}^*}$). If μ is invariant under the shift T, the source is said to be stationary.*

The name "source" comes from information theory where an element of $A^{\mathbb{Z}}$ or $A^{\mathbb{N}^*}$ is identified with a *message* whose length is infinite. The corresponding random variables are the $\xi_i(\cdots a_{-2}a_{-1}a_0a_1, a_2 \cdots) = a_i$. In general, the probability law of ξ_i depends on the value of the preceding ones $\xi_{i-1}, \xi_{i-2}, ...$ (think of the probability of the letters in some language). As in full generality the probability of ξ_i depends on the whole past history, we need consider arbitrarily long sequences in order to define an entropy.

Let μ be a probability measure on $A^{\mathbb{N}^*}$ (or $A^{\mathbb{Z}}$) which is invariant under the shift. Let $H_\mu^{<n>}$ be the entropy of the finite set A^n endowed with the probability measure defined by the measure of cylinders (recall definition in section 2.2), that is the direct image under the canonical projection $\pi_n : A^{\mathbb{Z}} \to A^n$ (or $A^{N^*} \to A^n$), $\pi(\cdots a_i \cdots) = a_1 a_2 \cdots a_n$ of the measure μ :

$$H_\mu^{<n>} = \sum_{j_1 j_2 \dots j_n \in \{1,2,\dots,r\}^n} \mu(A_{12\dots n}^{j_1 j_2 \cdots j_n}) \log \frac{1}{\mu(A_{12\dots n}^{j_1 j_2 \cdots j_n})}.$$

The entropy $H_\mu^{<n>}$ is a measure of the information obtained from the emission of a sequence of n successive symbols (or n successive experiments).

Lemma 62 (McMillan) *The "mean information content by symbol" $\frac{1}{n} H_\mu^{<n>}$ tends to a limit $H_\mu = H_\mu(T)$ when the length n of the sequence (the message) tends to infinity :*

$$H_\mu(T) = \lim_{n \to \infty} \frac{1}{n} H_\mu^{<n>} = \inf_n (\frac{1}{n} H^{<n>})$$

is by definition the entropy of the (stationary discrete) source[2].

The proof, which is given in ([C3] par. 8.3), consists in proving the subadditivity of the sequence $u_n = H_\mu^{<n>}$. For this, one decomposes the emission of a sequence of $n + m$ symbols into the emission of the n first symbols followed by the one of the m last symbols and one applies Shannon's inequality for the conditional entropy (see [C3] proposition 10).

[2]The notation $H_\mu(T)$ emphasizes the dependence on the shift T which comes from the stationarity. It is compatible with the more general definition of Kolmogorov entropy, given in the next section.

Theorem 61, which in the iid case is a direct consequence of the law of large numbers, admits a far reaching generalization to any stationary discrete ergodic measure; this is the so called *Shannon-McMillan-Breiman AEP theorem* a short proof of which can be found in [B2]:

Theorem 63 (Shannon-McMillan-Breiman) *The entropy of an ergodic stationary discrete source (T, μ) satisfies the strong Asymptotic Equipartition Property: for μ-almost every $\omega = \ldots a_1 a_2 \ldots a_k \ldots \in A^{\mathbb{Z}}$,*

$$\lim_{n\to\infty} \frac{1}{n} \log \frac{1}{\mu(a_1 a_2 \ldots a_n)} = H_\mu(T).$$

In the next section, I briefly allude to the remarkable generalization of Shannon – McMillan's entropy given by Kolmogorov in case the shift T is replaced by any measure preserving transformation of a probability space into itself.

7.3 Kolmogorov's entropy

The key to Kolmogorov's definition of the entropy of an arbitrary measure preserving map $T : (\Omega, \mathcal{F}, \mu) \to (\Omega, \mathcal{F}, \mu)$ from a probability space to itself is the translation of what we have just done into the language of finite measurable *partitions*: indeed, a random variable with finite values, a finite probability space and a finite measurable partition of a probability space are essentially the same object: this is illustrated on figure 10 where the partition

$$(\mathcal{E}) : \Omega = \Omega_1 + \Omega_2 \cdots + \Omega_r$$

is defined by $\Omega_i = \xi^{-1}(A_i)$ and the formula for entropy becomes

$$H_\mu(\mathcal{E}) = \sum_{i=1}^{r} \mu(\Omega_i) \log \frac{1}{\mu(\Omega_i)}.$$

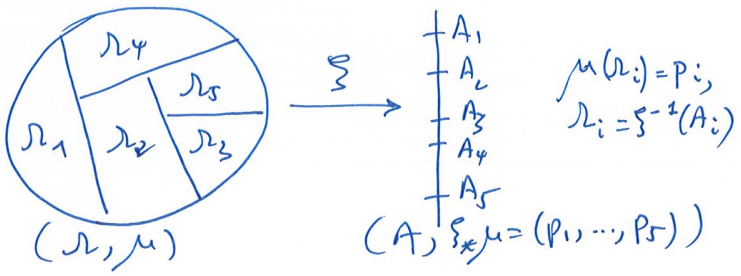


Figure 10 : finite partitions and random variables.

Now, to a measurable partition of Ω one can associate a *coding* of any map T by replacing an orbit $\{T^n(\omega)\}_{n\in\mathbb{N}}$ by the sequence $\{A_{i_n}\}_{n\in\mathbb{N}}$ of pieces of the partition which it visits, that is such that $\omega \in T^{-n}(A_{i_n})$. The finer the partition, the more faithful the coding. Such a coding replaces the transformation T by

a *stationary random process with finitely many states* whose entropy is defined by Shannon's formula.

Notations. Given a transformation $T : \Omega \to \Omega$ and a partition $\mathcal{E}$, we note $T^{-1}\mathcal{E}$ the algebra of subsets formed by the $T^{-1}(\Omega_i)$, $\Omega_i \in \mathcal{E}$. Given finite partitions $\mathcal{E}^{(1)}, \ldots, \mathcal{E}^{(m)}$ of Ω, we note $\vee_{i=1}^m \mathcal{E}^{(i)}$ the partition whose atoms are the intersections $\Omega_{k_1}^{(1)} \cap \Omega_{k_2}^{(2)} \cap \ldots \cap \Omega_{k_m}^{(m)}$, where $\Omega_{k_i}^{(i)}$ is an atom of $\mathcal{E}^{(i)}$.
In particular, if $T : A^{\mathbb{Z}}$ (or $A^{\mathbb{N}}$) is the shift, the partition into cylinders $A_{12\ldots n}^{j_1 j_2 \ldots j_n}$ can be written $\vee_{k=0}^{n-1} T^{-k}\mathcal{E}$, where the atoms of the partition $\mathcal{E}$ are the cylinders whose atoms are the cylinders $A_1^{j_1}$, that is

$$\Omega = \{\omega = \ldots a_1 a_2 \ldots |\, a_1 = A_1\} + \{\omega |\, a_1 = A_2\} + \ldots + \{\omega |\, a_1 = A_r\}$$

The best approximation to the definition of $H_\mu(T)$ for the shift contained in Lemma 62 is then

Definition 26 *The entropy $H_\mu(\mathcal{E}, T)$ of a partition $\mathcal{E}$ with respect to a measure preserving transformation $T : (\Omega, \mathcal{F}, \mu) \to (\Omega, \mathcal{F}, \mu)$ and the entropy $H_\mu(T)$ of the transformation T are respectively defined by*

$$H_\mu(\mathcal{E}, T) = \limsup_{n\to\infty} \frac{1}{n} H_\mu\big(\vee_{k=0}^{n-1} T^{-k}\mathcal{E}\big), \quad H_\mu(T) = \sup_{\mathcal{E}} H_\mu(\mathcal{E}, T),$$

where the sup is taken among all finite partitions $\mathcal{E}$ of Ω.

Explanation (see [B2]) : an element $A_{k_1}^{(1)} \cap A_{k_2}^{(2)} \cap \ldots \cap A_{k_m}^{(m)}$ of the partition $\vee_{i=1}^m \mathcal{E}^{(i)}$ may be considered as the realization of m experiments, corresponding to the m partitions $\mathcal{E}^{(i)}$. Given a partition $\mathcal{E}$, let us denote by $A = \{A_1, \ldots, A_r\}$ the set of atoms of the partition and by $x : \Omega \to A$ the random variable which, to an element $\omega \in \Omega$, associates the atom A_i to which it belongs. As T preserves the measure μ, the image measures of μ by the random variables $x \circ T^n$ are all the same (n is an integer or a relative integer if T is invertible). In other words, the experiments corresponding to the partitions $T^{-n}(\mathcal{E})$ have all the same probabilistic structure and hence they can be considered as realizations, a priori not independent, of one and the same experiment.

When $T : A^{\mathbb{Z}} \to A^{\mathbb{Z}}$ is the shift, the partition into cylinders is *generating*, which means that the sequence of partitions $\vee_{i=-n}^{n} T^{-i}(\mathcal{E})$, $n = 1, \cdots, \infty$, generate the borelian σ-algebra, which explains why the consideration of other partitions is not necessary. Indeed, more generally, the following theorem was proved by Kolmogorov and Sinai:

Theorem 64 *If $T : (\Omega, \mathcal{F}, \mu)$ is invertible and if there exists a finite partition $\mathcal{E}$ which is generating in the sense that the partitions $\vee_{i=-n}^{n} T^{-i}(\mathcal{E})$, $n = 1, \cdots, \infty$, genrate the σ-algebra $\mathcal{F}$, one has*

$$H_\mu(T) = H_\mu(\mathcal{E}, T).$$

An analogous statement holds true in the non invetible case if one replaces $\vee_{i=-n}^{n} T^{-i}(\mathcal{E})$ by $\vee_{i=0}^{n} T^{-i}(\mathcal{E})$.

7.4 Topological entropy

Replacing partitions by open covers, the expectation of $\log \frac{1}{p}$ by the log of a minimum number of pieces, it is possible by copying Kolmogorov's definition of metric entropy to define the *topological entropy* of a continuous map $\Phi : X \to X$ from a compact topological space to itself (see [AKM]) : the entropy $H_\mu(\mathcal{E})$ of a measurable finite partition is replaced by the entropy of an open cover $\mathcal{U}$:

$$H(\mathcal{U}) = \log \inf\{k|\ \exists \text{ a subcover of } \mathcal{U} \text{ with } k \text{ elements}, \}$$

and

$$H_{top}(\Phi,\mathcal{U}) = \lim_{n\to+\infty} \frac{1}{n} H(\big(\vee_{k=0}^{n-1}\Phi^{-k}\mathcal{U}\big), \quad H_{top}(\Phi) = \sup_{\mathcal{U}} H(\Phi,\mathcal{U}),$$

where existence of the limit follows, as in the metric case, from subadditivity.

The so-called *variational principle* asserts that the topological entropy $H_{top}(T)$ is the sup over all T-invariant Borel probability measures μ of the metric entropies $H_\mu(T)$.

References

[AB] A. Avila & J. Bochi, *On the subadditive ergodic theorem*, 2009

[A] V.I. Arnold, *Small denominators. I: Mappings of the circumference onto itself*, AMS Translations, Ser. 2, 46 (1965), 213-284.

[AA] V.I. Arnold & A. Avez, *Ergodic problems of classical mechanics*,

[AKM] R. Adler, A. Konheim, H. McAndrew, *Topological entropy*, Transactions of the merican Mathematical Society, 114 (1965), 309-319

[B1] P. Billingsley, *Probability and measure*, Wiley 1979

[B2] P. Billingsley, *Ergodic theory and information*,

[C0] A. Chenciner, *Perturbing a planar rotation: normal hyperbolicity and angular twist*, an introductory lecture to the course *Discrete Dynamical Systems*, Tsinghua, february 2017 `https://www.imcce.fr/fr/presentation/equipes/ASD/person/chenciner/preprint.html`

[C1] A. Chenciner, *Poincaré and the three-body problem*, in "Poincaré 1912-2012", Birkhauser 2014 `http://www.bourbaphy.fr/novembre2012.html`

[C2] A. Chenciner, *The planar circular restricted three body problem in the lunar case*, minicourse at the Chern Institute of Mathematics, Nankai University, may 2014 `https://www.imcce.fr/fr/presentation/equipes/ASD/person/chenciner/polys.html`

[C3] A. Chenciner, *Shannon's theorems: the strength of a simple idea*, minicourse at Universidade Catolica del Norte (Antofagasta, Chile) and at Capital Normal University (Beijing) `https://www.imcce.fr/fr/presentation/equipes/ASD/person/chenciner/polys.html` Sept. 2016

[CFS] I.P. Cornfeld, S.V. Fomin & Ya. G. Sinai, *Ergodic theory*, Springer 1982, Birkhauser 2014

[CT] T.C. Cover & J.A. Thomas, *Elements of Information Theory*, Wiley 1991

[Fa] A. Fathi, *Systèmes dynamiques*, Cours de l'École Polytechnique 1996

[F] W. Feller, *An introduction to probability theory and its applications*, vol. 1, Wiley

[H] P. R. Halmos, *Lectures on ergodic theory*, Chelsea 1956

[H2] P. R. Halmos, *Introduction to Hilbert spaces*, Chelsea 1957

[He1] M.R. Herman, *Sur la conjugaison différentiable des difféomorphismes du cercle à des rotations*, Pub. math. de l'I.H.É.S., tome 49 (1979).

[He2] M.R. Herman, *Mesure de Lebesgue et nombre de rotation*, Geometry and Topology (Proc. III Latin Amer. School of Math., IMPA, Rio de Janeiro, 1976) (Lecture Notes in Mathematics, 597) Springer, Berlin, 1977

[K] A. Katok, *Fifty Years of Entropy in Dynamics: 1958-2007*, Journal of Modern Dynamics, vol. 1, $n^0 4$, 2007, 545-596

[KH] A. Katok & B. Hasselblatt, *Introduction to the Modern Theory of Dynamical Systems*, Cambridge University Press 1995, section 4.1

[Kh1] A.I. Khinchin, *Mathematical foundations of information theory*, Dover 1957

[Ko] N. A. Kolmogorov, *Foundations of the theory of probabilities*, Chelsea 1960 (première édition, en allemand en 1933 sous le titre Grundbegriffe der Wahrscheinlichkeitrechnung)

[Ku] S. Kullback, *Information theory and statistics*, Wiley 1959, Dover 1968

[LC] P. Le Calvez, *Introduction to dynamical systems*, Tsinghua (2014).

[M1] R. Mañé, *Ergodic theory and Differentiable Dynamics*, Springer 1987

[O] J.C. Oxtoby, *Measure and Category*, Springer 1971

[Pe] K. Petersen, *Ergodic Theory*, Cambridge University Press 1983

[Po] H. Poincaré, *Les méthodes nouvelles de la mécanique céleste*, tome III, chapitre XXVI "Stabilité à la Poisson", Gauthier-Villars 1899, retirage librairie scientifique et technique Albert Blanchard 1987

[R] W. Rudin, *Real and complex analysis*, McGraw-Hill 1966

[Sh] C. Shannon, *A Mathematical Theory of Communication*, Bell System Technical Journal, vol. 27, pp. 379-423, 623-656, July, October 1948

[Si] Y Sinai, *Probability theory, an introductory course* (Moscou, 1985-1986, Springer 1992)

[St] H. Steinhaus, *Les probabilités dénombrables et leur rapport à la théorie de la mesure*, Fund. Math. 4, p. 286-310 (1923)

[T] T. Tao, *Poincaré's legacies, Part I*, AMS 2009

[Ta] S. Tabachnikov, *Billiards*, S.M.F. Panoramas et Synthèses 1 (1995)