Systèmes Dynamiques
Notes du cours de M2

Raphaël KRIKORIAN
Université Paris 6

Année 2011-2012

Table des matières

Chapitre 1

Introduction

1.1 Systèmes dynamiques

Un *système dynamique* (ou plus simplement une *dynamique*) est la donnée d'un ensemble X (l'espace des phases) et d'un groupe ou d'un semi-groupe G agissant [1] sur X. La plupart du temps, le groupe est soit le groupe additif $(\mathbb{Z}, +)$ et on parle alors de système dynamique *discret* soit le groupe additif $(\mathbb{R}, +)$ et la dynamique est dite *continue*. On peut aussi considérer des dynamiques associées à d'autres groupes, par exemple des groupes de Lie comme $SL(2, \mathbb{R})$ agissant sur le demi-plan de Poincaré par homographies [2]. Les dynamiques non-inversibles sont également intéressantes, mais alors l'action sur X est seulement associée à un semi-groupe.

Une action du semi-groupe additif $(\mathbb{N}, +)$ sur X est équivalente à l'itération d'une application $T : X \to X$: si $T(x) = 1{\cdot}x$ on a $n{\cdot}x = T^n(x)$ (T^n représente l'itéré n-ième de T) ; T est inversible si et seulement si l'action précédente s'étend naturellement en une action de $(\mathbb{Z}, +)$. Cet exemple sera le principal exemple que nous considérerons dans ce cours. Nous parlerons alors du système dynamique (X, T).

Une action du groupe additif $(\mathbb{R}, +)$ sur un ensemble X est équivalente à l'action d'une famille à un paramètre de bijection $(T^t)_{t\in\mathbb{R}}$ sur X : pour

1. L'action d'un groupe G sur un ensemble X est une application $G \times X \to X$, $(g, x) \mapsto g \cdot x$ telle que pour tout $(g, h) \in G^2$ et tout $x \in X$ on ait $g \cdot (h \cdot x) = (gh) \cdot x$ et si $e \in G$ est l'élément neutre $e \cdot x = x$; quand G et X sont munis d'une structure topologique et quand l'application $(g, x) \mapsto g \cdot x$ est continue, on dit que l'action est continue.

2. Le groupe $SL(2, \mathbb{R})$ est le groupe des matrices 2×2 à coefficients réels et de déterminant 1, tandis que le demi-plan de Poincaré $\mathcal{H}$ est l'ensemble des nombres complexes $z = x + iy$, où $y > 0$; si $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL(2, \mathbb{R})$, l'action est donnée par $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot z = \frac{az+b}{cz+d}$.

tous $t, s \in \mathbb{R}$, on a $T^t \circ T^s = T^{t+s}$. Des exemples de telles dynamiques sont naturellement donnés par des flots d'équations différentielles sur des variétés [3]

Afin d'obtenir des objets mathématiques dont l'étude est intéressante, il est nécessaire d'ajouter des structures supplémentaires à notre dynamique. Par exemple :

– Si X est un espace topologique et $T : X \to X$ est continue, nous parlerons de dynamique topologique.

– Quand $(X, \mathcal{B}, \mu)$ est un espace de probabilité (où $\mathcal{B}$ est une σ-algèbre de X et μ une probabilité sur $\mathcal{B}$) et T préserve cette structure : $T : X \to X$ est mesurable (c'est-à-dire pour tout $B \in \mathcal{B}$, $T^{-1}(B) \subset \mathcal{B}$) et préserve μ (c'est-à-dire pour tout $B \in \mathcal{B}$ on a $\mu(T^{-1}(B)) = \mu(B)$). L'étude de ces systèmes est l'objet de la théorie ergodique.

– Quand X est une variété et T est une application différentiable ; on parle alors de dynamique différentiable.

– *etc.*

1.2 Ensembles invariants

Un ensemble $A \subset X$ est dit *invariant* par $T : X \to X$ si $T(A) \subset A$, *fortement invariant* si $T(A) = A$ et *complètement invariant* si $T^{-1}(A) = A$. Quand T est une bijection ces deux dernières notions coïncident.

Construire des ensembles invariants est facile. Par exemple, si $x \in X$ alors son *orbite positive* $\mathcal{O}_+(x) = \{T^n(x) : n \in \mathbb{N}\}$ est un ensemble invariant qui en général n'est pas fortement invariant. Si T est inversible, l'orbite totale de x, $\mathcal{O}(x) = \{T^n(x) : n \in \mathbb{Z}\}$ est complètement invariante.

Un cas particulier important est quand l'orbite du point x est finie. Dans ce cas on dit que x est *pré-périodique* et il est facile de voir qu'il existe des entiers $p, k \geq 0$ tels que $T^p(T^k(x)) = T^k(x)$: ainsi pour tout $m \in \mathbb{N}$, $T^{k+mp}(x) = T^k(x)$. Quand $k = 0$, on dit que le point x est *périodique*. Le plus petit entier positif pour lequel $T^p(x) = x$ est appelé la *période* de x.

Exercice Soit $S : \mathbb{R}/\mathbb{Z} \circlearrowleft$ définie par $Sx = 2x$. Montrer que S est bien définie et montrer que pour tout entier n, $x = p/(2^n - 1)$ $(0 \leq p < 2^n - 1)$ est un point périodique de période n pourvu que $2^n - 1$ et p soient premiers entre eux. (Il s'agit de démontrer que les points $x, Sx, \ldots, S^{n-1}x$ sont distincts, ce

3. Si X est une variété (par exemple $\mathbb{R}^d$) et si F est un champ de vecteurs sur X, le temps t de F est (quand il est défini) l'application $\phi^t : X \to X$ qui à toute condition initiale $x \in X$ au temps $t = 0$ associe la valeur de la solution de l'équation différentielle $x'(t) = F(x(t))$, $x(0) = x$, au temps t.

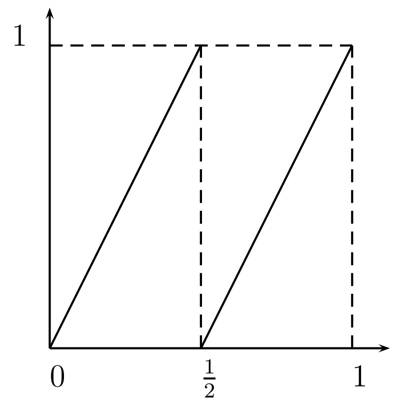

FIGURE 1.1 – L'application $T : x \mapsto 2x$ sur $[0, 1]$

qui revient à démontrer que $(2^b - 1)/(2^a - 1)$ n'est pas entier si a ne divise pas b ; on pourra faire la division euclidienne $b = qa + r$).

Dans le contexte de la dynamique topologique, on peut essayer de trouver des ensembles invariants qui ont en plus certaines propriétés topologiques, par exemple le fait d'être fermés.

1.3 Facteurs et problèmes de conjugaison

Une façon de comprendre une dynamique donnée est d'essayer de la réduire à une forme plus simple.

Nous dirons qu'un système dynamique (Y, S) est un *facteur* d'un système dynamique (X, T) s'il existe une application surjective $\pi : X \to Y$ (la "projection") telle que $S \circ \pi = \pi \circ T$. Observons que dans ce cas pour tout entier $n \in \mathbb{N}$ on a $S^n \circ \pi = \pi \circ T^n$. Dans un certain sens cela signifie que (Y, S) est une simplification de la dynamique de (X, T), une partie de l'information sur (X, T) étant perdue lors de ce procédé (si π n'est pas bijective). Quand π est bijective et $Y = X$, nous dirons que (X, T) est *conjugué* à (Y, S). Un problème général, qui a rarement une solution, est d'essayer de classifier les systèmes dynamiques à conjugaison près.

Chapitre 2

Dynamique Topologique

2.1 Espaces topologiques

2.1.1 Définitions (Rappels)

Un espace topologique $(X,\mathcal{U})$ est la donnée d'un ensemble X est d'une *topologie* $\mathcal{U}$ sur X ce qui signifie que $\mathcal{U}$ est une collection de sous-ensembles de X telle que : i) $\emptyset$ et X appartiennent à $\mathcal{U}$; ii) pour tous $U_1, U_2 \in \mathcal{U}$ on a $U_1 \cap U_2 \in \mathcal{U}$ et iii) pour toute famille $(U_i)_{i\in I}$, $U_i \in \mathcal{U}$ on a $\bigcup_{i\in I} U_i \in \mathcal{U}$. Les éléments U de $\mathcal{U}$ sont appelés les *ouverts* de X. Par définition un ensemble $F \subset X$, est *fermé* si son complémentaire $F^c = X \setminus F$ est ouvert. Si $(X,\mathcal{U})$ et $(Y,\mathcal{V})$ sont deux espaces topologiques, et $T : X \to Y$ est une application, on dit que T est *continue* si pour tout ouvert $V \in \mathcal{V}$, l'ensemble $T^{-1}(V) := \{x \in X : T(x) \in V\}$ de ses pré-images par T est un ouvert de X. Si T et bijective et T, T^{-1} sont toutes deux continues, on dit que T est un *homéomorphisme* de $(X,\mathcal{U})$ sur $(Y,\mathcal{V})$.

Si x est un point de X on dit que W est un *voisinage* de x s'il existe un ouvert $U \subset X$ contenant x et inclus dans W.

Une collection $\mathcal{C}$ de sous-ensembles de X est appelée une base de voisinage de X si tout $x \in X$ admet un élément de $\mathcal{C}$ comme voisinage.

Si A est un sous-ensemble de X, son *adhérence* est le sous-ensemble $\bar{A}$ de X défini de la façon suivante : $x \in \bar{A}$ si et seulement si pour tout voisinage W de x on a $W \cap A \neq \emptyset$. Il est facile de vérifier que c'est le plus petit (pour l'inclusion) ensemble fermé de X contenant A.

De façon similaire, *l'intérieur* $\mathring{A}$ de A (qui peut être vide) est l'ensemble défini de la façon suivante : $x \in \mathring{A}$ si et seulement si, il existe un voisinage W de x tel que $W \subset A$. L'ensemble ouvert $\mathring{A}$ est le plus grand (pour l'inclusion) ouvert de X contenant A.

Un ensemble $D \subset X$ est dense dans X si $\bar{D} = X$.

Un espace topologique est dit de *Hausdorff* ou *séparé* si pour tous $x, y \in X$ il existe des voisinages U et V de x et y respectivement qui sont disjoints. Dans ce cas le singleton $\{x\}$ est un fermé.

Un espace topologique est *séparable* si on peut trouver une base dénombrable de voisinage de cet espace.

Si Y est un sous-ensemble de X et si $(X, \mathcal{U})$ est un espace topologique, la *topologie induite* sur Y est la collection de tous les ensembles de la forme $Y \cap U$ où $U \in \mathcal{U}$ décrit tous les ouverts de X.

Un espace topologique est dit *compact* si pour tout recouvrement ouvert de X par des ouverts U_i, $i \in I$ ($X = \bigcup_{i \in I} U_i$) on peut extraire un sous-recouvrement fini : il existe un nombre fini d'ouverts $U_{i_1}, \ldots, U_{i_p}$ tels que $X = U_{i_1} \cup \cdots \cup U_{i_n}$. Un sous-ensemble K d'un espace compact X est compact si et seulement si par définition il est compact pour la topologie induite. De la définition de la compacité, il est facile de voir que *l'intersection d'une suite décroissante (pour l'inclusion) de compacts non-vides est un compact non-vide.* Tout ensemble fermé d'un compact est compact et si X est de Hausdorff alors si $K \subset X$ est compact, il est fermé.

N.B. Désormais, quand nous parlerons d'espace compact nous sous-entendrons que cet espace est de Hausdorff.

Un résultat utile est le suivant : *dans un espace compact, tout ouvert non vide contient un compact d'intérieur non vide.* [**Exercice** : Plus précisément démontrer que si X est compact, V est un ouvert de X et $x \in V$ il existe un ouvert Ucontenant x et un compact K tel que $x \in U \subset K \subset V$.]

Dans ces notes nous considérerons des espaces topologiques qui sont des espaces métriques. Un *espace métrique* (X, d) est la donnée d'un ensemble X et d'une *distance* $d : X \times X \to [0, \infty[$ telle que : i) pour tous $x, y \in X$ $d(x, y) = 0$ si et seulement si $x = y$; ii) (symétrie) pour tous $x, y \in X$, $d(x, y) = d(y, x)$; iii) (inégalité triangulaire) pour tous $x, y, z \in X$, $d(x, z) \leq d(x, y) + d(y, z)$. Si $x \in X$ nous noterons $B(x, \delta)$ la *boule* ouverte de centre x et de rayon δ : $B(x, \delta) := \{y \in X : d(x, y) < \delta\}$. La collection $\mathcal{U}$ des ensembles $U \subset X$ tels que $\forall x \in U$, $\exists \delta > 0$, $B(x, \delta) \subset U$, est une topologie sur X. D'après i) il est clair qu'un espace métrique est toujours séparé. Un espace métrique est séparable s'il existe un sous ensemble D dense et dénombrable de X ; par exemple, la collection des boules ouvertes $(B(x, r))_{x \in D, r \in \mathbb{Q}^+}$ est dans ce cas une base dénombrable de voisinages de X.

Sur un espace métrique (X, d) les énoncés suivants sont équivalents : "$K \subset X$ est compact" et "de toute suite de points de K on peut extraire une sous-suite convergente". La preuve de ce résultat est basé sur le Lemme de Lebesgue qui est intéressant par lui-même : *Si* (X, d) *est compact et* $(U_i)_{i \in I}$ *est un recouvrement ouvert de* X*, il existe un* $\delta > 0$ *tel que toute boule ouverte* *de rayon $\delta > 0$ est incluse dans au moins un des U_i.*

Un espace métrique (X, d) est *complet* si toute suite de Cauchy converge[1] Une conséquence de cette définition est que sur un espace complet, toute suite décroissante (pour l'inclusion) d'ensembles fermés dont le diamètre[2] tend vers zéro admet une intersection non vide.

Nous utiliserons dans ce chapitre le puissant Lemme de Baire. Un ensemble O (resp. F) est un G_δ (resp. F_σ) s'il est une intersection (resp. union) dénombrable d'ouverts (resp. de fermés).

Théorème 2.1.1 *Soit (X, d) un espace compact ou un espace métrique complet. Alors, l'intersection dénombrable de G_δ (en particulier d'ouverts) qui sont denses est un G_δ lui aussi dense. De même, l'union dénombrable d'une famille de F_σ (en particulier de fermés) d'intérieur vide est un F_σ d'intérieur vide.*

Démonstration. — Nous allons démontrer qu'une intersection $O = \bigcap_{i\in\mathbb{N}} U_i$ dénombrable d'ouverts dense U_i, $i \in \mathbb{N}$, est elle-meme dense. Le cas général et celui des F_σ s'en déduit. Nous devons démontrer que pour tout ouvert V de X, $O \cap V \neq \emptyset$. Puisque U_0 est dense, $U_0 \cap V$ est un ouvert non-vide et il existe, dans le cas où X est compact (resp. complet) un compact $K_0 \subset U_0 \cap V$ d'intérieur non-vide (resp. une boule ouverte B_0 non vide de rayon inférieur à 1 telle que $\bar{B}_0 \subset U_0 \cap V$). Comme U_1 est dense, l'ouvert $U_1 \cap \mathring{K}_0$ est non vide (resp. $B_0 \cap U_1 \neq \emptyset$) et il existe un compact $K_1 \subset U_1 \cap \mathring{K}_0$ d'intérieur non vide (resp. une boule ouverte B_1 de rayon $\leq 1/2$ telle que $\bar{B}_1 \subset B_0 \cap U_1$). En itérant ce procédé, nous pouvons construire une suite de compacts K_n d'intérieurs non vides telle que $U_{n+1} \cap \mathring{K}_n \supset K_{n+1} \supset \mathring{K}_{n+1} \neq \emptyset$ (resp. une suite de boules ouvertes B_n de rayon $\leq 1/(n+1)$ telle que $\bar{B}_{n+1} \subset B_n \cap U_n$). La suite K_n (resp. $\bar{B}_n$) est une suite décroissante de compacts (resp. fermés) non vide (resp. dont le diamètre tend vers 0) et donc admet une intersection non vide. En outre, pour tout n cette intersection est dans $U_n \cap V$. Nous avons donc démontré que $O \cap V$ est non vide. □

2.1.2 Exemples

Un exemple d'espace métrique (complet) est bien évidemment $\mathbb{R}^d$ muni d'une des métriques suivantes : si $x = (x_1, \ldots, x_d)$, $y = (y_1, \ldots, y_d)$, $d_\infty(x, y) = \max_{1\leq i\leq d} |x_i - y_i|$, $d_p(x, y) = \left(\sum_{i=1}^d |x_i - y_i|^p\right)^{1/p}$ où $p \in [1, \infty[$.

1. c'est une suite (x_n telle que pour tout $\epsilon > 0$ il existe un entier N tel que pour tous entiers $n, m \geq N$ on ait $d(x_n, x_m) < \epsilon$.
2. Le diamètre d'un ensemble A est $\sup_{x,y\in A} d(x, y)$.

D'autres exemples importants d'espaces topologiques sont donnés par des constructions faisant intervenir des recollements ou des passages au quotient.

Relation d'équivalence et topologie. Soient $(X, \mathcal{U})$ un espace topologique et $\sim$ une relation d'équivalence sur X. Nous notons $\tilde{X}$ l'ensemble des classes d'équivalence de $\sim$ et $\pi : X \to \tilde{X}$ la projection canonique qui à tout $x \in X$ associe sa classe d'équivalence dans $\tilde{X} = X/\sim$. Nous dirons qu'un ouvert U de X est *saturé* si $x \in U, y \sim x$ implique $y \in U$ ou de façon équivalente si $\pi^{-1}(\pi(U)) = U$. La collection $\tilde{U}$ des ensembles de $\tilde{X}$ de la forme $\pi(U)$ où U est un ouvert saturé de X est une topologie sur $\tilde{X}$. C'est la plus fine *i.e.* la plus grande (pour l'inclusion) topologie sur X pour laquelle l'application $\pi : X \to \tilde{X}$ est continue. Remarquons que dans certains cas cette topologie peut-être triviale, c'est-à-dire réduite à $\{\emptyset, X\}$ [3]

Illustrons la construction précédente dans le cas du tore de dimension d. Par définition c'est le quotient $\mathbb{R}^d/\sim$ où $x \sim y$ si et seulement si $x - y \in \mathbb{Z}^d$. De façon plus algébrique c'est l'espace quotient $\mathbb{R}^d/\mathbb{Z}^d := \{x + \mathbb{Z}^d : x \in \mathbb{R}^d\}$ du groupe additif $(\mathbb{R}^d, +)$ par son sous-groupe $\mathbb{Z}^d$. Les ensembles ouverts saturés de $\mathbb{R}^d$ sont les ensembles de la forme $U + \mathbb{Z}^d = \bigcup_{k \in \mathbb{Z}^d}(U + k)$ où U est un ouvert de $\mathbb{R}^d$ et par conséquent les ouverts de $\mathbb{R}^d/\mathbb{Z}^d$ sont les ensembles de la forme $U + \mathbb{Z}^d = \{x + \mathbb{Z}^d : x \in \mathbb{Z}^d\}$.

On peut en fait munir l'espace $\mathbb{T}^d$ d'une distance naturelle compatible avec la topologie précédente. La distance euclidienne plate $(dx_1)^2 + \cdots + (dx_d)^2$ sur $\mathbb{R}^d$ ne dépend pas du point de $\mathbb{R}^d$ où elle est évaluée et se projette ainsi sur $\mathbb{T}^d$; elle définit donc une métrique sur $\mathbb{T}^d$ et partant une distance géodésique. De façon plus concrète, la distance entre $\bar{x}, \bar{y} \in \mathbb{T}^d$ est $\min_{k \in \mathbb{Z}^d} d(x + k, y + l)$ $(x, y$ sont des représentants quelconques de $\bar{x}, \bar{y} : \pi(x) = \bar{x}, \pi(y) = \bar{y})$.

Dans l'exemple précédent, quand par exemple $d = 1$, on peut donner une autre construction de $\mathbb{T}^1$ (le cercle). L'espace $X = [0, 1]/\sim$, où $\sim$ est la relation d'équivalence dont les classes sont tous les ensembles (singletons) $\{x\}$ si $x \in]0, 1[$ et la paire $\{0, 1\}$, est ainsi homéomorphe à $\mathbb{T}^1$ quand on le munit de la topologie quotient associée à cette relation d'équivalence. Cette construction se généralise à $d > 1$ (Comment ?). [4]

Exercice :

a) Démontrer que X construit de la façon précédente, muni de la topologie quotient que l'on décrira, est homéomorphe à $\mathbb{R}/\mathbb{Z}$.

b) Définissons $T : X \to X$ par $Tx = 2x$ si $0 \leq x \leq 1/2$, et $Tx = 2x - 1$ si $1/2 \leq x \leq 1$. Prouver que T est bien définie et continue.

---

3. Ce sera le cas si on considère par exemple $X = \mathbb{R}$ et la relation $x \sim y$ si et seulement si $x - y$ est un rationnel.

4. Il existe une troisième façon de construire $\mathbb{T}^1$ (mais qui ne se généralise pas à la dimension plus grande) : $\mathbb{T}^1 \simeq \{z \in \mathbb{C} : |z| = 1\}$.

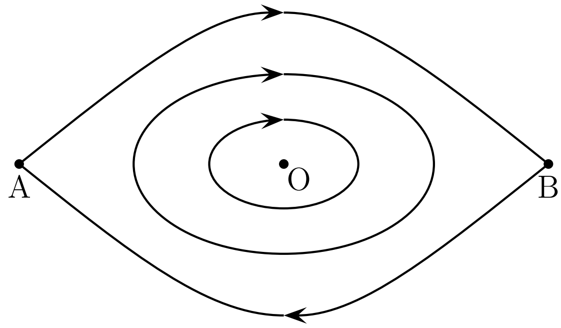

FIGURE 2.1 – $R(T)$ n'est pas fermé

c) T est-elle inversible ?

N.B. : Dans la suite nous supposerons que X est un espace compact, et $T : X \to X$ une application continue. Nous ferons souvent l'hypothèse que X est compact, métrique, séparable.

2.2 Récurrence

Nous introduisons à présent quelques notions relatives aux propriétés de récurrence d'un système dynamique topologique.

Point périodique Un point $x \in X$ est périodique s'il existe $n_0 \in \mathbb{N}^*$ tel que $T^{n_0}(x) = x$. Nous noterons $P_n(T)$ l'ensemble des points fixes de T^n : c'est un compact ; de même nous noterons $P(T) = \bigcup_{n\in\mathbb{N}} P_n(T)$. L'adhérence $\overline{P(T)}$ de $P(T)$ est évidemment un compact invariant par T. Noter que $P(T)$ peut être vide.

Point récurrent Un point x est dit récurrent ssi

$$\inf_{n\geq 1} d(x, T^n(x)) = 0.$$

L'adhérence $\overline{R_+(T)}$ de l'ensemble $R_+(T)$ des points récurrents est un compact invariant. Si T est inversible, on définit $R_-(T) = R_+(T^{-1})$ et $R(T) = R_-(T) \cup R_+(T)$. On verra dans la section suivante que si X est compact, $R_+(T)$ n'est jamais vide.

Exercice Prouver qu'en général l'ensemble $R(T)$ n'est pas fermé.

Ensembles α et ω-limite Si $x \in X$, l'ensemble des points ω-limite de x est l'ensemble $\omega(x)$ des $y \in X$ qui sont points d'accumulation de la suite $(T^n(x))_{n\geq 0}$: $\omega(x) = \bigcap_{N\geq 0} \overline{\{T^k(x) : k \geq N\}}$. Remarquons que $x \in R_+(T)$ si et seulement si $x \in \omega(x)$. Par définition, l'ensemble ω-limite de T est

$$L_+(T) = L_+(T) = \overline{\bigcup_{x\in X} \omega(x)}.$$

C'est un compact invariant. Si T est inversible on définit également l'ensemble α-limite de x qui est l'ensemble des points d'accumulation de la suite $(T^n(x))_{n\leq 0}$. Si T est inversible, nous notons

$$L(T) = \overline{\left(\bigcup_{x\in X} \omega(x) \cup \alpha(x)\right)}.$$

Point non errant Un point x_0 est dit errant ssi il existe un voisinage U de x_0 et un $N \in \mathbb{N}$ tels que pour tout $n \geq N$, $T^n(U) \cap U = \emptyset$; un point est non-errant s'il n'est pas errant. On note $\Omega(T)$ l'ensemble des points non errants. Si T est inversible $\Omega(T^{-1}) = \Omega(T)$.

Point récurrent par chaînes Un point x est récurrent par chaînes si et seulement si pour tout $\epsilon > 0$ il existe une suite x_i, $0 \leq i \leq n$ telle que $x_0 = x = x_n$ et telle que pour tout $0 \leq i \leq n-1$, on a $d(x_{i+1}, T(x_i)) \leq \epsilon$.

Exercice a) Prouver les inclusions

$$\overline{P(T)} \subset \overline{R(T)} \subset L(T) \subset \Omega(T) \subset C(T).$$

b) Trouver des exemples où ces inclusions sont strictes.

2.3 Irréductibilité

2.3.1 Minimalité

Notons $\mathcal{F}$ l'ensemble des compacts non-vides $K \subset X$ qui sont T-invariants. Un compact K est dit *minimal* s'il est minimal dans $\mathcal{T}$ pour l'inclusion (si $L \subset K$ avec L compact et $T(L) \subset L$ alors $L = K$). On dit que X est *minimal* s'il ne contient aucun compact non vide invariant autre que lui-même.

L'intérêt de cette notion réside dans la proposition suivante :

Proposition 2.3.1 *Un compact $K \subset X$ est minimal ssi pour tout $x \in K$, $\overline{\mathcal{O}_+(x)} = K$.*

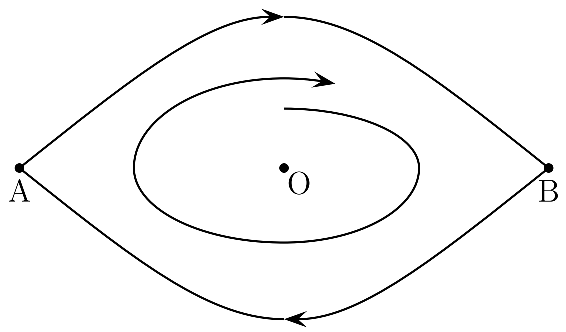

FIGURE 2.2 – Qu'est $\Omega(T)$?

Démonstration. — Supposons que K soit minimal. Alors pour tout $x \in K$ le compact $\overline{\mathcal{O}_+(x)} = K$ est T-invariant non vide inclus dans K et égale donc K. Réciproquement si pour tout $x \in K$ on a $\overline{\mathcal{O}_+(x)} = K$, alors si $L \subset K$ est un compact invariant on a pour $x \in L$, et tout $n \geq 0$, $T^n(x) \in L$ et partant $K = \overline{\mathcal{O}_+(x)} \subset L$ c'est-à-dire $L = K$. □

Le principal résultat est :

Théorème 2.3.1 *Si (X, d) est un espace métrique compact et $T : X \to X$ est continue, il existe toujours un compact $K \subset X$ minimal.*

Démonstration. — Nous aurons besoin pour la preuve du lemme de Zorn

Lemme 2.3.1 (Zorn) *Si $\mathcal{F}$ est un ensemble muni d'une relation d'ordre, il existe un sous-ensemble totalement ordonné et maximal pour cette propriété.*

Ainsi, $(\mathcal{F}, \subset)$ contient un sous-ensemble totalement ordonné maximal $\mathcal{G}$. Le compact $K = \cap_{K \in \mathcal{G}} L$ est non-vide, T-invariant et minimal. □

Définition 2.3.1 *On dit que T est minimale sur X ou que X est minimal pour T si X est T-minimal.*

Remarque : Si T est minimale et si X est infini alors T n'admet pas d'orbite périodique.

Exemples : a) Soit $X = \mathbb{R}/\mathbb{Z}$ le tore et $T(x) = x + \alpha$. Montrons que T est minimale si et seulement si $\alpha \notin \mathbb{Q}/\mathbb{Z}$. Si $\alpha = p/q$ il est clair que tout point est périodique ($T^q(x) = x \mod 1$). Réciproquement supposons α irrationnel : alors la suite $n\alpha$ est dense sur $\mathbb{T}$. En effet, la suite $n\alpha$ prend une infinité de valeurs sur le compact X et admet donc un point d'accumulation, disons x_0, c'est-à-dire qu'il existe une suite n_k strictement croissante telle que $n_k\alpha$ converge vers x_0. Ainsi, pour tout $\epsilon > 0$, il existe N tel que, dès que

$n_k > n_l \geq N$, on a $d(n_k\alpha, n_l\alpha) < \epsilon$ soit : $d(m\alpha, 0) < \epsilon$ où $m = n_k - n_l$. Il est facile de voir que si $z \in \mathbb{T}$ vérifie $d(z, 0) < \epsilon$ alors pour tout $x \in \mathbb{T}$ il existe $r \in \mathbb{N}$ tel que $d(x, rz) < \epsilon$. Si on pose $z = m\alpha$, on a ainsi $d(x, (rm)\alpha) < \epsilon$. Terminons la preuve de la minimalité : si x et y sont deux points il existe $r' \in \mathbb{N}$ tel que $d(y - x, r'\alpha) < \epsilon$ et donc $d(y, x + r'\alpha) < \epsilon$.

b) Sur $X = \mathbb{R}/\mathbb{Z}$, $f(x) = 2x$ n'est pas minimale car elle admet des points périodiques (les $k/(2^n - 1)$).

c) Un homéomorphisme de $\mathbb{R}$ n'est jamais minimal

d) Un homéomorphisme de $\mathbb{R}^2$ n'est jamais minimal (c'est une conséquence du théorème de translation de Brouwer : si un homéomorphisme de $\mathbb{R}^2$ est sans point fixe alors tout point est errant).

2.3.2 Transitivité

Dans ce qui suit nous notons $Acc(A)$ l'ensemble des points d'accumulation de l'ensemble A.

Définition 2.3.2 *On dit que T est transitive sur X si et seulement s'il existe un point $x \in X$ tel que $Acc(\mathcal{O}_+(x)) = X$. (Si T est inversible il est équivalent de dire que l'orbite de x est dense dans X).*

Voici une proposition qui justifie la terminologie :

Proposition 2.3.2 *Supposons que X soit compact séparable. Alors, T est transitive sur X si et seulement si pour tous ouverts U et V de X il existe n tel que $T^{-n}(V) \cap U \neq \emptyset$ (c'est-à-dire il existe un point de U qui visite un point de V).*

Démonstration. — Supposons donc qu'il existe un point $z \in X$ tel que l'ensemble des points d'accumulation de l'orbite positive de z soit X et soient U, V deux ouverts de X. Il existe donc $n > m$ entiers positifs tels $T^m(z) \in U$ et $T^{n-m}(T^m(z)) = T^n(z) \in V$. On a donc $T^m(z) \in U \cap T^{-(n-m)}(V)$.

Réciproquement, supposons que pour tous ouverts U, V de X il existe n tel que $U \cap T^{-n}(V) \neq \emptyset$ (appelons cela, la propriété de visite). Remarquons qu'en fait il existe une infinité de tels entiers n. Pour démontrer cela supposons tout d'abord que X soit sans point isolé[5]. Dans ce cas, fixons $x \in U$ et N un entier positif. Comme X est sans points isolés, il existe un voisinage ouvert $W \subset V$ de $\bigcup_{0 \leq k \leq N}\{T^k(x)\} \cap V$ tel que $V \setminus W \neq \emptyset$; on sait alors (cf. exercice dans les rappels sur la compacité) qu'il existe un compact K tel que $\bigcup_{0 \leq k \leq N}\{T^k(x)\} \cap V \subset \mathring{K} \subset K \subset W$. Comme les applications T^l,

5. On dit que $x \in X$ est isolé s'il existe un voisinage de ce point réduit à $\{x\}$.

$0 \leq l \leq N$ sont continues, il existe un voisinage ouvert $U_1 \subset U$ de x tel que $\bigcup_{l=0}^{N} T^l(U_1) \cap V \subset \mathring{K}$. Appliquons alors la propriété de visite aux ouverts non vides U_1 et $V_1 = V \setminus K$: il existe un entier $n \geq 0$ tel que $T^{-n}(V_1) \cap U_1 \neq \emptyset$. Mais, par construction, on a automatiquement $n > N$. Cela démontre bien que la propriété de visite a lieu pour des entiers aussi grands que l'on veut, au moins si X est sans point isolé. Supposons à présent que X admette un point isolé. Alors, ce point est automatiquement périodique et X est égal à l'orbite (finie donc fermée) de ce point périodique. En effet, $\{x\}$ est alors un ouvert et la propriété de visite montre qu'il existe n tel que $T^n(x) = x$. Soit V un ouvert quelconque de X ; la propriété de visite appliquée aux ouverts $\{x\}$ et V montre que $\mathcal{O}(x) \cap V \neq \emptyset$. Par conséquent cette orbite fermée est dense dans X et donc égale à X.

De ce qui précède il résulte que pour tout N l'ouvert

$$\bigcup_{n \geq N} T^{-n}(B)$$

est dense. Notons $\mathcal{B}$ une base dénombrable de voisinage de X. L'ensemble

$$\bigcap_{B \in \mathcal{B}} \bigcap_{N \geq 0} \bigcup_{n \geq N} T^{-n}(B)$$

est l'ensemble des points dont l'ensemble des points d'accumulation de l'orbite positive égale X. Mais, c'est une intersection dénombrable d'ouverts denses et d'après la propriété de Baire (X est compact) c'est un ensemble dense, donc en particulier non vide. $\square$

Remarque La preuve du théorème précédent montre que l'ensemble des points z dont l'orbite est dense est en fait un G_δ-dense de X.

Exercice : a) Montrer que T est minimale si et seulement si pour tout ouvert U de X il existe N tel que

$$X = \bigcup_{i=0}^{N} T^{-i}(U).$$

b) Montrer que T est transitive si et seulement pour tout ouvert U,

$$\bigcup_{i=0}^{\infty} T^{-i}(U)$$

est dense dans X.

c) Montrer que si T est transitive alors pour toute fonction continue $\phi : X \to \mathbb{R}$, $\phi \circ T = \phi$ implique que ϕ est constante. Que dire de la réciproque ?

d) Montrer qu'une action isométrique est transitive si et seulement si elle est minimale.

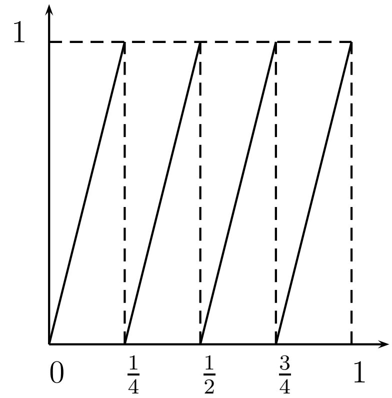

FIGURE 2.3 – L'application $T^2 : x \mapsto 2^2 x$ sur $[0,1]$

2.3.3 Mélange topologique

Définition 2.3.3 *On dit que T est topologiquement mélangeante si pour tous ouverts U, V de X il existe N tel que pour tout $n \geq N$, $T^{-N}(U) \cap V \neq \emptyset$.*

Remarque : Le mélange topologique, tout comme la minimalité, entraîne la transitivité, l'inverse étant faux.

Exemples a) Sur $X = \mathbb{R}/\mathbb{Z}$, $T(x) = 2x$ est topologiquement mélangeante (et donc transitive). En effet, il suffit de démontrer que pour tous intervalles dyadiques I et J, il existe N tel que pour tout $n \geq N$, $T^{-n}(I) \cap J$ est non vide. Or, dès que $n \geq N$, $T^{-n}(I)$ a une intersection non vide avec tout intervalle dyadique de longueur 2^{-N} donc avec J si $|J| > 2^{-N}$.

b) Une translation T_α sur $\mathbb{R}/\mathbb{Z}$ n'est jamais faiblement mélangeante : si α est rationnel c'est clair car T_α n'est pas transitive ; si α est irrationnel soient I et J deux intervalles de longueur 1/4 par exemple et notons I_0 un intervalle de longueur 1/4 disjoint de J. Comme T_α est minimale, on sait que $T_\alpha^{-n}(I)$ sera, pour une infinité de valeurs de n, proche de I_0 et sera donc disjoint de J.

c) L'application du tore $\mathbb{R}^2/\mathbb{Z}^2$ définie par $A(x, y) = (2x + y, x + y)$ est topologiquement mélangeante (mais pas minimale : elle admet une infinité de points périodiques).

2.4 Décalages (ou shifts)

2.4.1 Shift de Bernoulli

Notons $\Sigma_+ = \{0,1\}^{\mathbb{N}}$ (resp. $\Sigma = \{0,1\}^{\mathbb{Z}}$), l'ensemble des suites $(x_i)_{i\in\mathbb{N}}$ (resp, $(x_i)_{i\in\mathbb{Z}}$) $x_i \in \{0,1\}$. On munit Σ_+ et Σ de la distance (ultra-métrique)

$$d(x,y) = 2^{-m(x,y)}, \qquad m(x,y) = \inf\{|j| : x_j \neq y_j\}.$$

Afin de simplifier la discussion nous limiterons notre propos à Σ_+ (la discussion étant la même pour Σ). L'ensemble des *cylindres centrés* $C(m; a_0, \ldots, a_m) = \{(x_j)_{j\in\mathbb{N}} : x_0 = a_0, \ldots, x_m = a_m\}$ est une base dénombrable d'ouverts de la topologie définie par d. Le théorème de Tykhonov (ou encore un argument diagonal) montre que (Σ_+, d) est compact. On définit alors l'application de décalage ou shift σ par

$$(\sigma(x))_i = x_{i+1}.$$

Il est facile de voir que σ est continue sur (Σ_+, d).

Proposition 2.4.1 *L'application σ est faiblement mélangeante.*

Démonstration. — Il suffit de démontrer que pour tous cylindres $C = C(m; a_0, \ldots, a_m)$, $C' = C(m'; a'_0, \ldots, a'_{m'})$ il existe N tel que pour tout $n \geq N$ $\sigma^{-n}(C) \cap C' \neq \emptyset$. Mais $\sigma^{-n}(C)$ est l'ensemble des mots de la forme yaz où $a = a_0 \ldots a_m$ et y est une suite quelconque de longueur n, et $z \in \Sigma_+$ est également quelconque. Ainsi, si $n \geq m'$, tout mot de la forme $a'waz$ où $a' = a'_0 \ldots a'_{m'}$ et w de longueur $n - m'$ et $z \in \Sigma_+$ sont quelconques, appartient à $\sigma^{-n}(C) \cap C'$. □

On peut généraliser la construction et la proposition précédente au cas d'un shift sur un alphabet à r symboles $\mathcal{A} = \{1, \ldots, r\}$. On pose alors $\Sigma_+ = \mathcal{A}^{\mathbb{N}}$ (resp. $\Sigma = \mathcal{A}^{\mathbb{Z}}$) et on définit σ de la même manière.

2.4.2 Sous-shifts de type fini

Soit $\mathcal{A} = \{1, \ldots, r\}$ un alphabet à r symboles et A une matrice $r \times r$ à coefficients dans $\{0,1\}$ telle que pour tout $1 \leq i \leq r$ il existe j et j' dans $\mathcal{A}$ tels $(A)_{ij} = 1$ et $(A)_{j'i} = 1$. On associe à A un graphe orienté tel que pour tous sommets i, j il existe au plus une flèche de i vers j et pour tout sommet i il existe une flèche arrivant en i et une flèche sortant de i (cette association (matrice,graphe) est alors bijective). Nous noterons Γ_A le graphe *orienté associé à* A et $\tilde{\Gamma}_A$ le graphe non-orienté associé à Γ_A. Les hypothèses que nous avons faites sur A assure que le graphe non orienté $\tilde{\Gamma}_A$ associé à A est connexe. En revanche, la connexité du graphe orienté n'est pas automatique.

Définition 2.4.1 *Le graphe orienté Γ_A est dit fortement connexe si pour toute paire i, j dans l'alphabet $\mathcal{A}$ il existe un chemin allant de i à j (en suivant le sens des flèches).*

Remarque : Le nombre de chemins de longueur r allant de i à j est le coefficient $(A^r)_{ij}$. Le graphe Γ_A est donc fortement connexe si et seulement si

$$\forall\ i, j \quad \exists r(i, j), \quad (A^r)_{ij} > 0.$$

Nous noterons Σ_A l'ensemble des suites $(x_i)_{i\in\mathbb{N}}$ de $\mathcal{A}^{\mathbb{N}}$ qui vérifient la condition de compatibilité suivante : pour tout $i \geq 0$, $A_{x_i x_{1+1}} = 1$. Il est clair que Σ_A est encore un espace compact quand on le munit de la distance induite par l'inclusion et que Σ_A est un fermé invariant par σ.

Théorème 2.4.1 *Le système dynamique (Σ_A, σ) est transitif si et seulement si le graphe Γ_A est fortement connexe.*

Démonstration. — Remarquons déjà que si $i, j \in \mathcal{A}$ l'existence d'un chemin de i à j est équivalente au fait qu'il existe un entier n tel que $U_i \cap \sigma^{-n}(U_j) \neq \emptyset$ où on note $U_i = \{x : x_0 = i\}$ $U_j = \{x : x_0 = j\}$. On a donc clairement que si (Σ_A, σ) est transitif alors Γ_A est fortement connexe.

Réciproquement, supposons le graphe Γ_A fortement connexe. Pour démontrer la transitivité il suffit de prouver que pour tous cylindres $C = C(m; a_0, \ldots, a_m)$ et $C' = C(m'; a'_0, \ldots, a'_{m'})$ il existe n tel que $C \cap \sigma^{-n}(C') \neq \emptyset$. Notons a (resp. a') le mot $a_0, \ldots, a_m$ (resp. $a'_0, \ldots, a'_{m'}$). On sait qu'il existe un chemin allant de la fin du mot a_m de a au début du mot a'_0 de a'. On peut donc construire un mot b commençant par a_m et terminant par a'_0. Alors tout mot $aba'x$ est dans C et est tel que $\sigma^l(aba'x) \in C'$ pour l égal à la somme des longueurs des mots a et b. □

Caractérisons les sous-shifts de type fini topologiquement mélangeant :

Théorème 2.4.2 *Le système dynamique (Σ_A, σ) est topologiquement mélangeant si et seulement si Γ_A vérifie la propiété suivante : il existe un entier $r > 0$ tel que pour toute paire (i, j) il existe un chemin de longueur r allant de i à j (ce qui est équivalent à $(A^r)_{ij} > 0$).*

Démonstration. — Le début de la preuve du théorème précédent montre que si (Σ_A, σ) est topologiquement mélangeant, pour toute paire (i, j) il existe un entier $N_{i,j}$ tel que pour $n \geq N_{i,j}$, $U_i \cap \sigma^{-n}(U_j) \neq \emptyset$. Par conséquent si on pose $r = \max i, j N_{i,j}$ on a bien l'existence d'un chemin de longueur r dans le graphe Γ_A allant de i à j.

Réciproquement, supposons que le graphe vérifie la propriété du théorème. Alors, A^r a tous ses coefficients positifs strictement. Comme aucune ligne de

A n'est nulle (et comme A est à coefficients positifs ou nuls), A^{r+1} et plus généralement A^k, $k \geq r$ a tous ses coefficients strictement positifs. Il existe donc des chemins pour tout $k \geq r$ et toute paire i, j, des chemins de longueur k allant de i à j. Si on reprend la deuxième partie de la preuve du théorème précédent, on voit que les mots de la forme $aba'x$ où b est un mot de longueur $k \geq r$ sont dans $C \cap \sigma^{-k}(C')$. □

Il existe en fait une décomposition des sous-shifts de type fini transitifs en union disjointe de fermés ou la restriction d'une puissance de σ est topologiquement mélangeant.

Théorème 2.4.3 *Le système (Σ_A, σ) est transitif si et seulement si il existe $\Sigma_1, \ldots, \Sigma_m$ fermés de Σ_A disjoints et dont l'union est Σ_A tels que*

a) pour tout $1 \leq i < m$ on a $\sigma(\Sigma_i) = \Sigma_{i+1}$ et $\sigma(\Sigma_m) = \Sigma_1$;

b) $\sigma^m|\Sigma_1$ est topologiquement mélangeant.

Démonstration. — Montrons déjà le sens direct. Notons Λ_{ij} l'ensemble des longueurs de chemins allant de i à j ; on a pour $i, j, k \in \mathcal{A}$ $\Lambda_{ij} + \Lambda_{jk} \subset \Lambda_{ik}$. En particulier $\Lambda_{11} + \Lambda_{11} \subset \Lambda_{11}$ et si on note m le pgcd des éléments de Λ_{11} on peut dire que Λ_{11} contient tous les multiples de m assez grands (**Exercice**). Définissons alors les sous-ensembles de $\mathcal{A}$, $I_1, \ldots, I_m$ de la façon suivante : j appartient à I_l si et seulement s'il existe un chemin allant de 1 à j de longueurs congrue à $l - 1$ modulo m. On a alors,

Lemme 2.4.1 *Si $i \in I_l$ et $j \in I_{l'}$ tous les chemins allant de i à j sont de longueurs congrues à $l' - l$ modulo m*

Démonstration. — a) Démontrons déjà qu'il existe un chemin allant de i à 1 de longueur congrue à $-l$ modulo m. On sait qu'il existe un chemin allant de i à 1 dont on note s la longueur ; on peut donc construire un chemin allant de 1 à 1 en concaténant un chemin de longueur congrue à l modulo m et le chemin de longueur s. Comme tous les chemins allant de 1 à 1 sont de longueurs divisibles par m (par définition de m), s doit être congru à $-l$ modulo m.

b) Démontrons que tous les chemins allant de 1 à i ont une longueur congrue à l modulo m. En effet, étant donné un tel chemin de longueur t, puisqu'il en existe un autre allant de i à 1 de longueur congrue à $-l$ modulo m, on peut en concaténant construire un chemin de 1 à 1 de longueur congrue à $t - l$ modulo m. Mais un tel chemin a une longueur divisble par m si bien que t est congru à l modulo m.

c) Comme tout chemin allant de i à i concaténé à un chemin allant de i à 1 donne un chemin allant de i à 1, on déduit de b) que tout chemin de i à i est de longueur congrue à 0 modulo m.

d) D'après b), il existe un chemin allant de i à j de longueur congrue à $l' - l$ modulo m (en concaténant via 1) et un autre de j à i de longueur congrue à $l - l'$ modulo m. Pour tout chemin allant de i à j de longueur t on peut en concaténant en j construire un chemin de i à i de longueur congrue à $t+l-l'$ modulo m et comme cette longueur doit être un multiple de m d'après c) on a bien la conclusion du lemme. □

Définissons alors Σ_l comme étant l'ensemble des suites $x = (x_i)_{i\in\mathbb{N}}$ de Σ_A telles que $x_0 \in I_l$. Automatiquement, $x_1 \in I_{l+1 \mod m}$ car il existe un chemin de 1 à x_1 de longueur $1 + l$ (l de 1 à x_0 et 1 de x_0 à x_1) et de façon plus générale $x_k \in I_{l+k \mod m}$. On a donc bien $\sigma(\Sigma_l) = \Sigma_{l+1 \mod m}$. Vérifions que σ^m restreint à Σ_1 est topologiquement mélangeant. Pour cela on reprend la démonstration du théorème précédent : il suffit de démontrer que si a est un mot de longueur p commençant par une lettre de I_1 et a' un mot de longueur p' commençant par une lettre de I_1, on peut construire pour tout entier km un mot compatible de la forme $aba'x$ où b est un mot de longueur km. Or ceci est toujours possible dès que k est assez grand puisque Λ_{11} contient tous les multiples de m assez grands.

La réciproque est laissée en exercice au lecteur. □

2.5 Application à la preuve du théorème de van der Waerden

Nous nous proposons de démontrer par des méthodes de dynamique topologique le théorème suivant dû à van der Waerden :

Théorème 2.5.1 (van der Waerden) *Si $\mathbb{Z} = A_1 \cup \cdots \cup A_p$ est une partition de $\mathbb{Z}$ il existe $i \in \{1, \ldots, p\}$ tel que A_i contienne des progressions arithmétiques de longueur arbitraire i.e : pour tout $r \in \mathbb{N}$ il existe $a, b \in \mathbb{Z}$ tels que $a, a+b, \ldots, a+(r-1)b$ appartiennent à A_i.*

La preuve du théorème précédent est basée sur un théorème de récurrence multiple qui est dû à H. Furstenberg (la preuve que nous présentons est due à J.-P. Thouvenot).

Théorème 2.5.2 *Si (X, d) est un espace métrique compact et T un homéomorphisme de X il existe $x \in X$ tel que pour tout $r \geq 1$*

$$\inf_{n\geq 1} \max_{1\leq i\leq r} d(x, T^{in}x) = 0.$$

Démonstration. — La preuve se fait par récurrence sur $r \geq 1$.
i) Si $r = 1$ nous avons vu que le résultat est vrai : si K est un ensemble minimal de X, tout point x de K est d'orbite dense dans K et en particulier récurrent.
ii) Notons K un ensemble minimal fixé et notons pour $r \geq 1$, E_r l'ensemble des $x \in K$ pour lesquels

$$\inf_{n\geq 1} \max_{i\leq r} d(x, T^{in}x) = 0.$$

Lemme 2.5.1 *Si E_r est un G_δ-dense de K alors il en est de même de E_{r+1}.*

Démonstration. — [du lemme 2.5.1]

A) Remarquons tout d'abord que l'ensemble $\tilde{E}_r$ des $x \in K$ pour lesquels $T^n x \in E_r$ pour tout $n \in \mathbb{Z}$, $x \in E_r$ est encore un G_δ-dense (c'est l'ensemble $\bigcap_{n\in\mathbb{Z}} T^{-n}(E_r)$). Définissons alors par récurrence les suites ϵ_n de réels positifs et x_n d'éléments de $\tilde{E}_r$ de la façon suivante : $x_1 \in \tilde{E}_r$ et ϵ_1 étant choisis, il existe n_1 tel que

$$\max_{1\leq i\leq r} d(x, T^{-in_1}x_1) < \epsilon_1.$$

On pose alors $x_2 = T^{-(r+1)n_1}x_1$ et on choisit ϵ_2 suffisamment petit pour que $\epsilon_2 < \epsilon_1$ et pour que $d(y, x_2) < \epsilon_2$ implique $d(T^{n_1(r+1)}y, x_1) < \epsilon_1$. Nous noterons

$$x_2 \xrightarrow[\epsilon_1]{n_1} x_1$$

les relations

$$\max_{1\leq i\leq r+1} d(x_1, T^{in_1}x_2) < \epsilon_1.$$

Avec le choix que nous avons fait pour ϵ_2 il est clair que pour tout y tel que $d(y, x_2) < \epsilon_2$ on a

$$y \xrightarrow[\epsilon_1]{n_1} x_1.$$

On sait qu'il existe n_2 tel que

$$\max_{1\leq i\leq r} d(x, T^{-in_2}x_2) < \epsilon_2,$$

et on pose $x_3 = T^{-(r+1)n_2}x_2$. On a bien

$$x_3 \xrightarrow[\epsilon_2]{n_2} x_2$$

et si $\epsilon_3 < \epsilon_2$ est choisi suffisamment petit on a

$$y \xrightarrow[\epsilon_2]{n_2} x_2$$

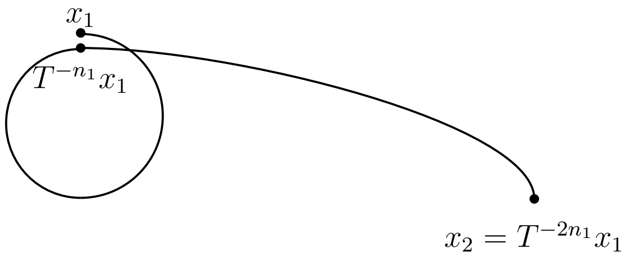

FIGURE 2.4 – Construction de la suite x_k.

pour tout y tel que $d(y, x_3) < \epsilon_3$. On a donc pour un tel y

$$y \xrightarrow[\epsilon_2]{n_2} x_2 \xrightarrow[\epsilon_1]{n_1} x_1.$$

Par récurrence on construit des suites infinies ϵ_k, x_k telles que

$$\cdots x_{k+1} \xrightarrow[\epsilon_k]{n_k} x_k \xrightarrow[\epsilon_{k-1}]{n_{k-1}} \cdots \xrightarrow[\epsilon_1]{n_1} x_1.$$

Comme K est compact, pour tout $\epsilon > 0$ il existe $x_k, x_l \in \tilde{E}_r$ avec $k > l$ et $d(x_k, x_l) < \epsilon$ tels que

$$x_k \xrightarrow[\epsilon_l]{n_{k,l}} x_l$$

où $n_{k,l} = n_{k-1} + \cdots + n_l$. En particulier, comme on peut choisir l tel que $\epsilon_l < \epsilon$, on a prouvé le lemme suivant :

Lemme 2.5.2 *Pour tout $\epsilon > 0$ il existe $x_\epsilon \in \tilde{E}_r$ et n_ϵ tels que*

$$x_\epsilon \xrightarrow[\epsilon]{n_\epsilon} x_\epsilon.$$

B) Pour terminer la preuve du Lemme 2.5.1 il suffit de voir que l'on peut inverser l'ordre des quantificateurs :

Lemme 2.5.3 *Il existe $x \in \tilde{E}_r$ tel que pour tout $\epsilon > 0$ il existe n_ϵ tel que*

$$x \xrightarrow[\epsilon]{n_\epsilon} x.$$

Démonstration. — B.1) Notons A_ϵ l'ensemble des $x \in \tilde{E}_r$ pour lesquels il existe n_ϵ tel que

$$x \xrightarrow[\epsilon]{n_\epsilon} x.$$

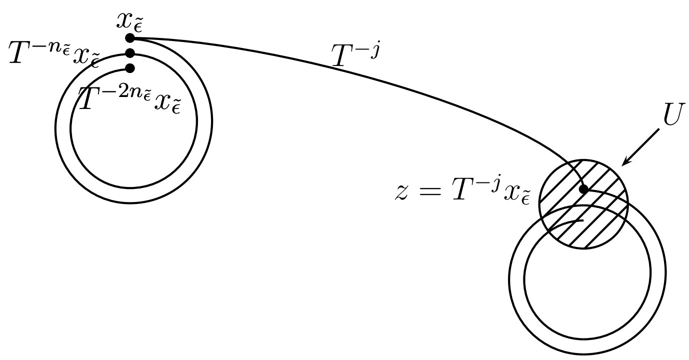

FIGURE 2.5 – A_ϵ est dense dans X.

Cet ensemble est un ouvert de $\tilde{E}_r \subset K$ et est donc un G_δ de K. Démontrons qu'il est dense dans K. Nous devons démontrer que A_ϵ intersecte tout ouvert U. Fixons cet ouvert. Comme T est minimale sur K, nous savons qu'il existe un entier N tel que tout point $x \in K$ se trouve après $j(x)$ itérations par T dans U où $0 \leq j(x) \leq N$ ($\exists 0 \leq j \leq N$, $T^j x \in U$) et comme les applications T^j, $j \in \mathbb{N}$ sont uniformément continues sur K on en déduit qu'il existe $\tilde{\epsilon} > 0$ tel que pour tout $x \in K$ et tout $0 \leq j \leq N$ on a $T^j(B(x, \tilde{\epsilon})) \subset B(T^j(x), \epsilon)$. Mais, nous savons qu'il existe un point $x_{\tilde{\epsilon}} \in K$ et un entier $n_{\tilde{\epsilon}}$ tels que

$$x_{\tilde{\epsilon}} \xrightarrow[\tilde{\epsilon}]{n_{\tilde{\epsilon}}} x_{\tilde{\epsilon}}.$$

Nous avons donc avec $j = j(x_{\tilde{\epsilon}})$

$$T^j x_{\tilde{\epsilon}}, T^j(T^{n_{\tilde{\epsilon}}} x_{\tilde{\epsilon}}), \ldots, T^j(T^{(r+1)n_{\tilde{\epsilon}}} x_{\tilde{\epsilon}}) \in B(T^j x, \epsilon).$$

Comme $T^j(T^{in_{\tilde{\epsilon}}}(x_{\tilde{\epsilon}})) = T^{in_{\tilde{\epsilon}}}(T^j(x_{\tilde{\epsilon}}))$, cela démontre que le point $z = T^j x_{\tilde{\epsilon}}$ est dans $U \cap A_\epsilon$. Nous avons démontré que A_ϵ est dense dans K et que c'est donc un G_δ-dense.

B.2) D'après le théorème de Baire

$$\bigcap_{m=1}^{\infty} A_{1/m}$$

est donc également un G_δ-dense ; mais cet ensemble n'est rien d'autre que E_{r+1}. La preuve du lemme est terminée. □

Terminons la preuve du théorème 2.5.2. L'intersection

$$E = \bigcap_{r \geq 1} E_r$$

est un G_δ-dense d'après le théorème de Baire. Si x est dans E la conclusion du théorème 2.5.2 est vérifiée. □

□

Preuve du théorème 2.5.1 Soit $\omega \in \{1, \ldots, p\}^{\mathbb{Z}}$ la suite définie par $\omega_i = k$ si $i \in A_k$ et notons X l'adhérence de l'orbite de ω sous l'action du décalage $\sigma : \{1, \ldots, p\}^{\mathbb{Z}} \to \{1, \ldots, p\}^{\mathbb{Z}}$. L'application σ est un homéomorphisme de X. Appliquons le théorème de récurrence multiple 2.5.2 à (X, σ) : il existe $x \in X$ tel que pour tout $r \geq 1$ et $\epsilon = 1/4$ on peut trouver un entier b pour lequel

$$\max_{1 \leq i \leq r} d(x, T^{ib}x) < (1/4).$$

Notons $i = x_0$. On déduit de l'inégalité précédente que $i = x_0 = (T^b x)_0 = \cdots = (T^{rb}x)_0$. Comme x est dans l'adhérence de $\{\sigma^k \omega\}_{k \in \mathbb{Z}}$, il existe a tel que $d(T^a \omega, x) < (1/4)$ et également

$$\max_{1 \leq i \leq r} d(T^a \omega, T^{ib}(T^a \omega)) < (1/4).$$

Ainsi, $\omega_a = x_0 = i$ et $\omega_a = \omega_{a+b} = \cdots = \omega_{a+rb}$. On a donc démontré que $a, a+b, \ldots, a+rb$ appartiennent à A_i. Ceci termine la preuve du théorème 2.5.1.

Chapitre 3

Mesures Invariantes

3.1 Mesures invariantes

3.1.1 Définition

Soit $(X, \mathcal{B}, \mu)$ un espace probabilisé (*cf.* l'appendice) et $T : X \to X$ une application messurable.

Quand X est un espace métrique compact, on choisira souvent en guise de tribu $\mathcal{B}$ la tribu borélienne de X (la plus petite tribu engendrée par les ouverts de X) et on sera souvent dans la situation où $T : X \to X$ est une application continue.

Définition 3.1.1 *Une mesure de probabilité μ sur $(X, \mathcal{B})$ est dite T-invariante si $T_*\mu = \mu$ c'est-à-dire si pour tout $A \in \mathcal{B}$*

$$\mu(T^{-1}(A)) = \mu(A).$$

Il est souvent utile de considérer des situations où l'application T est définie seulement en dehors d'un ensemble μ-négligeable N. Nous dirons par extension que $(X, \mathcal{B}, T, \mu)$ est un système dynamique si $T : X \setminus N \to X \setminus N'$ où N et N' sont des ensembles μ-négligeables de X. Dans ce cas on peut trouver un ensemble μ-négligeable N'' tel que $T : X - N'' \to X - N''$ (**Pourquoi**). Nous dirons souvent que (T, μ) est invariant.

Un critère commode pour vérifier qu'une mesure μ est T-invariante est le suivant.

Proposition 3.1.1 *La mesure de probabilité μ est T-invariante si et seulement si pour toute fonction $\phi \in L^\infty(X, \mu)$ (resp. $L^1(X, \mu)$) on a $\int_X \phi \circ T d\mu = \int_X \phi d\mu$.*

Démonstration. — Si pour tout $\phi \in L^\infty(X,\mu)$ on a $\int_X \phi \circ T d\mu = \int_X \phi d\mu$, alors en prenant $\phi = \mathbf{1}_A$ ($A \in \mathcal{B}$) on a $\mu(T^{-1}(A)) = \mu(A)$ puisque $\mathbf{1}_A \circ \phi = \mathbf{1}_{T^{-1}(A)}$.

Démontrons la réciproque. Si pour tout $A \in \mathcal{B}$ on a $\mu(T^{-1}(A)) = \mu(A)$ alors pour toute fonction simple ϕ on a $\int_X \phi \circ T d\mu = \int_X \phi d\mu$. Pa r le théorème de convergence monotone, cela est également vréie pour tout fonction positive et donc pour toute fonction dans$L^1(X,\mu)$. □

Exemple

La mesure de Haar sur le tore $\mathbb{R}/\mathbb{Z}$ est invariante par toute translation $T_\alpha : x \mapsto x + \alpha$. Nous verrons plus loin que dans le cas où α est irrationnel, c'est l'unique mesure invariante par T_α. En revanche, si $\alpha = p/q$, toute orbite périodique porte une mesure invariante.

La mesure de Haar sur le tore $\mathbb{R}/\mathbb{Z}$ est invariante par l'application $S : x \mapsto 2x$. Il existe une infinité d'autres mesures invariantes, en particulier les mesures portées par les orbites périodiques. Si x est un point périodique de période p, la mesure $(\delta_x + \cdots + \delta_{S^{p-1}(x)})/p$ est une mesure de probabilité S-invariante qui n'est pas équivalente à la mesure de Lebesgue.

Exercice : Construire une dynamique sur $[0,1]$ (non continue) qui n'admet pas de mesure de probabilité invariante.

3.1.2 Existence de mesures invariantes pour les dynamiques continues

L'ensemble $\mathcal{M}_T$ sera muni de la topologie faible* : une suite de mesures de probabilité μ_n converge vers μ si pour toute fonction continue $\phi \in C(X)$, $\int \phi d\mu_n = \langle \mu_n, \phi \rangle = \mu_n(\phi)$ converge vers $\int \phi d\mu = \langle \mu, \phi \rangle = \mu(\phi)$.

La proposition qui suit montre que lorsque T est continue il existe toujours des mesures T-invariantes sur X compact.

Proposition 3.1.2 *L'ensemble $\mathcal{M}_T$ est non vide, convexe et compact pour la topologie faible*.*

Démonstration. — La convexité de $\mathcal{M}_T$ est immédiate. Sa compacité pour la topologie faible* résulte du fait que $\mathcal{M}_T$ est fermé pour cette topologie et du fait que l'ensemble des mesures de probabilités sur X compact est compact pour cette topologie. Il reste donc à démontrer qu'il existe une mesure de probabilité T-invariante : soit $x \in X$ et considérons le barycentre des mesures de Dirac :

$$\mu_n = \frac{1}{n}\sum_{k=0}^{n-1} \delta_{T^k(x)}.$$

On a pour $\phi \in C(X)$
$$\mu_n(\phi) = \frac{1}{n}\sum_{k=0}^{n-1} \phi(T^k(x)).$$
On peut extraire de la suite μ_n une sous-suite μ_{n_k} qui converge pour la topologie faible* vers une mesure de probabilité μ. Comme,
$$\mu_n(\phi \circ T) = \frac{\phi(T^n(x)) - \phi(x)}{n} + \frac{1}{n}\sum_{k=0}^{n-1} \phi(T^k(x))$$
on a bien la conclusion en prenant $n = n_k$ et en faisant $k \to \infty$. □

3.1.3 Théorème de Récurrence de Poincaré

Savoir qu'une mesure est T-invariante donne des renseignements précieux sur les propriétés de récurrence de T :

Théorème 3.1.1 (de récurrence de Poincaré) *Si μ est une mesure T-invariante et $A \in \mathcal{B}$ un sous-ensemble de X de μ-mesure positive $\mu(A) > 0$ alors pour μ-presque tout point x de A il existe une suite infinie d'entiers n_k telle que $T^{n_k}(x) \in A$.*

Démonstration. — Notons B l'ensemble des $x \in A$ qui ne reviennent jamais dans A c'est-à-dire
$$B = A \cap \bigcap_{n \geq 1} T^{-n}(X - A).$$
Nous allons démontrer que $\mu(B) = 0$. Prouvons pour cela que les ensembles $B, T^{-1}(B), \ldots, T^{-k}(B), \ldots$ sont 2 à 2 disjoints. En effet pour $j > i$
$$T^{-i}(B) \cap T^{-j}(B) = T^{-i}(B \cap T^{-(j-i)}(B));$$
or si un point x appartient à $B \cap T^{-(j-i)}(B)$, son itéré $T^{j-i}(x)$ appartient à B donc à A ce qui contredit la définition de B puisque x est dans A et qu'un de ses itérés est dans A.

Puisque les $T^{-n}(B)$, $n \geq 0$ sont disjoints deux à deux on a
$$\sum_{n \geq 0} \mu(T^{-n}(B)) = \mu(\bigcup_{n \geq 0} T^{-n}(B));$$
L'inégalité $\mu(B) > 0$ est impossible puisque d'après l'invariance de μ par T le membre de gauche est infini, tandis que le membre de droite est inférieur ou égal à 1 (μ est une mesure de probabilité). Nous avons démontré que

$\mu(B) = 0$. D'après l'invariance de μ par T nous avons aussi $\mu(T^{-k}(B)) = 0$ et donc $\mu(\bigcup_{k\geq 0} T^{-k}(B)) = 0$. Mais cet ensemble contient les $x \in A$ pour lesquels il existe k_0 tel que $T^k(x) \notin A$ pour $k \geq k_0$. Le théorème est démontré. □

Corollaire 3.1.1 *Si (X, d) est métrique compact, l'ensemble des $x \in X$ qui sont T-récurrents est de μ-mesure totale (T n'est pas nécessairement continue).*

Démonstration. — Notons U_n, $n \geq 1$ une base dénombrable de voisinage de X. Notons C_n l'ensemble des $x \in U_n$ qui ne sont pas récurrents dans U_n. L'enemble des points non récurrents de X est la réunion des C_n et comme d'après le théorème précédent chaque C_n est de mesure nulle, le corollaire est démontré. □

3.1.4 Exemples

Translations sur les tores : Si $\mathbb{T}^n = \mathbb{R}^n/\mathbb{Z}^n$ est le tore de dimension n, la mesure de Haar $\mu = dx_1 \wedge \cdots \wedge dx_n$ (on identifie la forme volume avec une mesure) est l'unique mesure invariante par les translations T_α, $T_\alpha(x) = x+\alpha$, $\alpha \in \mathbb{T}^n$.

$T : x \mapsto 2x$ *sur* $[0, 1]$: La mesure de Lebesgue λ est invariante par T puisque pour tout intervalle dyadique $I = ([k/2^p, (k+1)/2^p[$, $T^{-1}([k/2^p, (k+1)/2^p[)$ est l'union disjointe de $[k/2^{p+1}, (k+1)/2^{p+1}[$ et de $(1/2) + [k/2^{p+1}, (k+1)/2^{p+1}[$. On a donc bien $\lambda(T^{-1}(I) = \lambda(I)$ et comme les intervalles dyadiques engendrent la tribu borélienne, la propriété d'invariance s'etend à tous les boréliens.

Décalage et sous-décalage de type fini : Si $p \in [0, 1]$ on peut définir la mesure μ sur $\{0, 1\}^{\mathbb{Z}}$ (muni de la tribu engendrée par les cylindres) de la façon suivante : pour tout cylindre $C = C(\epsilon_0, \ldots, \epsilon_{n-1})$ on pose $\mu(C) = p^r(1-p)^{n-r}$ où r est le nombre de ϵ_i égaux à 1. Le théorème de Carathéodory (ou Kolmogorov) permet d'étendre cette mesure à la tribu borélienne toute entière. La mesure obtenue est clairement invariante par le shift σ (c'est clair sur les cylindres et on utilise la partie unicité du théorème de Carathéodory). Ceci fournit une famille entière de mesures invariantes par le décalage, les mesures de Bernoulli. Ce ne sont pas les seules : chaque orbite périodique de σ définit naturellement une mesure σ-invariante : la moyenne des mesures de Dirac portées par l'orbite périodique.

Pour les sous-shifts de type fini sur un alphabet à r symboles et de matrice de transition A, la construction de mesures naturelles invariantes par σ se fait de la manière suivante : soit P_{ij} $(1 \leq i, j \leq r)$ les coefficients d'une matrice stochastique P :

a) $P_{ij} \geq 0$

b) $\sum_{j=1}^{r} P_{ij} = 1$

et faisons l'hypothèse que P est compatible avec A c'est-à-dire que $A_{ij} = 0$ si et seulement si $P_{ij} = 0$ et supposons que A soit irréductible c'est-à-dire que (Σ_A, σ) soit transitif.

La condition b montre que le vecteur dont toutes les composantes valent 1 est vecteur propre de P associé à la valeur propre 1. Par conséquent 1 est valeur propre de ${}^tP$ et il est possible de démontrer que l'espace propre correspondant est engendré par un vecteur dont tous les coefficeients $p_1, \ldots, p_r$ sont positifs ou nuls. On peut supposer que la somme $p_1 + \cdots + p_r = 1$. On a donc pour tout $1 \leq j \leq r$

$$\sum_{i=1}^{r} p_i P_{ij} = p_j. \tag{3.1}$$

Pour tout cylindre $C = C(\epsilon_0, \ldots, \epsilon_n)$ de Σ_A définissons

$$\mu(C) = p_{\epsilon_0} P_{\epsilon_0\epsilon_1} \cdots P_{\epsilon_{n-1}\epsilon_n}.$$

La condition b) garantit la cohérence au sens du théorème de Kolmogorov :

$$\mu(C(\epsilon_0, \ldots, \epsilon_{n-1})) = \sum_{l=1}^{r} \mu(C(\epsilon_0, \ldots, \epsilon_{n-1}, l)).$$

Par conséquent, on peut étendre μ en une mesure (de probabilité) à la tribu entière. L'invariance de μ par σ se fait sur les cylindres en utilisant (3.1) et

$$\mu(\sigma^{-1}(C(\epsilon_0, \ldots, \epsilon_n) = \sum_{l=1}^{r} \mu(C(l, \epsilon_0, \ldots, \epsilon_n)).$$

Automorphismes des tores Si $A \in SL(n, \mathbb{Z})$, l'application $\bar{A}$ de $\mathbb{R}^n/\mathbb{Z}^n$ dans lui même définie par $\bar{A}(x + \mathbb{Z}^n) = Ax + \mathbb{Z}^n$ est inversible et s'appelle un automorphisme du tore $\mathbb{T}^n$. Puisque $\det(A) = 1$ on voit que la mesure de Haar est invariante par $\bar{A}$ (**Exercice**). Bien évidemment ce n'est pas la seule puisque A admet une infinité de points périodiques.

3.2 Ergodicité

Définition 3.2.1 *Un système dynamique $(X, \mathcal{B}, \mu, T)$ est dit ergodique ssi tout ensemble $A \in \mathcal{B}$ tel que $\mu(A \Delta T^{-1}(A)) = 0$ est de μ mesure 0 ou 1. (On note $A \Delta B = (A \cup B) \setminus (A \cap B)$.)*

En d'autres termes, d'un point de vue mesurable, les seuls ensembles invariants sont $\emptyset$ ou X. On peut reformuler la définition précédente :

Proposition 3.2.1 *Le système dynamique $(X, \mathcal{B}, \mu, T)$ est ergodique si et seulement si les seules fonctions $\phi \in L^\infty(X, \mu)$ vérifiant $\phi \circ T = \phi$ sont les fonctions constantes*

Démonstration. — Prouvons que si T est ergodique les seules fonctions invariantes par T sont les constantes : introduisons $E_\lambda = \{x \in X : \phi(x) \le \lambda\}$. Comme $\phi \circ T = \phi$ on a $\phi^1(E_\lambda) = E_\lambda$ (μ p.p) et d'après l'ergodicité $F_\phi(\lambda) := \mu(\phi \le \lambda) \in \{0, 1\}$. Comme la fonction de répartition F_ϕ est croissante et continue à droite il existe λ_0 tel que pour tout $\lambda \ge \lambda_0$, $\mu(\phi \le \lambda) = \mu(\phi = \lambda_0) = 1$. Ainsi, ϕ est μ-p.p constante (égale à λ_0).

La réciproque est claire (observer que $\phi = \mathbf{1}_A$ est T-invariante ssi A est un ensemble T-invariant).

□

3.2.1 Premiers exemples

Translation sur des tores

Si $X = \mathbb{T}$ et μ est la mesure de Haar, la transformation $T : x \mapsto x + \alpha$ est μ-ergodique si et seulement si α est irrationnel. En effet, soit ϕ est une fonction L^∞ telle que $\phi \circ T = \phi$. Puisque ϕ est L^2, on a l'identité dans L^2

$$\phi(x) = \sum_{k \in \mathbb{Z}} \hat{\phi}(k) e^{2\pi i k x}$$

où les $\hat{\phi}(k)$ sont les coefficients de Fourier de ϕ. De l'unicité de la décomposition de Fourier et de l'identité dans L^2, $\phi \circ T = \phi$ on tire

$$\hat{\phi}(k) e^{2\pi i k \alpha} = \hat{\phi}(k).$$

Si α est irrationnel on a alors pour tout $k \neq 0$ $\hat{\phi}(k) = 0$ et par conséquent, ϕ est constante. Si $\alpha = p/q$ il existe clairement des fonctions ϕ qui sont T-invariantes et qui ne sont pas constantes, par exemple $\phi(x) = e^{2\pi i q x}$.

On peut donner une deuxième preuve de ce résultat qui est plus géométrique et ne fait pas appel à la décomposition en série de Fourier. Si $\alpha = p/q$ ($p \wedge q = 1$) est rationnel, l'orbite de 0 est un ensemble discret. Il est clair qu'il existe un petit intervalle ouvert I contenant 0 et dont tous les itérés (qui sont au nombre de q) sont disjoints deux à deux. L'union de ces intervalles est un borélien de mesure de Haar différente de 0 et de 1 et invariant par T ce qui prouve que T n'est pas ergodique pour la mesure de Haar. Supposons à présent α irrationnel et faisons l'hypothèse qu'il existe un ensemble borélien A T-invariant, de mesure de Lebesgue comprise entre 0 et 1 strictement. D'après le théorème de densité de Lebesgue, presque tout point de A est un point de densité de A, ce qui signifie que pour Lebesgue presque tout $x \in A$ on a

$$\lim_{\epsilon \to 0} \frac{\mathrm{Leb}(]x-\epsilon, x+\epsilon[\cap A)}{2\epsilon} = 1.$$

Soit x un point de densité et $I =]x-\epsilon, x+\epsilon[$ tel que $\frac{\mathrm{Leb}(]x-\epsilon,x+\epsilon[\cap A)}{2\epsilon} \geq 1-\delta$. Comme T est une isométrie minimale, il est clair que l'on peut trouver une suite d'entiers n_k, $0 \leq k \leq r$ telle que les T^{n_k}, $1 \leq k \leq r$ soient disjoints deux à deux et couvrent un ensemble de mesure plus grande que $1-\delta$. Dans chaque $T^{n_k} I$ la proportion de points de A est supérieure à $1-\delta$ puisque A est T-invariant , si bien que $A \cap \cup_{0 \leq k \leq r} T^{n_k} I$ a une mesure supérieure ou égale à $(1-\delta)^2$. Par conséquent la mesure de A est supérieure ou égale à $(1-\delta)^2$. Comme ceci vaut pour tout δ, on en déduit que la mesure de A égale 1, ce qui est une contradiction.

Exercice Montrer qu'une translation T_α sur le tore de dimension n est ergodique pour la mesure de Haar, si et seulement si $\langle k, \alpha \rangle \in \mathbb{Z}$, $k \in \mathbb{Z}$ entraîne $k = 0$.

Ergodicité de $x \mapsto 2x$

Démontrons que $T : x \mapsto 2x$ est ergodique pour la mesure de Haar sur $\mathbb{T}$. Soit ϕ une fonction T-invariante. Les coefficients de Fourier de ϕ vérifient $\hat{\phi}(2k) = \hat{\phi}(k)$ si bien que $\hat{\phi}(2^p k) = \hat{\phi}(k)$ pour tous entiers k, p. Si $k \neq 0$ $\lim_{p\to\infty} \hat{\phi}(2^p k) = 0$ car les coefficients de Fourier d'une fonction L^2 sont dans $l^2(\mathbb{Z})$ et donc tendent vers 0 à l'infini. Ainsi, ϕ est constante, ce qui prouve l'ergodicité.

Exercice Démontrer que si $X = \mathbb{T}^2$, l'automorphisme du tore $T(x, y) = (2x+y, x+y)$ préserve la mesure de Haar et est ergodique pour cette mesure.

Exercice Démontrer que le décalage est ergodique pour les mesures de Bernoulli.

3.3 Les Théorèmes ergodiques

3.3.1 Le point de vue spectral et le théorème de von Neumann

Considérons un système dynamique mesurable $(X, \mathcal{B}, T, \mu)$. L'espace $L^2(X, \mathcal{B}, \mu)$ muni du produit scalaire

$$\langle \phi, \psi \rangle = \int_X \phi \bar{\psi} d\mu, \qquad \phi, \psi \in L^2(X, \mathcal{B}, \mu)$$

est un espace de Hilbert. Le point de vue spectral consiste à étudier l'opérateur linéaire U_T (que nous noterons souvent T) agissant sur $L^2(X, \mathcal{B}, \mu)$:

$$\begin{aligned} U_T : L^2(X, \mathcal{B}, \mu) &\to L^2(X, \mathcal{B}, \mu) \\ \phi &\mapsto \phi \circ T \end{aligned}$$

Notons U^* l'adjoint de U défini par

$$\langle U^* \phi, \psi \rangle = \langle \phi, U \psi \rangle;$$

puisque T préserve μ il est facile de voir que U_T est une isométrie c'est-à-dire préserve la norme (ou le produit scalaire) $\|U_T \phi\| = \|\phi\|$ et par conséquent

$$U^* U = Id.$$

Si en outre T est inversible T est unitaire c'est-à-dire que $U^* = U^{-1}$ puisque, T préservant μ

$$\langle \phi \circ T, \psi \rangle = \int_X (\phi \cdot \bar{\psi} \circ T^{-1}) \circ T d\mu = \langle \phi, \psi \circ T^{-1} \rangle.$$

Essayons de comprendre la situation quand T n'est pas inversible. La sigma algèbre $T^{-1}B$ est incluse dans B et il est possible de définir l'espace $L^2(X, T^{-1}\mathcal{B}, \mu)$ des fonctions $T^{-1}\mathcal{B}$-mesurables qui sont L^2 pour la (restriction à $T^{-1}\mathcal{B}$ de la) mesure μ. C'est un sous-espace fermé de $L^2(X, \mathcal{B}, \mu)$ et on peut introduire la projection orthogonale $P : L^2(X, \mathcal{B}, \mu) \to L^2(X, T^{-1}\mathcal{B}, \mu)$ (rappelons qu'une projection orthogonale est caractérisée par $P^* = P$ et $P^2 = P$). Cette projection s'appelle *l'espérance conditionnelle* par rapport à la tribu $T^{-1}\mathcal{B}$ et se note $\mathbf{E}(\cdot | T^{-1}B)$. On a le lemme facile suivant :

Lemme 3.3.1 *L'espace $L^2(X, T^{-1}\mathcal{B}, \mu)$ est l'ensemble des fonctions $\phi \in L^2(X, \mathcal{B}, \mu)$ qui s'écrivent sous la forme $\phi = \psi \circ T$ où $\psi \in L^2(X, \mathcal{B}, \mu)$.*

Démonstration. — Il est clair qu'une fonction de la forme $\phi = \psi \circ T$ où $\psi \in L^2(X, \mathcal{B}, \mu)$ est mesurable par rapport à la tribu $T^{-1}\mathcal{B}$ et appartient donc à $L^2(X, T^{-1}\mathcal{B}, \mu)$. Réciproquement, si une fonction ϕ est dans $L^2(X, T^{-1}\mathcal{B}, \mu)$, il est possible de trouver une suite de fonctions étagées $\phi_n = \sum_i \lambda_{i,n} \mathbf{1}_{B_{i,n}}$ avec $B_{i,n} \in T^{-1}\mathcal{B}$ convergeant μ-pp et L^2 vers ϕ. Les $B_{i,n}$ sont par définition de la forme $T^{-1}A_{i,n}$ et $\mathbf{1}_{T^{-1}A_{i,n}} = \mathbf{1}_{1_{i,n}} \circ T$. Remarquons que

$$\| \sum_i \lambda_i \mathbf{1}_{T^{-1}C_i} \|^2 = \sum_i |\lambda_i|^2 \mu(T^{-1}C_i) = \sum_i |\lambda_i|^2 \mu(C_i) = \| \sum_i \lambda_i \mathbf{1}_{C_i} \|^2.$$

Ainsi, la suite $\phi_n = \sum_i \lambda_{i,n} \mathbf{1}_{B_{i,n}}$ converge dans L^2 si et seulement si la suite $\psi_n = \sum_i \lambda_{i,n} \mathbf{1}_{A_{i,n}}$ converge dans L^2. Notons ψ sa limite. Puisque $\phi_n = \psi_n \circ T = U_T \psi_n$ et que U_T est continue dans L^2 on a $\phi = \psi \circ T$.

□

Le lemme précédent montre que $\mathrm{Im}P = \mathrm{Im}U$. Comme P est un projecteur $\ker P = (\mathrm{Im}P)^\perp$ et comme de façon générale $(\overline{\mathrm{Im}U})^\perp = \ker U^*$ on a $\ker P = \ker U^*$. L'opérateur UU^* est symétrique borné et vérifie $(UU^*)^2 = UU^*UU^* = UU^*$ (car U est une isométrie) si bien que UU^* est une projection orthogonale ; son noyau $\ker UU^*$ est $\ker U^*$ (**exercice**) i.e $\ker P$. Cela suffit pour affirmer que $UU^* = P$. Nous avons donc démontré

Lemme 3.3.2 *On a $U_T^* U_T = Id$ et $U_T U_T^* = \mathbf{E}(\cdot | T^{-1}\mathcal{B})$.*

La tribu des invariants Notons $\mathcal{I}$ la tribu constituée des $A \in \mathcal{B}$ tels que $T^{-1}A = A \mod 0$. L'espace $L^2(X, \mathcal{I}, \mu)$ est l'ensemble des $\phi \in L^2$ telles que $\phi \circ T = \phi$ (**exercice**). On définit comme précédemment $\mathbf{E}(\cdot|\mathcal{I})$ la projection orthogonale sur $L^2(X, \mathcal{I}, \mu)$. On peut énoncer le théorème de Von Neumann

Théorème 3.3.1 *Si $(X, \mathcal{B}, \mu, T)$ est un système dynamique, alors pour toute fonction $\phi \in L^2(X, \mathcal{B}, \mu)$ la suite*

$$\frac{1}{n} S_n \phi = \frac{1}{n} \Big(\phi + \phi \circ T + \cdots + \phi \circ T^{n-1} \Big)$$

converge dans $L^2(X, \mathcal{B}, \mu)$ vers $\mathbf{E}(\phi|\mathcal{I})$.

Démonstration. — Nous allons démontrer que pour toute fonction $\phi \in L^2(X, \mathcal{B}, \mu)$ et tout $\epsilon > 0$ il existe $\psi \in L^2(X, \mathcal{B}, \mu)$ et $\eta_\epsilon \in L^2(X, \mathcal{B}, \mu)$ tels que$\|\eta_\epsilon\| < \epsilon/2$ et

$$\phi = \psi_\epsilon \circ T - \psi_\epsilon + \mathbf{E}(\phi|\mathcal{I}) + \eta_\epsilon.$$

Il suffit pour cela de démontrer que $\phi - \mathbf{E}(\phi|\mathcal{I})$ est dans l'adhérence L^2 de $U - I$. Un calcul simple montre $\overline{\mathrm{Im}(I-U)}^\perp = \ker(I - U^*)$. Or, si $U^*\phi = \phi$

on a $P\phi = UU^*\phi = U\phi$. Mais, $\|U\phi\| = \|\phi\|$ i.e $\|P\phi\| = \|\phi\|$. Ceci n'est possible que si $\phi \in \mathrm{Im}P$ c'est-à-dire si $P\phi = \phi$ d'où l'on déduit $U\phi = \phi$. Réciproquement, si $U\phi = \phi$ on a $\phi = U^*U\phi = U^*\phi$. On a donc prouvé

$$\overline{\mathrm{Im}(U-I)}^{\perp} = \ker(U^* - I) = \ker(U - I) = L^2(X, \mathcal{I}, \mu).$$

On a donc,

$$\overline{\mathrm{Im}(U-I)} = L^2(X, \mathcal{I}, \mu)^{\perp}$$

Mais par définition pour tout $\phi \in L^2(X, \mathcal{B}, \mu)$, $\phi - \mathbf{E}(\phi|\mathcal{I}) \in L^2(X, \mathcal{I}, \mu)^{\perp}$

Concluons la preuve du théorème :

$$\frac{1}{n} S_n \phi = \frac{1}{n}(\phi \circ T^n - \phi) + \mathbf{E}(\phi|\mathcal{I}) + \frac{1}{n} S_n \eta_\epsilon;$$

mais $\|\frac{1}{n} S_n \eta_\epsilon\| \leq \epsilon/2$ si bien que

$$\begin{aligned}\|\frac{1}{n} S_n \phi - \mathbf{E}(\phi|\mathcal{I})\| &\leq \frac{2\|\psi_\epsilon\|}{n} + \|\eta_\epsilon\| \\ &\leq \epsilon\end{aligned}$$

si n est assez grand. □

3.3.2 Convergence presque sûre

Théorème 3.3.2 *a) Si $(X, \mathcal{B}, \mu, T)$ est un système dynamique, alors pour toute fonction $\phi \in L^1(X, \mathcal{B}, \mu)$ la suite*

$$\frac{1}{n} S_n \phi = \frac{1}{n}\Big(\phi + \phi \circ T + \cdots + \phi \circ T^{n-1}\Big)$$

converge vers $\mathbf{E}(\phi|\mathcal{I})$

i) μ-presque sûrement

ii) dans $L^1(X, \mathcal{B}, \mu)$.

b) Si T est inversible on a la même conclusion pour $n \to -\infty$.

Démonstration. —

Un ingrédient crucial de la preuve de ce théorème est le lemme suivant

Lemme 3.3.3 (ergodique maximal de Hopf) *Si ϕ est mesurable, notons $S_n^*\phi(x) = \max_{1\leq k\leq n}(S_k\phi(x))$, $E_n = \{x \in X : S_n^*\phi(x) \geq 0\}$ et $E = \cup E_n$. Alors, $\int_E \phi d\mu \geq 0$.*

Démonstration. — On observe que

$$-\phi(x) + S_{n+1}^*\phi(x) = \max(S_n^*\phi(Tx), 0)$$

si bien que

$$\phi(x) = S_{n+1}^*\phi(x) - (S_n^*\phi)^+ \circ T(x)$$

et

$$\begin{aligned}
\int_{S_{n+1}^*\phi \geq 0} \phi d\mu &= \int_{S_{n+1}^*\phi \geq 0} (S_{n+1}^*\phi) d\mu - \int_{S_{n+1}^*\phi \geq 0} ((S_n^*\phi) \circ T)^+ d\mu \\
&\geq \int_X (S_{n+1}^*\phi)^+ d\mu - \int_X (S_n^*\phi)^+ \circ T d\mu \\
&\geq \int_X (S_{n+1}^*\phi)^+ d\mu - \int_X (S_n^*\phi)^+ d\mu \\
&\geq 0
\end{aligned}$$

la dernière inégalité provenant du fait que $\max(0, \phi(x), \ldots, S_{n-1}\phi(x), S_n\phi(x)) \geq \max(0, \phi(x), \ldots, S_{n-1}\phi(x))$. On a donc $\int_{E_{n+1}} \phi d\mu \geq 0$ pour tout n et on conclut par convergence dominée. $\square$

Un corollaire du lemme précédent est le suivant : *Si* $A \in \mathcal{B}$ *vérifie* $T^{-1}A = A$ *alors* $\int_{E \cap A} \phi d\mu \geq 0$. Démontrons à présent le théorème de Birkhoff : soit $A_{\alpha,\beta}$ l'ensemble des $x \in X$ tels que

$$\liminf_{n\to\infty} \frac{1}{n} S_n\phi(x) \leq \alpha < \beta \leq \limsup_{n\to\infty} \frac{1}{n} S_n\phi(x)$$

L'ensemble $A_{\alpha,\beta}$ est T invariant. Le lemme de Hopf appliqué à $\alpha - \phi$ et $\phi - \beta$ montre que (**pourquoi ?**)

$$\int_{A_{\alpha,\beta}} \phi \leq \alpha\mu(A_{\alpha,\beta}), \qquad \int_{A_{\alpha,\beta}} \phi \geq \beta\mu(A_{\alpha,\beta})$$

ce qui n'est possible que si $\mu(A_{\alpha,\beta}) = 0$. Par conséquent, pour μ-presque tout $x \in X$ $\liminf_{n\to\infty} \frac{1}{n} S_n\phi(x) = \limsup_{n\to\infty} \frac{1}{n} S_n\phi(x)$ et donc $\frac{1}{n} S_n\phi(x)$ converge.

Nous avons donc montré l'existence d'une fonction $\tilde{\phi}$ telle que pour μ-pp

$$\lim_{n\to\infty} \frac{1}{n} S_n\phi(x) = \tilde{\phi}(x).$$

On a nécessairement $\tilde{\phi} \circ T = \tilde{\phi}$ et pour toute fonction $\psi \in L^\infty(\mathcal{I})$ (i.e $\psi \circ T = \psi$) on a $\int_X \psi\tilde{\phi} d\mu = \int_X \psi\phi d\mu$, c'est-à-dire que $\mathbf{E}(\phi|\mathcal{I}) = \tilde{\phi}$ (cf. rappels sur l'espérance conditionnelle dans l'appendice).

Prouvons à présent ii) : d'après le théorème de convergence dominée c'est évident si ϕ est dans $L^\infty(X,\mu)$. Sinon, pour tout $\epsilon > 0$ il existe $\phi_\epsilon \in L^\infty$ qui est ϵ-L^1-proche de ϕ. Comme $\|\mathbf{E}(\phi - \phi_\epsilon|\mathcal{I})\|_{L^1} \leq \epsilon$ et que l'on a convergence L^1 de $n^{-1}S_n\phi_\epsilon$ vers $\mathbf{E}(\phi_\epsilon|\mathcal{I})$ on conclut aisément. □

Exercice Démontrer le point b) du théorème. (On pourra utiliser la convergence L^1 donnée par le théorème de Birkhoff.)

3.4 Liens avec la dynamique topologique

Dans cette section on suppose que (X, d) est un espace métrique compact. La tribu avec laquelle on travaille est la tribu borélienne. Nous noterons $\mathcal{M}_T$ l'ensemble des mesures de probabilité T-invariantes. De façon équivalente, μ est T-invariante si et seulement si pour toute fonction continue $\phi \in C(X)$, $\mu(\phi \circ T) = \mu(\phi)$

3.4.1 Existence de mesures ergodiques

Définition 3.4.1 *Si K est un ensemble convexe, on dit qu'un point $x \in X$ est extrémal si $x = tx_1 + (1-t)x_2,\ 0 < t < 1,\ x_1 \in K,\ x_2 \in K$ implique $x = x_1 = x_2$.*

Notons $\mathcal{M}$ l'ensemble des mesures invariantes T-invariantes. C'est un ensemble convexe non vide et compact pour la topologie faible*. Le théorème suivant permet de caractériser les mesures T-invariantes ergodiques

Théorème 3.4.1 *Si (X, d) est un espace métrique compact et si $T : X \to X$ est continue, une mesure de probabilité T-invariante est ergodique si et seulement si elle est extrémale.*

Démonstration. — Supposons que μ est T-ergodique et qu'il existe une décomposition $\mu = t\mu_1 + (1-t)\mu_2$, $0 < t < 1$, μ_1, μ_2 étant T-invariantes. La mesure μ_1 est absolument continue par rapport à la mesure μ et d'après le théorème de Radon-Nikodym, il existe une fonction $f \in L^1(\mu)$, $f \geq 0$ telle que $d\mu_1 = f d\mu$. Montrons que f est T-invariante. Observons que comme μ_1 est T-invariante on a pour toute fonction $\phi \in L^\infty$

$$\int \phi d\mu_1 = \int \phi \circ T d\mu_1$$

c'est-à-dire

$$\int \phi f d\mu = \int \phi \circ T \cdot f d\mu$$

et comme μ est T-invariante

$$\int \phi \circ T \cdot f \circ T d\mu = \int \phi \circ T \cdot f d\mu.$$

Posons à présent $\phi = \mathbf{1}_{f>\lambda}$. On a

$$\int_{f\circ T>\lambda} f \circ T d\mu = \int_{f\circ T>\lambda} f d\mu$$

et puisque μ est T-invariante

$$\int_X (f-\lambda)^+ d\mu = \int_{f\circ T>\lambda} (f\circ T-\lambda)^+ d\mu = \int_{f\circ T>\lambda} (f-\lambda) d\mu.$$

On a donc pour tout λ

$$\{f > \lambda\} = \{f \circ T > \lambda\},$$

modulo un ensemble de μ-mesure nulle. Par conséquent, $f = f \circ T$, μ-pp.

La réciproque est plus facile à démontrer. Supposons que μ ne soit pas T-ergodique. Il existe donc un ensemble A dans la tribu, T-invariant et tel que $0 < \mu(A) < 1$. Si on pose $\mu_1 = \mu(\cdot \cap A)/\mu(A)$ et $\mu_2 = \mu(\cdot \cap A^c)/\mu(A^c)$, on a $\mu = t\mu_1 + (1-t)\mu_2$ avec $t = \mu(A)$, ce qui contredit le fait que μ est un point extrémal. □

Un corollaire du théorème précédent est l'existence de mesure ergodique pour toute transformation continue sur un espace compact.

Corollaire 3.4.1 *Si (X,d) est un espace métrique compact et $T : X \to X$ une application continue, il existe une mesure qui est T-ergodique.*

Démonstration. — Cela résulte du théorème de Krein-Milman qui affirme que tout compact, convexe d'un espace vectoriel topologique admet des points extrémaux[1]. Dans le cas qui nous intéresse, on peut le démontrer directement. Choisissons $(\phi_n)_n$ une suite de fonctions continues dense dans $C^0(X)$. L'ensemble $\mathcal{M}_0$ des mesures de probabilités μ sur X telles que $\langle \mu, \phi_0 \rangle = \sup_{\nu\in\mathcal{M}} \langle \nu, \phi_0 \rangle$ est non vide puisque $\mathcal{M}$ est compact pour la topologie faible* et est un espace convexe compact pour la topologie faible*. Par récurrence on construit $\mathcal{M}_p$ qui est l'ensemble non vide convexe compact pour la topologie faible* constitué des mesures $\nu \in \mathcal{M}_{p-1}$ telles que $\langle \mu, \phi_p \rangle = \sup_{\nu\in\mathcal{M}_{p-1}} \langle \nu, \phi_p \rangle$. Notons $\mathcal{M}_\infty$ l'intersection des $\mathcal{M}_p$, $p \geq 0$. C'est toujours un ensemble non vide convexe compact pour la topologie faible*. Démontrons qu'il est constitué

1. et qu'il est l'enveloppe convexe de ses points extrémaux

de points extrémaux. Supposons par l'absurde que ce ne soit pas le cas et que l'on ait une écriture $\mu = t\mu_1 + (1-t)\mu_2$ avec $0 < t < 1$. Il est facile de voir que pour tout p, $\langle \mu, \phi_p \rangle = \langle \mu_1, \phi_p \rangle = \langle \mu_2, \phi_p \rangle$ (utiliser la définition de $\mathcal{M}_p$) et comme la suite ϕ_p est dense dans $C^0(X)$ on a $\mu = \mu_1 = \mu_2$. □

3.4.2 Points génériques

Supposons que (X, d) soit un espace métrique compact, et soient T une transformation mesurable sur X, μ une mesure T-invariante ergodique et $(\phi_k)_{k \geq 0}$ une suite de fonctions continues sur X dense dans $C^0(X)$. Le Théorème de Birkhoff nous apprend qu'il existe un ensemble E_k de μ-mesure 1 tel que pour tout $x \in E_k$, $(1/n)S_n\phi_k(x)$ converge vers $\int_X \phi_k d\mu$. L'ensemble $E = \cap_k E_k$ est de μ-mesure 1 et comme la suite des ϕ_k est C^0-dense dans $C^0(X)$, il est clair que pour toute $\phi \in C^0(X)$ et tout $x \in X$, la suite $(1/n)S_n\phi(x)$ converge également vers $\int_X \phi d\mu$. On dit que l'ensemble E est un ensemble générique pour (T, μ).

Un corollaire du résultat précédent est le suivant :

Corollaire 3.4.2 *Si μ et ν sont deux mesures de probabilité T-invariantes et ergodiques elles sont mutuellement singulières[2] ou égales.*

Démonstration. — Supposons que ce ne soit pas le cas. L'intersection E de l'ensemble des points réguliers de μ et de l'ensemble des points réguliers de ν est alors de μ-mesure et de ν-mesure positive. Pour toute fonction continue ϕ et $x \in E$, on a donc convergence de $(1/n)S_n\phi(x)$ vers $\int_X \phi d\mu$ et $\int_X \phi d\nu$. Par conséquent, $\mu = \nu$. □

3.4.3 Unique ergodicité

Définition 3.4.2 *On dit qu'un système dynamique (X, T, μ) est uniquement ergodique, si μ est l'unique mesure de probabilité invariante par T.*

Puisque l'ensemble des mesures ergodiques est l'ensemble des mesures extrémales on a

Proposition 3.4.1 *Si (X, T, μ) est uniquement ergodique, il est ergodique.*

Exemple : La translation $R_\alpha : x \mapsto x + \alpha$ sur le cercle $\mathbb{R}/\mathbb{Z}$ avec α irrationnel admet la mesure de Lebesgue comme unique mesure invariante. En effet, si $(R_\alpha)_*\mu = \mu$, on a $(R_{n\ a})_*\mu = \mu$ et comme la suite $n\alpha$ est dense sur le cercle,

2. pour tout borélien, $\mu(A) > 0$ implique $\nu(A) = 0$ et $\nu(A) > 0$ implique $\mu(A) = 0$

il est facile de voir que pour tout β sur le cercle $(R_\beta)_*\mu = \mu$. Mais la mesure de Haar est l'unique mesure invariante par toute translation.

Théorème 3.4.2 *Soit (X, d) un espace métrique compact et $T : X \to X$ une transformation continue et μ une mesure T-invariante. Les trois propriétés suivantes sont équivalentes :*

i) μ est l'unique mesure de probabilité invariante par T ;

ii) pour toute fonction $\phi \in C(X)$ et tout $\epsilon > 0$ il existe des fonctions $\psi, \eta \in C(X)$ telles que

$$\phi = \psi \circ T - \psi + \eta + \int_X \phi d\mu, \qquad \|\eta\|_0 \leq \epsilon;$$

iii) pour toute fonction continue $\phi \in C(X)$, les moyennes de Birkhoff de ϕ, $(1/n)S_n\phi(\cdot)$ convergent uniformément vers $\int_X \phi d\mu$.

Le fait que pour toute fonction continue $\phi \in C(X)$, les moyennes de Birkhoff de ϕ, $(1/n)S_n\phi(\cdot)$ convergent uniformément vers une constante est équivalent à l'unique ergodicité de (T, μ).

Démonstration. —

$i) \Longrightarrow ii)$: Notons $\mathcal{E}$ l'ensemble des fonctions continues de X de μ-intégrale nulle et $\mathcal{F}$ l'adhérence pour la topologie C^0 de l'ensemble des fonctions de la forme $\psi \circ T - \psi$. Si $\mathcal{E} \neq \mathcal{F}$, le théorème de Hanh-Banach nous enseigne qu'il existe une forme linéaire non nulle $\Lambda \in \mathcal{E}^*$ dont la restriction à $\mathcal{F}$ est nulle. D'après le théorème de représentation de Riesz, il existe une mesure (réelle) ν telle que pour tout $\phi \in \mathcal{E}$, $\Lambda\phi = \int_X \phi d\nu$. Par conséquent, pour toute fonction $\phi = \psi \circ T - \psi$ où ψ est continue on a $\int_X \phi d\nu = 0$, et donc $\int_X \psi \circ T d\nu = \int_X \psi d\nu$. On a donc $T_*\nu = \nu$. La mesure ν admet une écriture unique de la forme $\nu = \nu_+ - \nu_-$ où $\nu_\pm$ sont des mesures boréliennes positives telles que pour tout borélien A, $\nu_\pm(A) = \pm\nu(A \cap E_\pm)$, où $E_\pm$ sont des boréliens. Démontrons que $\nu_\pm$ sont T-invariantes. Déjà,

$$\nu_+(T^{-1}E^+) \geq \nu(T^{-1}E_+) = \nu(E_+) = \nu_+(E_+)$$

et donc $\nu_+(T^{-1}E_+ \Delta E_+) = 0$. En outre,

$$\nu_+(T^{-1}E_+) - \nu_-(T^{-1}E_+) = \nu(T^{-1}E_+) = \nu(E_+) = \nu_+(E_+)$$

et comme $\nu_+(E_+) = \nu_+(T^{-1}E_+)$ on en déduit $\nu_-(T^{-1}E_+) = 0$ et $\nu(T^{-1}E_+ \Delta E_+) = 0$. Pour tout borélien A

$$\nu_+(T^{-1}A) = \nu(T^{-1}A \cap E^+) = \nu(T^{-1}A \cap T^{-1}E_+) = \nu(A \cap E_+) = \nu_+(A)$$

. On verrait de même que ν_- est T-invariante. Comme par hypothèse (T,μ) est uniquement ergodique, on a $\nu_+ = \mu$. Mais alors, Λ est nulle sur $\mathcal{E}$, ce qui est une contradiction.

$ii) \Longrightarrow iii)$: C'est clair.

$iii) \Longrightarrow i)$: Si ν est une mesure de probabilité T-invariante, $\int_X (1/n)S_n\phi d\nu = \int_X \phi d\nu$. De la convergence uniforme des moyennes de Birkhoff de ϕ vers $\int_X \phi d\mu$ on déduit $\mu = \nu$.

□

Skew-shift

Pour $\alpha \in \mathbb{T}$ irrationnel, notons $T = T_\alpha : \mathbb{T}^2 \to \mathbb{T}^2$ défini par $T(x,y) = (x+\alpha, y+x)$. Les itérés de T se calculent aisément : pour $n \geq 0$,

$$T^n(x,y) = (x+n\alpha, y+nx+n(n-1)\alpha/2).$$

Si m désigne la mesure de Lebesgue, il est clair que m est T-invariante puisque le jacobien de T est constant égal à 1. Démontrons que (T,m) est ergodique : si f est une fonction bornée T-invariante, il est facile de voir que ses coefficients de Fourier vérifient $\hat{f}(k+l,l) = e^{2\pi ik\alpha}\hat{f}(k,l)$; comme ils sont dans $l^2(\mathbb{Z}^2)$ on voit facilement que f est constante.

Démontrons à présent

Théorème 3.4.3 *Si α est irrationnel, T_α est uniquement ergodique.*

Démonstration. — Il suffit pour cela de démontrer que si ν est une mesure de probabilité T-invariante elle est égale à m. Si $\pi : \mathbb{T}^2 \to \mathbb{T}$ est la projection suivant la première variable, on voit que $\pi \circ T_\alpha = R_\alpha \circ \pi$ où R_α est la translation d'angle α sur $\mathbb{T}$. On a donc $(R_\alpha)_*(\pi_*\nu) = (\pi_*\nu)$ et comme la mesure de Lebesgue est l'unique mesure invariante par R_α (α est irrationnel) on a nécessairement $\pi_*\nu = \mathrm{Leb}$: ν se projette par le premier facteur sur la mesure de Lebesgue. Ainsi, pour tout borélien $I \subset \mathbb{T}$, $\nu(\pi^{-1}I) = \mathrm{Leb}(I)$. Notons E_m (resp. E_ν) l'ensemble des points m-génériques (resp. ν-générique) pour T. Le théorème de Fubini implique qu'il existe un ensemble $I \subset \mathbb{T}$ de mesure de Lebesgue 1 tel que pour tout $x \in I$ la mesure de Lebesgue de la fibre $\pi^{-1}(x) \cap E_m$ égale 1.Si on note J l'ensemble des $x \in I$ pour lesquels $\pi^{-1}(x) \cap E_\nu = \emptyset$, on a $\mathrm{Leb}_1(J) = \nu(\pi^{-1}(J)) = 0$; en notant $\tilde{I} = I - J$ on voit que $\nu(\pi^{-1}(\tilde{I})) = 1$ si bien que $E_\nu = \pi^{-1}(\tilde{I}) \mod \nu$. Comme pour $x \in \tilde{I}$, $\pi^{-1}(x) \in E_m$ on a que pour tout $(x,y) \in \mathbb{T}^2$ avec $x \in \tilde{I}$ il existe une suite $y_n \in \mathbb{T}$ convergeant vers y telle que chaque (x,y_n) est m-générique. On a donc pour toute fonction continue $f : \mathbb{T}^2 \to \mathbb{R}$,

$$\lim_{N\to\infty} \frac{1}{N}\sum_{k=0}^{N} f(x+k\alpha, y+kx+k(k-1)\alpha) = \tilde{f}_\nu(x,y),$$

$$\lim_{N\to\infty} \frac{1}{N}\sum_{k=0}^{N} f(x+k\alpha, y_n+kx+k(k-1)\alpha) = \int f dm,$$

et du fait de l'uniforme continuité de f

$$\limsup_{N\to\infty} \frac{1}{N}\sum_{k=0}^{N} |f(x+k\alpha, y+kx+k(k-1)\alpha) - f(x+k\alpha, y_n+kx+k(k-1)\alpha)| \le \delta(|y-y_n|),$$

où δ est une fonction tendant vers 0 en 0. Ceci démontre que pour tout $x \in \tilde{I}$, la fibre entière $\pi^{-1}(x)$ est dans E_m et donc que ν-presque tout point de E_ν est appartient à E_m. Mais ceci entraîne de façon claire que $\nu = m$ puisque le théorème de convergence dominée montre que pour toute fonction continue f, $\int_X f d\nu = \int_X (1/n) S_n f d\nu$ converge vers $\int_X f dm$.

□

On peut donner une preuve qui s'inspire du point de vue spectral. Supposons que ν soit une mesure invariante par T et notons $U : L^2(\mathbb{T}^2, \nu) \to L^2(\mathbb{T}^2, \nu)$ l'opérateur unitaire[3] défini par $Uf = f \circ T$. Pour $r = (k,l) \in \mathbb{Z}^2$ $z = (x,y)$, notons $e_r(z) = e_{k,l}(x,y) = \exp(2\pi(kx+ly))$ et $\hat{\nu}(r) = \langle \nu, e_{-r}\rangle = \int_{\mathbb{T}^2} e^{-2\pi i\langle r,z\rangle} d\nu(z)$ les coefficients de Fourier de ν (rappelons qu'une mesure est caractérisée par ses coefficients de Fourier).

Si $r \neq 0$ est fixé, les vecteurs $(U^n e_r)_{n\ge1}$ sont deux à deux orthogonaux. En effet, $\langle U^n e_r, U^m e_r\rangle = \langle U^{n-m} e_r, e_r\rangle$. Comme $Ue_{(k,l)} = e^{2\pi ik\alpha} e_{(k+l,l)}$ on voit que $U^{(n-m)} e_r$ est de la forme $e^{2\pi i\beta} e_{(k',l)}$ (avec $k' = k + (n-m)l$) pour un certain β réel, si bien que $\langle U^{n-m} e_r, e_r\rangle = e^{2\pi i\beta}\langle e_{(k',l)}, e_{(k,l)}\rangle$. Remarquons que $\langle e_{(k',l)}, e_{(k,l)}\rangle = \langle e_{(k',0)}, e_{(k,0)}\rangle$. Mais on a par ailleurs

$$\langle e_{(k',0)}, e_{(k,0)}\rangle = \langle Ue_{(k',0)}, Ue_{(k,0)}\rangle = e^{(2\pi i(k'-k)\alpha)}\langle e_{(k',0)}, e_{(k,0)}\rangle$$

ce qui entraîne que $\langle e_{(k',0)}, e_{(k,0)}\rangle$ égale 0 si $k \neq k'$ et 1 sinon. On a donc bien démontré que les vecteurs $(U^n e_r)_{n\ge1}$ sont deux à deux orthogonaux. Projetons le vecteur 1 sur l'espace engendré par les $(U^n e_r)_{n\ge1}$: d'après l'inégalité de Parseval-Bessel

$$\sum_{n\ge0} |\langle U^n e_r, 1\rangle|^2 \le \|1\|^2_{L^2(\nu)} = 1.$$

3. on munit $L^2(X,\nu)$ du produit hermitien $\langle f,g\rangle = \int_{\mathbb{T}^2} f(z)\bar{g}(z)d\nu(z)$

Mais, $\langle U^n e_r, 1\rangle = \langle U^n e_r, U^n 1\rangle = \langle e_r, 1\rangle = \hat{\nu}(-r)$. On a donc démontré que pour tout $r \in \mathbb{Z}^2$, $\hat{\nu}(r) = 0$ si $r \neq 0$ et $\hat{\nu}(0) = 1$. Ces relations caractérisent la mesure de Lebesgue sur $\mathbb{T}^2$.

Equirépartition

Définition 3.4.3 *Une suite (x_n) de points dans $[0,1]^d$ est équirépartie si pour tout pavé $I \subset [0,1]^d$ on a*

$$\lim_{N\to\infty} \frac{\#\{k \in \{1,\dots,N\} : x_k \in I\}}{N} = \mathrm{vol}_d(I)$$

Un critère d'équirépartition est le suivant (preuve laissée en exercice au lecteur).

Théorème 3.4.4 *Les propriétés suivantes sont équivalentes :*
i) La suite (x_n) est équirépartie ;
ii) Pour toute fonction Riemann intégrable f sur $[0,1]^d$ la suite $\frac{1}{n}\sum_{k=1}^n f(x_k)$ converge vers $\int_{[0,1]^d} f(x)dx$.
iii) Pour toute fonction continue f sur $[0,1]^d$ la suite $\frac{1}{n}\sum_{k=1}^n f(x_k)$ converge vers $\int_{[0,1]^d} f(x)dx$.
iv) Pour tout polynôme trigonométrique P la suite $\frac{1}{n}\sum_{k=1}^n P(x_k)$ converge vers $\int_{\mathbb{T}^d} P(x)dx$.

Comme corollaire des deux sous-sections précédentes on a

Corollaire 3.4.3 *Si α est irrationnel, la suite $n(n-1)\alpha/2$ est équirépartie sur $[0,1]$*

Démonstration. — Il suffit de poser $f(x,y) = e^{2\pi i r}y$ et d'appliquer les théorèmes 3.4.2, 3.4.3, 3.4.4 □

On peut démontrer par cette méthode que si $P(n)$ est un polynôme dont le coefficient du monôme du plus haut degré est irrationnel, alors la suite $(P(n))$ est équirépartie. On introduira pour cela $T : \mathbb{T}^d \to \mathbb{T}^d$ défini par

$$T : (\theta_1, \theta_2, \dots, , \theta_d) \mapsto (\theta_1 + \alpha, \theta_2 + \theta_1, \dots, \theta_d + \theta_{d-1}),$$

et on démontrera que l'unique mesure de probabilité T-invariante est la mesure de Lebesgue sur $\mathbb{T}^2$. Si P est de degré d, on pose $P_d = P$, et $P_j = P_{j+1}(X+1) - P_{j+1}(X)$, $j = N-1, \dots, 0$. On a $P_0(X) = \alpha$ où on a noté $\alpha/N!$ le coefficient dominant de P. Si on pose $\theta_n = P_1(n), \dots, P_d(n))$ on a $T^n\theta_0 = \theta_n$. On conclut alors comme précédemment.

3.5 Mélange

Définition 3.5.1 *Un système dynamique* $(X, \mathcal{A}, T, m)$ *est dit mélangeant si pour tous boréliens* $A, B \in \mathcal{A}$ *on a*

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

Il n'est pas difficile de prouver que

Proposition 3.5.1 *Un système dynamique est mélangeant si et seulement si pour toutes fonctions* $f, g \in L^2(X, m)$ *on a*

$$\lim_{n\to\infty} \int_X f \circ T^n \bar{g} dm = \int_X f dm \int_X \bar{g} dm$$

Théorème 3.5.1 *Si un système dynamique est mélangeant il est ergodique.*

Démonstration. — En effet, si A est un borélien T-invariant on a $\lim_{n\to\infty} m(T^{-n}A \cap A) = m(A)m(A)$ ce qui s'écrit $m(A) = m(A)^2$ et donc $m(A)$ égale 0 ou 1. □

Dans un cadre probabiliste il s'agit de la loi du 0,1 de Kolmogorov.

Un exemple important de systèmes mélangeants est fourni par les sous-shifts de type fini.

Théorème 3.5.2 (Perron-Frobenius) *Si* $P \in M_r(\mathbb{R})$ *est une matrice stochastique irréductible*[4] *alors il existe une unique mesure stationnaire, c'est-à-dire un unique vecteur* $p \in \mathbb{R}^n$ *à coordonnées positives et dont la somme des composantes vaut 1 tel que* $pP = p$. *En outre, si* P *est apériodique*[5] *alors pour toute mesure de probabilité* q *sur* $\{1, \ldots, r\}$, ν, $\lim_{n\to\infty} qP^n = p$.

Démonstration. — Supposons P irréductible, et faisons l'hypothèse qu'il existe deux mesures de probabilités p_1, p_2 sur $\{1, \ldots, r\}$ différentes telles que $p_1 = p_1 P$ et $p_2 = p_2 P$. Si on appelle x le vecteur ligne $x = p_1 - p_2$ on a $x = xP$ et $\sum_{i\in E} x_i = 0$; en particulier il existe deux indices i, j tels que x_i et x_j sont de signes opposés. Puisque la matrice P est irréductible il existe

4. ce qui signifie que le graphe orienté de P est connexe
5. $\exists m \geq 0,\ \forall i, j,\ (P^m)_{ij} > 0$

un exposant m tel que $(P^m)_{ij} > 0$. Ecrivons $x = xP^m$ puis,

$$\begin{aligned}\sum_{j\in E} |x_j| &= \sum_{j\in E} |\sum_{k\in E} x_k (P^m)_{kj}| \\ &\le \sum_{j\in E}\sum_{k\in E} |x_k||(P^m)_{kj}| \\ &\le \sum_{k\in E}\sum_{j\in E} |x_k|(P^m)_{kj} \\ &\le \sum_{k\in E} |x_k|\end{aligned}$$

puisque la matrice P est stochastique (noter que P^m l'est aussi et donc $\sum_{j\in E}(P^m)_{ij} = 1$). Or, cette dernière inégalité est une égalité ; par conséquent dans la suite d'inégalités que nous avons écrites nous avions en fait déjà des égalités. Il en résulte que pour tout j,

$$|x_j| = |\sum_{k\in E} x_k(P^m)_{kj}| = \sum_{k\in E} |x_k|(P^m)_{kj},$$

ce qui n'est possible que si les $x_k(P^m)_{kj}$, sont tous de même signe quand k varie dans le sous-ensemble de E constitué des k pour lesquels $|x_k|(P^m)_{kj}$ est non nul. Puisque $x_i \neq 0$ et que par définition de m le coefficient $(P^m)_{ij} > 0$, ceci entraîne que x_j est de même signe que x_i ce qui est une contradiction et démontre l'unicité.

L'existence se démontre en choisissant une probabilité q sur $\{1, \dots, r\}$ et en extrayant de la suite $(1/n)(q + \cdots + P^{n-1}q)$ une sous-suite convergente vers un certain p qui vérifie nécessairement $pP = p$.

Supposons à présent P apériodique. Notons $\mathcal{M}_0$ l'ensemble des vecteurs lignes x tels que $\sum_{i\in E} x_i = 0$. Définissons pour $x \in \mathcal{M}_0$ la norme suivante :

$$\|x\| = \frac{1}{2}\sum_{i\in E} |x_i| = \sum_{i\in E} (x_i)^+.$$

(Avec la notation $z^+ = \max(0, z)$). L'application $x \mapsto xP$ envoie $\mathcal{M}_0$ dans lui même. On a le lemme suivant

Lemme 3.5.1 *Soit Q une matrice stochastique telle que $\alpha = \min_{i,j\in E} Q_{ij} > 0$. Alors, Q est une $(1-\alpha)$-contraction : pour tout $x \in \mathcal{M}_0$*

$$\|xQ\| \le (1-\alpha)\|x\|.$$

Démonstration. — On a

$$\begin{aligned}
\|xQ\| &= \sum_{i:(xQ)_i>0} (xQ)_i \\
&= \sum_{i:(xQ)_i>0} \sum_{k\in E} x_k (Q)_{ki} \\
&\leq \sum_{i:(xQ)_i>0} \sum_{k:x_k>0} |x_k|(Q)_{ki}| \\
&\leq \sum_{k:x_k>0} \sum_{i:(xQ)_i>0} x_k (Q)_{ki}
\end{aligned}$$

Comme $xQ \in \mathcal{M}_0$ on a $\sum_{i\in E}(xQ)_i = 0$ et l'ensemble des $i \in E$ pour lesquels $(xQ)_i > 0$ n'est pas E tout entier. Par conséquent,

$$\begin{aligned}
\sum_{i:(xQ)_i>0} (Q)_{ki} &= \sum_{i\in E}(Q)_{ki} - \sum_{i:(xQ)_i\leq 0} (Q)_{ki} \\
&= 1 - \sum_{i:(xQ)_i\leq 0} (Q)_{ki} \\
&\leq (1-\alpha).
\end{aligned}$$

Revenant aux inégalités précédentes

$$\|xQ\| \leq (1-\alpha)\|x\|.$$

□

Puisque P est apériodique il existe un entier m tel que $Q = P^m$ soit à coefficients strictement positifs. L'application $x \mapsto xP^m$ de $(\mathcal{M}_0, \|\cdot\|)$ dans lui même est donc une $(1-\alpha)$-contraction et par conséquent pour tout entier l, $\|xP^{lm}\| \leq (1-\alpha)^l\|x\|$. Si p est l'unique mesure de probabilité stationnaire et q une mesure de probabilité $p - q \in \mathcal{M}_0$ et donc pour tous entiers l, c

$$\|(p-q)P^{lm+c}\| \leq (1-\alpha)^l \|(p-q)P^c\|$$

ce qui s'écrit (en utilisant $pP^n = p$)

$$\|p - qP^{lm+c}\| \leq (1-\alpha)^l \|(p-q)P^c\|.$$

Comme tout entier n s'écrit de façon unique $n = lm + c$ avec $0 \leq c < m$, $l \geq n/m$ (division euclidienne de n par m) et comme $0 \leq (1-\alpha) < 1$, la suite $p - qP^n$ converge vers 0.

□

Comme corollaire on obtient :

Théorème 3.5.3 *Si P est une matrice stochastique et A la matrice de transition associée à P ($A_{i,j} = 1$ ssi $P_{ij} > 0$) notons p l'unique mesure stationnaire telle que $pP = p$ et μ_p la mesure correspondante sur $(\Sigma_A, \mathcal{B}or)$. Le système dynamique $(\Sigma_A, \mathcal{B}or, \sigma, \mu_p)$ est ergodique si P est transitive et mélangeant si P est apériodique.*

Démonstration. — Démontrons la partie mélange. Il suffit de démontrer que pour tous cylindres $C = C(\epsilon_0, \ldots, \epsilon_n)$, $C' = C(\epsilon'_0, \ldots, \epsilon'_{n'})$ on a

$$\lim_{k\to\infty} \mu_p(C \cap \sigma^{-k}C') = \mu_p(C)\mu_p(C').$$

Si k est assez grand ($k \geq n$), $C \cap \sigma^{-k}C'$ est le cylindre $C'' = C(\epsilon_0, \ldots, \epsilon_n, *, \epsilon_k, \ldots, \epsilon_{k+n'})$ dont la μ_p mesure vaut

$$\begin{aligned}\mu_p(C'') &= p_{\epsilon_0} P_{\epsilon_0\epsilon_1} \cdots P_{\epsilon_{n-1}\epsilon_n} (P^{k-n})_{\epsilon_n\epsilon_k} P_{\epsilon_k\epsilon_{k+1}} \cdots P_{\epsilon_{k+n'-1}\epsilon_{k+n'}} \\ &= p_{\epsilon_0} P_{\epsilon_0\epsilon_1} \cdots P_{\epsilon_{n-1}\epsilon_n} (P^{k-n})_{\epsilon_n\epsilon'_0} P_{\epsilon'_0\epsilon'_1} \cdots P_{\epsilon'_{n'-1}\epsilon'_{n'}}\end{aligned}$$

Mais $q_j(P^{k-n})_{ji}$ converge vers p_i quand k tend vers l'infini ; ainsi, quand k tend vers l'infini $m(C'')$ tend vers

$$p_{\epsilon_0} P_{\epsilon_0\epsilon_1} \cdots P_{\epsilon_{n-1}\epsilon_n} p_{\epsilon'_0} P_{\epsilon'_0\epsilon'_1} \cdots P_{\epsilon'_{n'-1}\epsilon'_{n'}}$$

qui vaut $\mu_p(C)\mu_p(C')$.

La partie (T, μ_p) ergodique se démontre de la façon suivante. La preuve précédente et le fait que pour tout q, $\lim_{n\to\infty}(1/n)(q + qP + \cdots + qP^{n-1}) = p$ (**Exercice :** Pourquoi ?) permet de démontrer que pour tous cylindres $C = C(\epsilon_0, \ldots, \epsilon_n)$, $C' = C(\epsilon'_0, \ldots, \epsilon'_{n'})$ on a

$$\lim_{n\to\infty} \sum_{k=0}^{n-1} \mu_p(C \cap \sigma^{-k}C') = \mu_p(C)\mu_p(C').$$

Cela suffit pour démontrer l'ergodicité.

Exercice : Démontrer le. [On pourra démontrer que cela implique que pour toutes fonctions $\varphi, \psi \in L^2(\mu_p)$ on a

$$\lim_{n\to\infty} \sum_{k=0}^{n-1} \int_\Sigma \varphi\psi d\mu_p = \int_X \varphi d\mu_p \int_\Sigma \psi d\mu_p.$$

En appliquant le théorème ergodique (de von Neumann) on démontrera que cela implique l'ergodicité).]

□

Exercice Démontrer que si (X, T) est une dynamique topologique, et μ une mesure T-invariante alors le fait que (T, μ) est ergodique (resp. uniquement ergodique, resp. mélangeante) implique que T restreinte à $supp \mu$ est transitive (resp. minimale, resp. topologiquement mélangeante).

Chapitre 4

Théorie spectrale

Dans ce qui suit $(X, \mathcal{B}, \mu, T)$ est un système dynamique mesurable, X étant un espace de Lebesgue. L'idée de la théorie specrale est d'étudier ce système dynamique *via* les propriétés spectrales de l'isométrie $U_T : L^2(X, \mu) \to L^2(X, \mu)$, $\phi \mapsto \phi \circ T$. Nous supposerons dans la suite que T est inversible (T^{-1} étant $\mathcal{B}$ mesurable) de façon que U_T est en fait un opérateur unitaire $U_T^* = U_T^{-1}$. On peut reformuler les notions d'ergodicité, de mélange ou de mélange faible au moyen de U_T. Si on note P la projection orthogonale de $L^2(X, \mu)$ sur $\ker(U_T - Id) = L^2(\mathcal{I}, \mu)$ ($\mathcal{I}$ étant la tribu des invariants de T) :
i) $(X, \mathcal{B}, \mu, T)$ est *ergodique* si et seulement si, la tribu des invariants $\mathcal{I}$ égale la tribu triviale $\{\emptyset, X\}$; d'après le théorème de Von Neumann ceci est équivalent au fait que pour tout $\phi \in L^2(X, \mu)$

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} U_T^k \phi = P\phi = \langle \phi, 1 \rangle 1$$

au sens L^2, où P est la projection orthogonale sur l'espace des fonctions constantes. En fait, $(X, \mathcal{B}, \mu, T)$ est *ergodique* si et seulement si pour tout $\phi \in L^2(X, \mu)$, $S_n\phi/n$ converge faiblement vers $P\phi$: pour tout $\psi \in L^2(X, \mu)$

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \langle U_T^k \phi, \psi \rangle = \langle P\phi, \psi \rangle.$$

ii) $(X, \mathcal{B}, \mu, T)$ est mélangeant si pour tous $\phi, \psi \in L^2(X, \mu)$

$$\lim_{n\to\infty} \langle U_T^n \phi, \psi \rangle = \langle \phi, 1 \rangle \overline{\langle \psi, 1 \rangle} = \langle P\phi, \psi \rangle,$$

(c'est-à-dire $U_T^n \phi$ converge faiblement vers $P\phi$.

4.1 Le théorème spectral

Le théorème fondamental est le suivant :

Théorème 4.1.1 (Théorème spectral) *Si H est un espace de Hilbert (séparable) et $U : H \to H$ un opérateur unitaire ($U* = U^{-1}$) alors il existe*

a) $r \in \mathbf{N} \cup \{\infty\}$ et des mesuresν_i, $1 \le i \le r$ telles que $\nu_1 >> \nu_2 >> \dots$

b) une isométrie bijective Λ entre H et $\mathcal{E} = \bigoplus_{i=1}^{r} L^2(S^1, \mathcal{B}or, \nu_i)$ qui conjugue U (qui agit sur H) à l'opérateurM_z de multiplication par z (qui agit sur $\mathcal{E}$) : si $\Lambda v = \phi(\cdot)$ alors $(\Lambda H v)(z) = z\phi(z)$.

En outre cette décomposition est essentiellement unique : si (H, U) est isométriquement équivalent à $(\bigoplus_{i=1}^{\tilde{r}} L^2(S^1, \mathcal{B}or, \tilde{\nu}_i), M_z)$ avec $\tilde{\nu}_1 >> \tilde{\nu}_2 >> \dots$, alors $\tilde{r} = r$ et $\tilde{\nu}_i$ est équivalente à ν_i pour tout i.

La démonstration du théorème spectral que nous ne ferons pas complètement repose sur les résultats suivants

Théorème 4.1.2 (Existence des mesures spectrales) *1) Pour tout $x \in H$ il existe une mesure de probabilité ν_x telle que pour tout $n \in \mathbb{Z}$*

$$\langle U^n x, x \rangle = \hat{\nu}_x(n) := \int_{\mathbb{S}^1} z^{-n} d\nu(z) = \int_{\mathbb{R}/\mathbb{Z}} e^{-2\pi i n \theta} d\tilde{\nu}(\theta).$$

2) Pour tout $x \in H$ l'opérateur unitaire U restreint à l'espace cyclique $C(x) = \mathrm{Adh}\{U^k x, k \in \mathbf{Z}\}$ est isométriquement équivalent à l'opérateur de multiplication M_z agissant sur $L^2(\mathbb{S}^1, Bor(\mathbb{S}^1, \nu)$.

Démonstration. — 1) Nous travaillerons plutôt sur $\mathbf{R}/\mathbf{Z} \equiv S^1$ (on note $z = e^{2\pi i\theta}$). Posons pour $\|x\| = 1$

$$\nu_N(\theta) = \frac{1}{N} \left\| \sum_{k=0}^{N-1} (U^k x) e^{2\pi i k\theta} \right\|^2.$$

Par définition $\nu_N \ge 0$ et puisque

$$\begin{aligned}\nu_N(\theta) &= \frac{1}{N} \sum_{0 \le k, l \le N-1} \langle U^k, U^l x \rangle e^{2\pi i(k-l)\theta} \\ &= \|x\|^2 + \frac{1}{N} \sum_{0 \le k \ne l \le N-1} \langle U^{k-l} x, x \rangle e^{2\pi i(k-l)\theta}\end{aligned}$$

on voit que $\int_{\mathbf{T}} \nu(\theta)d\theta = 1$. Ainsi la mesure $\nu_N = \nu_N(\theta)d\theta$ est de probabilité. Le calcul précédent montre que

$$\begin{aligned}\hat{\nu}_N(n) &= \frac{1}{N} \sum_{\substack{0\leq k\neq l\leq N-1 \\ k-l=n}} \langle U^{k-l}x, x\rangle \\ &= \frac{N-n}{N}\langle U^n x, x\rangle\end{aligned}$$

et donc pour tout $n \in \mathbf{Z}$, $\lim_{n\to\infty} \hat{\nu}_N(n) = \langle U^n x, x\rangle$. Ceci implique que ν_N converge faiblement vers une mesure de probabilité ν et que $\hat{\nu}(n) = \langle U^n x, x\rangle$. [1]

2) Pour tout polynôme trigonométrique P on a $\nu(P) = \langle P(U)x, x\rangle$. Par ailleurs,

$$\begin{aligned}\|P(U)x\|^2 &= \langle (P(U))^* P(U)x, x\rangle \\ &= \langle \bar{P}(U^{-1})P(U)x, x\rangle \\ &= \nu(Q) \qquad (Q(z) := \bar{P}(z^{-1})P(z)) \\ &= \nu(|P|^2)\end{aligned}$$

car pour $z \in \mathbf{S}^1$ $(|z| = 1)$ $\bar{P}(z^{-1})P(z) = |P|^2(z)$. On a donc démontré que pour tout polynôme trigonométrique

$$\|P(U)x\|^2 = \int_{\mathbf{S}^1} |P(z)|^2 d\nu(z).$$

Si $P_n(U)x$ est une suite de $C(x)$ convergeant vers $y \in C(x)$, la suite $P_n(U)x$ est de Cauchy et d'après l'inégalité précédente il en est de même de $P_n(\cdot) \in L^2(\mathbf{S}^1, \nu)$. Il existe donc $\phi \in L^2(\mathbf{S}^1, \nu)$ tel que $\lim_{n\to\infty} \|P_n - \phi\|_{L^2(\nu)} = 0$. Si Q_n est une autre suite de polynôme trigonométrique telle que $Q_n(U)x$ converge vers y il est facile de voir que Q_n qui est de Cauchy dans $L^2(\mathbf{S}^1, \nu)$ converge également vers ϕ [2]. Si on pose $\Lambda y = \phi$ il est facile de voir que Λ est linéaire et par définition $\|y\| = \|\Lambda y\|_{L^2(\mathbf{S}^1,\nu)}$.

□

Proposition 4.1.1 *Si x et y sont dans H, notons μ_x et μ_y des mesures spectrales associées à x et y comme cela a été fait dans le théorème précédent. Alors,*

1. Pour tout polynôme trigonométrique la limite $\lim_{N\to\infty} \nu_N(P)$ existe ; en outre, pour toutes fonctions continues $\phi, P : \mathbf{S}^1 \to \mathbf{C}$, on a $|\nu_N(\phi) - \nu_{N'}(\phi)| \leq |\nu_N(P) - \nu_{N'}(P)| + \sup_{\mathbf{S}^1} |\phi(z) - P(z)|$. Par conséquent, pour toute fonction continue ϕ la suite $\nu_N(\phi)$ converge vers un réel $\nu(\phi)$; d'après le théorème de Banach-Steinhaus ν est une mesure de probabilité.

2. car $\|P_n - Q_n\|_{L^2(\mathbf{S}^1,\nu)}$ converge vers 0

a) si $\mu_x \perp \mu_y$ *alors* $C(x) + C(y) \subset C(x+y)$.

b) si $C(x) \subset C(y)$ *alors* $\mu_x << \mu_y$.

Enfin mentionnons le théorème de Wiener :

Théorème 4.1.3 *Si* $H = L^2(\mathbf{S}^1, \mathcal{B}, \nu)$ *les sous-espaces fermés invariants par* $M : \phi(z) \mapsto z\phi(z)$ *sont les* $\mathbf{1}_B \cdot L^2(\mathbf{S}^1, \nu) = \{f \in L^2(\mathbf{S}^1, \mathcal{B}, \nu) : f|S^1 - B = 0\}$ *où* B *parcourt les boréliens* $B \subset \mathcal{B}$.

Démonstration. — Soit L un sous-espace fermé invariant par M e écrivons $1 = f + g$ avec $f \in L$, $g \in L^\perp$. On a $z^n f$
perpg pour tout n et donc pour tout polynôme trtigonométrique

$$\int_{\mathbf{S}^1} P(z) f(z) \overline{g(z)} d\nu(z) = 0$$

si bien que $f\bar{g} = 0$ ν-pp. Ainsi, $f = \mathbf{1}_B$, $g = \mathbf{1}_{B^c}$ pour $B \in Bor(\mathbf{S}^1)$. En outre, $z^n \mathbf{1}_B \in L$ et $z^n \mathbf{1}_{B^c} \in L^\perp$ pour tout n entraîne que $\mathbf{1}_B L^2(\mathbf{S}^1, \nu) \subset L$ et $\mathbf{1}_{B^c} L^2(\mathbf{S}^1, \nu) \subset L^\perp$. Comme $\mathbf{1}_B L^2(\mathbf{S}^1, \nu) + \mathbf{1}_{B^c} L^2(\mathbf{S}^1, \nu) = L^2(\mathbf{S}^1, \nu)$ on a $L = \mathbf{1}_B L^2(\mathbf{S}^1, \nu)$. □

4.2 Transformations à spectre discret

Définition 4.2.1 *On dit que* $(X, \mathcal{B}, \nu, T)$ *est à spectre discret si le spectre de* U_T *est purement ponctuel (la mesure spectrale maximale est une somme de mesures de Dirac).*

Proposition 4.2.1 *Si une transformation est à spectre discret et ergodique alors toutes ses valeurs propres sont simples, sur le cercle unité et constituent un sous-groupe multiplicatif de* $(\mathbb{T}, \cdot)$ *; les fonctions propres peuvent être choisies à valeurs dans* $\mathbb{T}$.

Démonstration. — Supposons en effet que $Uf = \lambda f$ et $Ug = \lambda g$. De façon claire f, g sont de module constant (d'après l'ergodicité) et on peut les supposer à valeurs sur $\mathbb{T}$. La fonction f/g est invariante par U et est donc constante. □

On dit que deux systèmes dynamiques $(X_i, \mathcal{B}_i, \nu_i, T_i)$, $i = 1, 2$ sont *spectralement isomorphes* s'il existe un opérateur unitaire $V : L^2(X_1, \nu_1) \to L^2(X_2, \nu_2)$ tel que

$$U_{T_2} = V \circ U_{T_1} \circ V^{-1}.$$

On dit qu'ils sont *métriquement isomorphes* s'il existe une application inversible $h : (X_1, \mathcal{B}_1, \nu_1) \to (X_2, \mathcal{B}_2, \nu_2)$ telle que

$$T_2 = h \circ T_1 \circ h^{-1}.$$

D'après le théorème spectral et la propriété précédente, deux systèmes dynamiques ergodiques et à spectre discret sont spectralement isomorphe si et seulement si ils ont le même spectre.

Théorème 4.2.1 *Si deux systèmes dynamiques à spectre discret sont ergodiques et spectralement isomorphes (ont le même spectre) alors ils sont mesurablement isomorphes.*

Théorème 4.2.2 *Un système dynamique ergodique et à spectre discret est métriquement isomorphe à une translation sur un groupe abelien compact.*

4.3 Mélange faible

Définition 4.3.1 *On dit que $(X, \mathcal{B}, \nu, T)$ est faiblement mélangeant (en abrégé f.m.) si 1 est l'unique valeur propre de U_T et si elle est simple (les seules fonctions propres sont les constantes).*

Ainsi, une transformation f.m est ergodique. Le théorème spectral permet de caractériser les transformations faiblement mélangeante. :

Théorème 4.3.1 *Les assertions suivantes sont équivalentes :*
i) $(X, \mathcal{B}, \mu, T)$ est faiblement mélangeant;
ii) pour tous $A, B \in \mathcal{B}$

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

iii) pour tous $A, B \in \mathcal{B}$ il existe un sous-ensemble $\mathcal{N}$ de $\mathbf{N}$ de densité 1[3] *tel que*

$$\lim_{\substack{n\to\infty \\ n\in\mathcal{N}}} \mu(T^{-k}A \cap B) = \mu(A)\mu(B);$$

iv) les mêmes assertions que ii) et iii) mais en remplaçant "tous $A, B \in \mathcal{B}$ par "pour tous A, B dans un ensemble engendrant la tribu $\mathcal{B}$"

3. un ensemble $A \in \mathbf{N}$ est de densité 1 si $(1/n)\sum_{k=0}^{n-1} \mathbf{1}_A(k)$ converge vers 1 quand $n \to \infty$

v) Pour toutes $f, g \in L^2(X,\mu)$

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} |\langle U^k f, g\rangle|^2 = \Big(\int_X f d\mu\Big)\Big(\int_X \bar{g} d\mu\Big).$$

Démonstration. — a) Nous allons d'abord démontrer $i)ssiv)$.

Supposons que $i)$ soit vraie : par polarisation il suffit de démontrer que pour toute $f \in L^2(X,\mu)$

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} |\langle U^k f, f\rangle|^2 = \Big(\int_X f(x) d\mu(x)\Big)^2$$

et il est facile de voir qu'il suffit de démontrer ceci pour f telle que $\int_X f(x)d\mu(x) = 0$. Utilisons le théorème spectral pour U restreinte à l'espace cyclique engendré par f : il existe une mesure de probabilité ν borélienne sur $\mathbf{S}^1$ et une fonction $\phi \in L^2(\mathbf{S}^1, \nu)$ telles que pour tout k

$$\langle U^k f, f\rangle = \int_{\mathbf{S}^1} z^k \phi(z)\bar{\phi}(z) d\nu(z).$$

On a donc

$$\begin{aligned} \frac{1}{n} \sum_{k=0}^{n-1} |\langle U^k f, f\rangle|^2 &= \frac{1}{n} \sum_{k=0}^{n-1} \Big| \int_{\mathbf{S}^1} z^k \phi(z)\bar{\phi}(z) d\nu(z)\Big|^2 \\ &= \frac{1}{n} \sum_{k=0}^{n-1} \Big| \int_{\mathbf{S}^1} z^k d\sigma(z)\Big|^2 \\ &= \frac{1}{n} \sum_{k=0}^{n-1} |\hat{\sigma}(k)|^2 \end{aligned}$$

où μ est la mesure de probabilité $d\sigma(z) = |\phi(z)|^2 d\nu(z)$. Or, on a le théorème suivant :

Théorème 4.3.2 (Wiener) *Si σ est une mesure de probabilité sur $\mathbf{S}^1$*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} |\hat{\sigma}(k)|^2 = \sum_{a \text{ atome de } \sigma} |\sigma(\{a\})|^2$$

Démonstration. — En effet

$$\begin{aligned} \frac{1}{n} \sum_{k=0}^{n-1} |\hat{\mu}(k)|^2 &= \frac{1}{n} \sum_{k=0}^{n-1} \Big(\int_{\mathbf{S}^1} z^k d\sigma(z)\Big)\Big(\int_{\mathbf{S}^1} w^{-k} d\sigma(w)\Big) \\ &= \int_{\mathbf{S}^1\times\mathbf{S}^1} \Big(\frac{1}{n} \sum_{k=0}^{n-1} (zw^{-1})^k\Big) d(\sigma\otimes\sigma)(z,w). \end{aligned}$$

Or,

$$\left|\frac{1}{n}\sum_{k=0}^{n-1}(zw^{-1})^k\right| \leq 1 \in L^2$$

converge simplement vers $\mathbf{1}_\Delta(z,w)$ où Δ est la diagonale $z = w$ dans $\mathbf{S}^1 \times \mathbf{S}^1$. Le théorème de convergence dominée montre donc que

$$\begin{aligned}\lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}|\hat{\mu}(k)|^2 &= \int_{\mathbf{S}^1\times\mathbf{S}^1}\mathbf{1}_\Delta(z,w)d(\sigma\otimes\sigma)(z,w)\\ &= \int_{\mathbf{S}^1}\sigma(\{z\})d\sigma(z)\\ &= \sum_{a \text{ atome de } \sigma}|\sigma(\{a\})|^2\end{aligned}$$

□

Utilisant ce théorème de Wiener on voit que

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}|\langle U^k f, f\rangle|^2 = \sum_{a \text{ atome de } \nu}|\phi(a)|^2|\nu(\{a\})|^2.$$

Mais, si $\nu(\{a\}) > 0$ cela signifie que a est valeur propre de U_T restreint à $C(f)$ (**Pourquoi ?**) ; comme T est f.m $a = 1$ et la fonction propre associée est une constante non nulle. Ceci est impossible car toute fonction de $C(f)$ est de μ moyenne nulle.

L'implication réciproque est facile : si $Uf = \lambda f$ alors λ est de module 1 (car U est une isométrie) et $U^k f = \lambda^k f$. Par conséquent,

$$\left(\int_X f d\mu\right)^2 = \lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}|\langle U^k f, f\rangle|^2 = \lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}\lambda^{2k} = 0.$$

b) Les autres implications découleront facilement du lemme important suivant.

Lemme 4.3.1 *Soit (a_k) une suite de réels positifs. On a*

$$(a) \qquad \lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}a_k = 0$$

si et seulement si : (b) la suite a_k tend vers 0 le long d'un ensemble de densité 1 : il existe $Z \subset \mathbf{N}$ de densité 1 tel que

$$\lim_{\substack{n\to\infty\\ n\in Z}} a_n = 0.$$

Démonstration. — Le fait que (b) implique (a) est trivial et laissé au lecteur. Montrons donc que (a) entraîne (b). Pour tout $p \geq 1$ l'ensemble

$$Z_p = \{k : a_k \geq \frac{1}{p}\}$$

est de densité 0 (son complémentaire est de densité 1). Par conséquent il existe n_p tel que pour tout $n \geq n_p$

$$\frac{1}{n} \sum_{k=0}^{n_p - 1} \mathbf{1}_{Z_p}(k) \leq \frac{1}{p}.$$

Posons alors

$$Z = \bigcup_{p \geq 1} (Z_p \cap [n_p, \infty[).$$

Le long de Z^c la suite a_n tend vers 0 ; en effet, si $n \notin Z$ alors pour $n \geq n_p$ on a $a_n \leq (1/p)$. En outre, Z est de densité 0 : puisque les ensembles Z_p croissent avec p, pour $n \geq 1$, $Z \cap [0, n] \subset Z_p$ où $n_p \leq n < n_{p+1}$. Donc,

$$\frac{1}{n} \sum_{k=0}^{n-1} \mathbf{1}_Z(k) \leq \frac{1}{n} \sum_{k=0}^{n-1} \mathbf{1}_{Z_p}(k) \leq \frac{1}{p}.$$

□

Le fait que v) implique iii) se fait en utilisant le lemme précédent et en faisant dans v) $f = \mathbf{1}_A - \mu(A)$, $g = \mathbf{1}_B - \mu(B)$. L'équivalence de iii) et ii) est encore due au lemme précédent. Montrons que iii) entraîne v) : Le lemme et iii) impliquent que v) a lieu pour f et g de la forme $f = \mathbf{1}_A - \mu(A)$, $g = \mathbf{1}_B - \mu(B)$ et donc v) a lieu pour f et g fonctions indicatrices puis fonctions simples puis L^2. L'équivalence de iv) avec le reste est facile est laissée au lecteur.

□

Exercice : Démontrer que T est faiblement mélangeante si pour tous $A, B, C \in \mathcal{B}$

$$\frac{1}{n} \sum_{k=0}^{n-1} \mu(A \cap T^{-k} B \cap T^{-2k} C) = \mu(A)\mu(B)\mu(C).$$

Le théorème précédent a pour conséquence le suivant :

Théorème 4.3.3 *Les assertions suivantes sont équivalentes*
i) $(X \times X, \mathcal{B} \otimes \mathcal{B}, \mu \otimes \mu, T \times T)$ *est ergodique*

ii) $(X, \mathcal{B}, \mu, T)$ *est faiblement mélangeant*
iii) $(X \times X, \mathcal{B} \otimes \mathcal{B}, \mu \otimes \mu, T \times T)$ *est faiblement mélangeant*
iv) pour tout $(Y, \mathcal{C}, \nu, S)$ *le système dynamique* $(X \times Y, \mathcal{B} \otimes \mathcal{C}, \mu \otimes \nu, T \times S)$ *est ergodique.*

Démonstration. —
Montrons que i) implique ii) : Si ϕ est une fonction propre de T de v.p λ, $T\phi = \lambda\phi$ on a $T\bar{\phi} = \bar{\ } bar\phi$ si bien que si on pose $\psi = \phi \otimes \bar{\phi}$ ($\psi(x, y) = \phi(x)\bar{\psi}(y)$) on a $(T \times T)\psi = \bar{\lambda}\lambda\psi$. Or, $|\lambda| = 1$ (U_T est une isométrie) si bien que ψ est constante ($T \times T$ est ergodique) et par conséquent ϕ est constante μ-pp (**Pourquoi ?**).
ii) implique iii) : soient $A, B, C, D \in \mathcal{B}$ et Z un ensemble de densité 1 sur lequel $\mu(T^{-k}A \cap C) \to \mu(A)\mu(C)$ et $\mu(T^{-k}B \cap D) \to \mu(B)\mu(D)$. On a par définition de la mesure produit

$$(\mu \otimes \mu)((T^{-k}A \times T^{-k}B) \cap (C \times D)) = \mu(T^{-k}A \cap C)\mu(T^{-k}B \cap D).$$

et quand $k \in Z$ cette dernière quantité tend vers $\mu(A)\mu(C)\mu(B)\mu(D)$ qui est $(\mu \otimes \mu)(A \times B) \cdot (\mu \otimes \mu)(C \times D))$.
iii) implique i) est immédiat.
L'équivalence de iv) avec le reste se fait comme "ii) implique iii)". □

Remarque : Quand on étudie des actions de groupes autres que **Z** on définit le faible mélange comme étant l'absence de sous-espaces vectoriels de dimension finie invariants en dehors de l'ensemble des constantes.

4.4 Facteur de Kronecker

Supposons que $(X, \mathcal{B}, \mu, T)$ soit ergodique et notons F l'adhérence dans $L^2(X, \mu)$ de l'espace vectoriel engendré par les *fonctions propres* de U_T. Notons par ailleurs $\mathcal{K}$ la tribu engendrée par les fonctions propres de U (c'est-à-dire engendrée par les $\{f > \lambda\}$, $\lambda \in \mathbb{R}$, f fonction propre de U). Il est clair que $F \subset L^2(X, \mathcal{K})$. Démontrons l'inclusion réciproque. Remarquons que si f est une fonction propre de U de valeur propre associée λ alors f^n est fonction propre de valeur propre λ^n. Les valeurs propres de U sont simples, si bien que les fonctions f^n, $n \in \mathbb{Z}$ sont deux à deux orthogonales dans $L^2(X, \mu)$. Si $P(z) = \sum_{|k| \leq N} a_k z^k$ est un polynôme en z, z^{-1} on a donc $\|P(f)\|_{L^2(\mu)} = \sum_{|k| \leq N} |a_k|^2$ et donc

$$\|P(f)\|^2_{L^2(X,\mu)} = \int_{\mathbb{T}} |P(z)|^2 dz.$$

Si A est un borélien de $\mathbb{T}$ il existe une suite de tels polynômes trigonométriques P_n qui converge vers $\mathbf{1}_A$ dans $L^2(\mathbb{T}, dz)$, si bien que $\mathbf{1}_A \circ f$ est limite dans $L^2(X, \mu)$ de $P_n(f)$. Ceci démontre que toute fonction de $L^2(X, \mu)$ de la forme $\mathbf{1}_A \circ f$ est dans F. Mais il est clair que $L^2(\mathcal{K}, \mu)$ est engendré par les $\mathbf{1}_A \circ f$ quand A décrit les boréliens de $\mathbb{T}$ et f les fonctions propres de U.

Notons à présent Λ le groupe abélien constitué des valeurs propres de U et Γ son groupe dual, c'est-à-dire le groupe des caractères de Λ.

4.5 Couplages

Etant donnés deux systèmes dynamiques $(X_1, \mathcal{B}_1, \mu_1, T_1)$ et $(X_2, \mathcal{B}_2, \mu_2, T_2)$, on appelle couplage, toute mesure ν définie sur la tribu $\mathcal{B} = \mathcal{B}_1 \otimes B_2$, invariante par $T_1 \times T_2 : X_1 \times X_2 \to X_1 \times X_2$ et qui se projette sur le premier facteur sur μ_1 et sur le deuxième facteur sur μ_2.

Exemple a) La mesure produit $\nu = \mu_1 \otimes \mu_2$ définie par $\nu(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2)$ est toujours un couplage.

b) Si $(X_i, \mathcal{B}_i, \mu_i, T_i) = (X, \mathcal{B}, \mu, T)$, $i = 1, 2$, la mesure diagonale $\Delta(A_1 \times A_2) = \mu(A_1 \cap A_2)$ est un couplage.

Définition 4.5.1 *On dit que les systèmes dynamiques sont disjoints* $(X_1, \mathbf{B_1}, \boldsymbol{\mu_1}, \mathbf{T_1})$ *et* $(X_2, \mathbf{B_2}, \boldsymbol{\mu_2}, \mathbf{T_2})$ *si le seul couplage qu'ils admettent est la mesure produit.*

On a alors la propriété suivante simple mais très utile :

Proposition 4.5.1 *L'identité est toujours disjointe des ergodiques : avec les notations précédentes, si* $T_1 = Id$ *et si* $(X_2, \mathbf{B_2}, \boldsymbol{\mu_2}, \mathbf{T_2})$ *est ergodique, alors le seul couplage possible est la mesure produit.*

Démonstration. — Soit ν est une mesure invariante par $I \times T_2$; pour $f \in C(X_1)$, $g \in C(X_2)$ la moyenne $\frac{1}{n}\sum_{k=0}^{n-1} f(x_1)g(T_2^k x_2)$ converge pour tout $x_1 \in X_1$ et μ_2-p.t $x_2 \in X_2$ vers $f(x_1)\int_{X_2} g(y)d\mu_2(y)$. En particulier, puisque $(\pi_2)_*\nu = \mu_2$ la convergence a lieu ν-p.p. et d'après le théorème de convergence dominée

$$\int_{X_1 \times X_2} f(x_1)g(x_2)d\nu(x_1, x_2) = \int_{X_1 \times X_2} f(x_1)d\nu(x_1, x_2) \int_{X_2} g(x_2)d\mu_2(x_2).$$

Comme $(\pi_1)_*\nu = \mu_1$ on a donc

$$\int_{X_1 \times X_2} f(x_1)g(x_2)d\nu(x_1, x_2) = \int_{X_1} f(x_1)d\mu_1(x_1) \int_{X_2} g(x_2)d\mu_2(x_2),$$

ce qui prouve que ν est la mesure produit. $\square$

Exercice On suppose que $(X, \mathbf{B}, \mu, \mathbf{T})$ est faiblement mélangeant.
1) On note $S = I \times T \times T^2$ et Δ la mesure diagonale sur $(X^3, \mathcal{B}^{3\otimes})$ définie par $\Delta(A \times B \times C) = \mu(A \cap B \cap C)$. Soit λ une limite pour la topologie faible-$*$ de la suite $(1/N)\sum_{k=0}^{N}(S_*)^k\Delta$. Démontrer que λ est un couplage de (I, μ) et $(T \times T^2, \mu \otimes \mu)$.
2) En déduire que pour tous $A, B, C \in \mathcal{A}$

$$\lim_{N\to\infty} \frac{1}{N}\sum_{k=0}^{N} \mu(A \cap T^{-k}B \cap T^{-2k}C) = \mu(A)\mu(B)\mu(C).$$

[Comme (T^2, μ) est ergodique ((T, μ) est f.m) et que (T, μ) est f.m., on voit que $(T \times T^2, \mu \otimes \mu)$ est ergodique]
3). Démontrer que pour tous $A, B, C \in \mathcal{A}$ il existe $\mathcal{N} \subset \mathbf{N}$ de densité 1 tel que

$$\lim_{k\to\infty, N\in\mathcal{N}} \mu(A \cap T^{-k}B \cap T^{-2k}C) = \mu(A)\mu(B)\mu(C).$$

[Appliquer le résultat du 2. à $(T \times T, \mu \otimes \mu)$, $A \times A$, $B \times B$, $C \times C$ pour obtenir

$$\lim_{N\to\infty} \frac{1}{N}\sum_{k=0}^{N}\Big(\mu(A \cap T^{-k}B \cap T^{-2k}C) - \mu(A)\mu(B)\mu(C)\Big)^2.$$

]

4.6 Mélange faible d'ordre supérieur

On dit qu'une transformation est *p-mélangeante* si pour tous $A_1, \ldots, A_p \in \mathcal{B}$

$$\lim_{n\to\infty} \mu(A_1 \cap T^{-n}A_2 \cdots \cap T^{-n(r-1)}A_p) = \mu(A_1)\mu(A_2)\cdots\mu(A_p).$$

Une question toujours non résolue qui est l'objet de recherches actives est la suivante : *le* 2*-mélange implique-t-il le* 3*-mélange ?* Il est remarquable que l'analogue *faible* de cette question admette une réponse positive dont la démonstration est non triviale. On dit qu'une transformation est *faiblement mélangeante d'ordre* p si pour tous $A_1, \ldots, A_p \in \mathcal{B}$ il existe un ensemble $Z \subset \mathbf{N}$ de densité 1 tel que

$$\lim_{\substack{n\to\infty \\ n\in Z}} \mu(A_1 \cap T^{-n}A_2 \cdots \cap T^{-n(r-1)}A_p) = \mu(A_1)\mu(A_2)\cdots\mu(A_p).$$

Il est équivalent de dire que pour tous $A_1, \ldots, A_p \in \mathcal{B}$

$$\lim_{n\to\infty} \sum_{k=0}^{n-1} |\mu(A_1 \cap T^{-k} A_2 \cdots \cap T^{-k(r-1)} A_p) - \mu(A_1)\mu(A_2)\cdots\mu(A_p)| = 0.$$

On a

Théorème 4.6.1 *Si $(X, \mathcal{B}, \mu, T)$ est faiblement mélangeant alors pour tout $p \geq 3$ il est faiblement mélangeant d'ordre p.*

Il est facile d'adapter la preuve de l'exercice de la section précédente pour démontrer ce résultat. On peut également en donner une preuve en utilisant la méthode de Van der Corput.

4.7 Argument de Hopf et Théorie spectrale

Soit T un automorphisme ergodique du tore $\mathbb{T}^2$. Nous illustrons dans ce qui suit une méthode (due à Martine Babillot) qui permet de démontrer qu'un tel automorphisme est mélangeant. Soit f une fonction continue sur $\mathbb{T}^2$ de moyenne nulle. Il existe une suite n_k d'entiers pour laquelle $f \circ T^{n_k}$ converge faiblement dans $L^2(\mathbb{T}^2)$ vers une fonction de carré intégrable g et nous devons démontrer que g est nulle. Supposons par l'absurde que g ne soit pas nulle. D'après le Théorème spectral, il existe une mesure ν sur le cercle et un isomorphisme unitaire V de $\overline{\{f \circ T^k : k \in \mathbb{Z}\}}^{L^2(\mathbb{T}^2)}$ sur $L^2(\nu)$ tel que $V(l \circ T) = zV(l)$. Notons $\phi = V(f)$, $\psi = V(g)$ et A le sous-espace de $L^2(\nu)$ constitué des limites faibles de la suite de fonctions z^n, $n \to \pm\infty$. Il est facile de voir que A est stable par conjugaison complexe et par produit. Posons alors $\theta = |\psi|^2$ qui est non identiquement nulle et dans A et posons $h = V^{-1}(\theta)$. Par définition il existe une suite $n_k \to \infty$ pour laquelle $f \circ T^{n_k}$ converge faiblement vers h et comme h est invariante par conjugaison il existe une suite $n'_k \to \infty$ telle que $f \circ T^{-n'_k}$ converge faiblement vers h. D'après le Théorème de Banach-Sachs on sait que l'on peut extraire une sous-suite n_{i_k} de n_k telle que les moyennes de Cesaro $M_K := (1/K) \sum_{i=1}^{K} f \circ T_{n_{i_k}}$ converge dans $L^2(\mathbb{T}^2)$ vers h. Il existe donc une sous-suite K_p pour laquelle M_{K_p} converge p.p vers h. Par un argument classique on voit que h est constante sur presque toute feuille stable de T et on voit de la même manière que h est constante sur presque toute feuille instable de T. On en déduit que h est constante et comme h n'est pas identiquement nulle cette constante est également non-nulle. Mais cela est une contradiction car comme f est de moyenne nulle, toute limite faible de la suite $f \circ T^n$ doit être de moyenne nulle.

Cet argument présent l'intérêt de donner une démonstration simple du mélange du flot géodésique.

Chapitre 5

Entropie

La notion d'entropie métrique (resp. topologique) en théorie ergodique est un invariant très utile de conjugaison mesurable (resp. topologique).

5.1 Entropie métrique

Dans tout ce qui suit $(X, \mathcal{B}, \mu, T)$ est un système dynamique.

5.1.1 Entropie d'une partition finie

Si ξ est une partition mesurable finie de X c'est-à-dire si ξ est une partition finie de X en ensembles $C_i \in \mathcal{B}$, $1 \leq i \leq r$ $(r = \#\xi)$ de μ-mesure non-nulle[1] on définit *l'entropie de la partition* ξ par rapport à μ par

$$H_\mu(\xi) = -\sum_{i=1}^{r} \mu(C_i) \log(\mu(C_i)).$$

Dans la suite nous omettrons souvent l'indice μ.

Il est commode d'introduire la *fonction d'information* définie pour tout $x \in X$ par

$$I(\xi)(x) = -\sum_{i=1}^{r} \log(\mu(C_i))\mathbf{1}_{C_i}(x).$$

On a

$$H_\mu(\xi) = \int_X I(\xi)(x) d\mu(x).$$

Si ξ et η sont deux partitions mesurables finies on introduit la partition $\xi \vee \eta$ qui est la plus petite partition raffinant ξ et η (contenant les atomes de

1. Nous dirons que les C_i sont les atomes de la partition ξ

ξ et de η) : c'est la partition dont les atomes sont les $C_i \cap D_j$, $C_i \in \xi$, $D_j \in \eta$. Essayons de calculer l'entropie de la partition $\xi \vee \eta$ en fonction de celles de ξ et de η :

$$\begin{aligned}
H(\xi \vee \eta) &= -\sum_{C_i \in \xi, D_j \in \eta} \mu(C_i \cap D_j) \log(\mu(C_i \cap D_j)) \\
&= -\sum_{C_i \in \xi, D_j \in \eta} \mu(C_i \cap D_j) \log(\mu(C_i|D_j)\mu(D_j)) \\
&= -\sum_{C_i \in \xi, D_j \in \eta} \mu(C_i \cap D_j) \log(\mu(D_j)) - \sum_{C_i \in \xi, D_j \in \eta} \mu(C_i \cap D_j) \log(\mu(C_i|D_j)) \\
&= H(\eta) - \sum_{C_i \in \xi, D_j \in \eta} \mu(C_i \cap D_j) \log(\mu(C_i|D_j))
\end{aligned}$$

où on a noté $\mu(C_i|D_j) = \mu(C_i \cap D_j)/\mu(D_j)$. Si on introduit *l'entropie conditionnelle* $H(\xi|\eta)$ de ξ par rapport à η

$$H(\xi|\eta) = -\sum_{C_i \in \xi, D_j \in \eta} \mu(C_i \cap D_j) \log(\mu(C_i|D_j))$$

on obtient la formule très importante

$$H_\mu(\xi \vee \eta) = H_\mu(\xi|\eta) + H_\mu(\eta).$$

Il est utile parfois d'écrire

$$H_\mu(\xi|\eta) = \sum_{D_j \in \eta} \mu(D_j) H_{\mu(\cdot|D_j)}(\xi)$$

où $\mu(\cdot|D_j)$ est la mesure conditionnelle par rapport à D_j (i.e $\mu(\cdot|D_j) = \mu(\cdot \cap D_j)/\mu(D_j)$).

A ce stade il est pertinent d'introduire la *fonction d'information conditionnelle*. Supposons que ξ est une partition mesurable finie et que $\mathcal{A}$ est une sous-tribu de $\mathcal{B}$; dans la situation précédente, $\mathcal{A}$ sera la tribu $\hat{\eta}$ engendrée par les atomes de la partition η. Introduisons l'espérance conditionnelle $\mathbf{E}(\cdot|\mathcal{A})$ par rapport à la tribu $\mathcal{A}$; dans le cas où $\mathcal{A} = \hat{\eta}$ on a pour toute fonction $f \in L^1(X, \mathcal{B}, \mu)$

$$\mathbf{E}(f|\mathcal{A})(\cdot) = \sum_{D_j \in \eta} \left(\frac{1}{\mu(D_j)} \int_{D_j} f(x) d\mu(x) \right) \mathbf{1}_{D_j}(\cdot).$$

Nous définissons alors

$$I(\xi|\mathcal{A})(\cdot) = -\sum_{C_i \in \xi} \log\Big(\mathbf{E}(\mathbf{1}_{C_i}|\mathcal{A})(\cdot) \Big) \mathbf{1}_{C_i}(\cdot).$$

On définit alors *l'entropie conditionelle de ξ par rapport à la tribu $\mathcal{A}$*

$$H(\xi|\mathcal{A}) = \int_X I(\xi|\mathcal{A})(x)d\mu(x).$$

Remarquons que puisque $\int_X \mathbf{E}(f|\mathcal{A})d\mu = \int_X f d\mu$ on a

$$H(\xi|\mathcal{A}) = -\int_X \Big(\sum_{C_i\in\xi} \log\Big(\mathbf{E}(\mathbf{1}_{C_i}|\mathcal{A})(\cdot)\Big)\mathbf{E}(\mathbf{1}_{C_i}|\mathcal{A})(\cdot)\Big)d\mu$$

ou encore en notant $\phi(t) = -t\log t$

$$H(\xi|\mathcal{A}) = \int_X \sum_{C_i\in\xi} \phi\Big(\mathbf{E}(\mathbf{1}_{C_i}|\mathcal{A})(\cdot)\Big)d\mu \tag{5.1}$$

Dans la cas où $\mathcal{A} = \hat{\eta}$ on retrouve le résultat précédent.

Remarque : Si $\mathcal{A} = \{\emptyset, X\}$ est la tribu triviale $I(\xi|\mathcal{A})(\cdot) = I(\xi)(\cdot)$ tandis que si $\mathcal{A} = \mathcal{B}$ on a $I(\xi|\mathcal{A})(\cdot) = 0$ μ-pp. En effet, si $\mathcal{A}$ est la tribu triviale le résultat est clair tandis que si $\mathcal{A} = \mathcal{B}$ on a $\mathbf{E}(\mathbf{1}_{C_i}|\mathcal{A})(\cdot) = \mathbf{1}_{C_i}(\cdot)$ et donc $I(\xi|\mathcal{A})(x) = \sum_{C_i\in\xi}\phi(\mathbf{1}_{C_i}(x))$ où $\phi(t) = -t\log t$ (observer que $\phi(0) = \phi(1) = 0$).

Nous regroupons dans la proposition qui suit quelques propriétés utiles de l'entropie :

Proposition 5.1.1 *Si ξ et η sont des partitions mesurables finies*
1) $H(\xi\vee\eta) = H(\eta) + H(\xi|\eta) = H(\xi) + H(\eta|\xi)$ [2]
2) $H(\xi\vee\eta) \leq H(\xi) + H(\eta)$.
3) Si $T : X \to X$ est $\mathcal{B}$-mesurable et préserve μ

$$H(T^{-1}\xi|T^{-1}\eta) = H(\xi|\eta).$$

4) Si $\xi_1 < \xi_2$ (ξ_2 est plus fine que ξ_1 [3]*) alors $H(\xi_1|\mathcal{A}) \leq H(\xi_2|\mathcal{A})$.*
5) Si $\mathcal{A}_1 \subset \mathcal{A}_2$ sont des tribus, $H(\xi|\mathcal{A}_2) \leq H(\xi|\mathcal{A}_1)$ [4]*. (En particulier $H(\xi|\eta) \leq H(\xi)$.)*
6) On a toujours $H(\xi) \leq \log(\#\xi)$.

2. et plus généralement $H(\xi\vee\eta|\mathcal{A}) = H(\eta|\mathcal{A}) + H(\xi|\hat{\eta}\vee\mathcal{A}) = H(\xi|\mathcal{A}) + H(\eta|\hat{\xi}\vee\mathcal{A})$: c'est facile à démontrer si $\mathcal{A}$ a un nombre fini d'atomes, plus délicat sinon.

3. i.e. tout atome de ξ_2 est inclus dans un atome de ξ_1

4. $H(\xi|\mathcal{B}) = 0$, $H(\xi|\{\emptyset, X\}) = H(\xi)$

Démonstration. — Le point 1) a déjà été démontré. Le point 3) est évident. Le 2) est un cas particulier de 5) qui repose sur l'inégalité de Jensen [5] et le fait que la fonction $t \mapsto -t \log t$ est concave : puisque $\mathcal{A}_1 \subset \mathcal{A}_2$ l'opérateur $\mathbf{E}(\cdot|\mathcal{A}_1)$ restreint à $L^1(\mathcal{A}_2)$ est l'identité si bien que d'après l'inégalité de Jensen

$$\mathbf{E}\Big(\phi\Big(\mathbf{E}(\mathbf{1}_{\mathbf{C_i}}|\mathcal{A}_\mathbf{2})(\cdot)\Big)|\mathcal{A}_\mathbf{1}\Big) \leq \phi\Big(\mathbf{E}(\mathbf{1}_{\mathbf{C_i}}|\mathcal{A}_\mathbf{1})(\cdot)\Big);$$

comme

$$\int_X \phi\Big(\mathbf{E}(\mathbf{1}_{C_i}|\mathcal{A}_2)(\cdot)\Big)d\mu = \int_X \mathbf{E}\Big(\phi\Big(\mathbf{E}(\mathbf{1}_{\mathbf{C_i}}|\mathcal{A}_\mathbf{2})(\cdot)\Big)|\mathcal{A}_\mathbf{1}\Big)\mathbf{d}\mu$$

l'égalité (5.1) permet de conclure la preuve de 5).

Le point 4) (et son analogue conditionnel) se démontre en remarquant que puisque $\xi_1 < \xi_2$ on a $\xi_1 \vee \xi_2 = \xi_2$ et donc,

$$H(\xi_2) = H(\xi_1) + H(\xi_2|\xi_1) \geq H(\xi_1).$$

Le point 6) est facile (utiliser la concavité de ϕ). □

Exercice Démontrer que
$I(T^{-1}\xi) = I(\xi) \circ T$ et
$I(\xi \vee \eta) = I(\xi|\eta) + I(\eta)$.

5.1.2 Entropie d'une transformation

Définition

Nous sommes en mesure de définir l'entropie d'une transformation (T, μ).

Théorème 5.1.1 *Si $(X, \mathcal{B}, \mu, T)$ est un système dynamique et si ξ est une partition mesurable finie de X la suite $H(\bigvee_{k=0}^{n-1} T^{-k}\xi)$ est sous-additive et on note*

$$h_\mu(T, \xi) = \lim_{n\to\infty} \frac{1}{n} H(\bigvee_{k=0}^{n-1} T^{-k}\xi).$$

Démonstration. — Notons $\xi_n = \bigvee_{k=0}^{n-1} T^{-k}\xi$. Puisque $\xi_{n+m} = \xi_n \vee \xi_m$ on a d'après le 2 de la proposition 5.1.1

$$H(\xi_{n+m}) \leq H(\xi_n) + H(\xi_m).$$

□

5. c'est-à-dire pour ϕ fonction concave, $f \in L^1(X, \mathcal{B}, \mu)$ et $\mathcal{A}$ sous-tribu de $\mathcal{B}$ on a $\phi\Big(\mathbf{E}(f|\mathcal{A})\Big) \geq \mathbf{E}(\phi \circ f|\mathcal{A})$; cela se démontre facilement d'abord dans le cas où f est étagée

Définition 5.1.1 *On définit l'entropie de* (T, μ) *comme étant*

$$h_\mu(T) = \sup_\xi h(T, \xi)$$

le sup étant pris sur toutes les partitions mesurables finies d'entropie finie.

Exemples : Calculons l'entropie d'une translation rationnelle sur $\mathbf{R}/\mathbf{Z}$ d'angle p/q. Pour toute partition mesurable finie ξ d'entropie finie, le nombre d'atomes de la partition $\bigvee_{k=0}^{n-1} T^{-k}\xi$ est inférieur à $(\#\xi)^q$ et est donc borné. Par conséquent $h(T, \xi) = 0$ pour toute partition finie et donc $h(T) = 0$.

Notons le théorème suivant

Théorème 5.1.2 *Pour toute partition mesurable finie* ξ

$$h(T, \xi) = \lim_{n\to\infty} H(\xi | \bigvee_{k=1}^{n-1} T^{-k}\xi)$$

ou encore

$$h(T, \xi) = H(\xi | \bigvee_{k=1}^{\infty} T^{-k}\xi).$$

Démonstration. — Pour la première partie, il suffit d'écrire

$$\begin{aligned}
H(\bigvee_{k=0}^{n-1} T^{-k}\xi) &= H(\xi \vee \bigvee_{k=1}^{n-1} T^{-k}\xi) \\
&= H(\xi | \bigvee_{k=1}^{n-1} T^{-k}\xi) + H(\bigvee_{k=1}^{n-1} T^{-k}\xi) \\
&= H(\xi | \bigvee_{k=1}^{n-1} T^{-k}\xi) + H(T^{-1} \bigvee_{k=0}^{n-2} T^{-k}\xi) \\
&= H(\xi | \bigvee_{k=1}^{n-1} T^{-k}\xi) + H(\bigvee_{k=0}^{n-2} T^{-k}\xi)
\end{aligned}$$

où on a utilisé le 3) de la proposition 5.1.1. Si on itère la relation précédente on trouve que

$$H(\xi | \bigvee_{k=0}^{n-1} T^{-k}\xi) = H(\xi | T^{-1}\xi) + \cdots + H(\xi | T^{-1}\xi \vee T^{-2}\xi) + \cdots + H(\xi | \bigvee_{k=1}^{n-1} T^{-k}\xi).$$

La suite $H(\xi|\bigvee_{k=1}^{n-1} T^{-k}\xi)$ est décroissante par rapport à k (cf. 5) de la proposition 5.1.1) et par conséquent converge ; le théorème de Césaro[6] permet de conclure.

La deuxième partie du théorème se démontre en utilisant le lemme suivant :

Lemme 5.1.1 *Si les tribus $\mathcal{A}_n$ v'érifient $\mathcal{A}_n \uparrow \mathcal{A}$, alors*

$$\lim_{n\to\infty} H(\xi|\mathcal{A}_n) = H(\xi|\mathcal{A}).$$

Démonstration. — D'après le théorème de convergence des martingales[7] $\mathbf{E}(\mathbf{1}_{C_i}|\mathcal{A}_n)(\cdot)$ converge μ-p.p vers $\mathbf{E}(\mathbf{1}_{C_i}|\mathcal{A})(\cdot)$ et donc $\phi\Big(\mathbf{E}(\mathbf{1}_{C_i}|\mathcal{A}_n)(\cdot)\Big)$ converge μ-p.p vers $\phi\Big(\mathbf{E}(\mathbf{1}_{C_i}|\mathcal{A}_n)(\cdot)\Big)$. Comme la fonction $\phi = -t\log t$ est continue (donc bornée) sur $[0,1]$ les hypothèses du théorème de convergence dominée sont vérifiées et on a

$$\lim_{n\to\infty}\int_X \phi\Big(\mathbf{E}(\mathbf{1}_{C_i}|\mathcal{A}_n)(\cdot)\Big)d\mu = \int_X \phi\Big(\mathbf{E}(\mathbf{1}_{C_i}|\mathcal{A})(\cdot)\Big)d\mu.$$

Puisque ξ est *finie* et que

$$H(\xi|\mathcal{A})(\cdot) = \int_X \sum_{C_i\in\xi} \phi\Big(\mathbf{E}(\mathbf{1}_{C_i}|\mathcal{A})(\cdot)\Big)d\mu$$

on a convergence μ-p.p de $H(\xi|\mathcal{A}_n)(\cdot)$ vers $H(\xi|\mathcal{A})(\cdot)$. L □

□

Distance de Rokhlin

Théorème 5.1.3 *Si ξ et η sont deux partitions mesurables finies on a toujours*

$$|h_\mu(T,\xi) - h_\mu(T,\eta)| \leq H(\xi|\eta) + H(\eta|\xi).$$

6. si a_n converge alors la moyenne $(a_1 + \cdots + a_n)/n$ converge vers la même limite
7. Si f est une fonction $L^1(\mathcal{B},\mu)$ $\mathbf{E}(f|\mathcal{A}_n)$ converge L^1 et μ-p.p vers $\mathbf{E}(f|\mathcal{A})$.

Démonstration. — En effet

$$\begin{aligned}
H(\bigvee_{k=0}^{n-1} T^{-k}\xi) &\leq H(\bigvee_{k=0}^{n-1} T^{-k}\xi \vee \bigvee_{k=0}^{n-1} T^{-k}\eta) \\
&\leq H(\bigvee_{k=0}^{n-1} T^{-k}\eta) + H(\bigvee_{k=0}^{n-1} T^{-k}\xi | \bigvee_{k=0}^{n-1} T^{-k}\eta) \\
&\leq H(\bigvee_{k=0}^{n-1} T^{-k}\eta) + \sum_{l=0}^{n-1} H(T^{-l}\xi | \bigvee_{k=0}^{n-1} T^{-k}\eta) \\
&\leq H(\bigvee_{k=0}^{n-1} T^{-k}\eta) + \sum_{l=0}^{n-1} H(T^{-l}\xi | \mathbb{T}^{-l}\eta) \\
&\leq H(\bigvee_{k=0}^{n-1} T^{-k}\eta) + \sum_{l=0}^{n-1} H(\xi | \eta) \\
&\leq H(\bigvee_{k=0}^{n-1} T^{-k}\eta) + nH(\xi | \eta).
\end{aligned}$$

Diviser par n and faire tendre $n \to \infty$ donne

$$h(T,\xi) - h(T,\eta) \leq H(\xi|\eta).$$

L'inégalité inverse est claire. □

Définition 5.1.2 *La quantité $d(\xi,\eta) = H(\xi|\eta) + H(\eta|\xi)$ définit une distance appelée distance de Rokhlin. Ainsi $\xi \mapsto h(T,\xi)$ est 1-lipschitzienne.*

Mentionnons un corollaire très utile du théorème précédent :

Théorème 5.1.4 *Si ξ_n est une suite de partition mesurables finies croissante telle que $\hat{\xi}_n \uparrow \mathcal{B}$ alors*

$$\lim_{n\to\infty} h(T,\xi_n) = h(T).$$

Démonstration. — Remarquons que pour $n \leq m$

$$h(T,\xi_m) \leq h(T,\xi_n) + H(\xi_m|\xi_n);$$

mais comme $\xi_n < \xi_m$ on a $H(\xi_m|\xi_n) = 0$. Ainsi la suite $h(T,\xi_n)$ est décroissante et admet donc une limite. En outre, on a vu que $H(\eta|\xi_n)$ convergeait vers $H(\eta|\mathcal{B}) = 0$. Comme

$$h(T,\eta) \leq h(T,\xi_n) + H(\eta|\xi_n),$$

on en déduit que pour tout η

$$h(T,\eta) \leq \lim_{n\to\infty} h(T,\xi_n),$$

c'est-à-dire $h(T) \leq \lim_{n\to\infty} h(T,\xi_n)$. L'inégalité inverse est évidente. $\square$

Générateur, générateur fort

Définition 5.1.3 *Si le système dynamique (T,μ) est inversible, on dit qu'une partition est un générateur pour T (resp. générateur fort) si $\mathcal{B}$ coïncide avec la tribu engendrée par $\bigvee_{k=-\infty}^{\infty} T^{-k}\xi$ (resp. $\bigvee_{k=0}^{\infty} T^{-k}\xi$). Si T n'est pas inversible seule la définition d'un générateur fort (au sens précédent) est pertinente.*

Le théorème suivant permet de calculer facilement des entropies :

Théorème 5.1.5 *Si ξ est un générateur (resp. générateur fort) pour T*

$$h_\mu(T) = h_\mu(T,\xi).$$

Démonstration. — Ecrivons,

$$\begin{aligned}
H(\bigvee_{k=0}^{n-1} T^{-k}\eta) &\leq H(\bigvee_{k=0}^{n-1} T^{-k}\eta \vee \bigvee_{k=0}^{m+n-1} T^{-k}\xi) \\
&\leq H(\bigvee_{k=0}^{m+n-1} T^{-k}\xi) + H(\bigvee_{k=0}^{n-1} T^{-k}\eta | \bigvee_{k=0}^{m+n-1} T^{-k}\xi) \\
&\leq H(\bigvee_{k=0}^{m+n-1} T^{-k}\xi) + \sum_{l=0}^{n-1} H(T^{-l}\eta | \bigvee_{k=0}^{m+n-1} T^{-k}\xi) \\
&\leq H(\bigvee_{k=0}^{m+n-1} T^{-k}\xi) + \sum_{l=0}^{n-1} H(T^{-l}\eta | \bigvee_{k=l}^{m+l} T^{-k}\xi) \\
&\leq H(\bigvee_{k=0}^{m+n-1} T^{-k}\xi) + \sum_{l=0}^{n-1} H(\eta | \bigvee_{k=0}^{m} T^{-k}\xi).
\end{aligned}$$

Or,

$$\lim_{m\to\infty} H(\eta|\bigvee_{k=0}^{m} T^{-k}\xi) = 0$$

puisque la tribu engendrée par $\bigvee_{k=0}^{m} T^{-k}\xi$ converge en croissant vers $\mathcal{B}$. Par conséquent, si on choisit m assez grand pour que $H(\eta|\bigvee_{k=0}^{m} T^{-k}\xi) \leq \epsilon$ on a pour n assez grand

$$\frac{1}{n} H(\bigvee_{k=0}^{n-1} T^{-k}\eta) \leq \frac{n+m}{n}\frac{1}{n+m} H(\bigvee_{k=0}^{m+n-1} T^{-k}\xi) + n\epsilon.$$

Si on fait tendre n vers l'infini on obtient

$$h(T,\eta) \leq h(T,\xi) + \epsilon$$

ceci pour tout ϵ et tout η. Par conséquent $h(T)$, qui est le sup des $h(T,\eta)$, égale $h(T,\xi)$. $\square$

Remarque :

1) Si T est inversible et admet un générateur fort alors $h(T) = 0$.

2) Toute transformation d'entropie finie admet une partition génératrice finie qui a au plus $[e^{h(T)}] + 1$ éléments. (Krieger).

5.1.3 Exemples

Entropie d'une translation sur un tore

Calculons l'entropie d'une translation $x \mapsto x + \alpha$ sur $\mathbf{R}/\mathbf{Z}$ par rapport à la mesure de Haar (qui est clairement invariante). On a déjà vu que $h(T) = 0$ si α est rationnel. Supposons donc α irrationnel.

Première méthode : Soit ξ une partition finie en intervalles du cercle et notons A l'ensemble des extrémités de ces intervalles. Un instant de réflexion montre que le joint $\xi_n := \bigvee_{k=0}^{n-1} T^{-k}\xi$ est la partition en intervalles qui sont les composantes connexes du complémentaire de $\bigcup_{k=0}^{n-1} T^k A$ dans le cercle. Ainsi, ξ_n comporte au plus $n\#\xi$ atomes et donc

$$h(T,\xi) = \lim_{n\to\infty} \frac{\log(n\#\xi)}{n} = 0.$$

A présent, si $\xi^{(m)}$ est la partition en intervalles m-adiques il est clair que $\xi^{(m)} \uparrow Bor$ et le théorème 5.1.4 permet de dire que $h(T) = 0$.

Deuxième méthode : On peut raisonner de la façon suivante ; la partition $\xi = \{[0,1/2),[1/2,1)\}$ est génératrice quand α est irrationnel car pour tout m, la tribu engendrée par les intervalles m-adiques est incluse dans la tribu engendrée par les atomes du joint $\bigvee_{k=0}^{n-1} T^{-k}\xi$ pour n assez grand (le sup des diamètres des atomes tend vers 0). Il suffit donc de démontrer que $h(T,\xi) = 0$ ce qui s'effectue comme précédemment. On peut aussi procéder de la façon suivante : $h(T,\xi) = \lim_{n\to\infty} H(\xi | \bigvee_{k=1}^{\infty} T^{-k}\xi) = 0$ car le joint précédent engendre la tribu borélienne.

On peut généraliser les résultats précédents aux cas des translations sur le tore $\mathbf{T}^d$. (**Exercice**).

5.1.4 Théorème de Shannon

Théorème 5.1.6 (Shannon-Mc-Millan-Breiman) *Si T est μ ergodique et si ξ est une partition mesurable finie*

$$\lim_{n\to\infty} \frac{1}{n} I(\bigvee_{k=0}^{n-1} T^{-k}\xi)(\cdot) = h(\xi, T)$$

la convergence précédente ayant lieu μ-p.s et $L^1(\mu)$.

Démonstration. — Notons $\mathcal{B}_n = \bigvee_{k=1}^{n-1} T^{-k}\xi$ pour $n \in \mathbf{N} \cup \{\infty\}$ et $\mathcal{B}_0 = \{\emptyset, X\}$. On a d'après l'exercice de l'exercice suivant la proposition 5.1.1

$$\begin{aligned} I(\bigvee_{k=0}^{n-1} T^{-k}\xi) &= I(\xi | \bigvee_{k=1}^{n-1} T^{-k}\xi) + I(\bigvee_{k=1}^{n-1} T^{-k}\xi) \\ &= I(\xi | \bigvee_{k=1}^{n-1} T^{-k}\xi) + I(\bigvee_{k=0}^{n-2} T^{-k}\xi) \circ T \end{aligned}$$

et par conséquent

$$I(\bigvee_{k=0}^{n-1} T^{-k}\xi) = I(\xi|\mathcal{B}_0) \circ T^n + I(\xi|\mathcal{B}_1) \circ T^{n-1} + \cdots + I(\xi|\mathcal{B}_n).$$

Posons

$$g_n(\cdot) = |I(\xi|\mathcal{B}_n)(\cdot) - I(\xi|\mathcal{B}_\infty(\cdot)|, \qquad G_M(\cdot) = \sup_{n\geq M} g_n(\cdot).$$

On a vu que $G_M(\cdot) \to 0$ μ-p.p et $L^1(\mu)$ quand $M \to \infty$. Pour $M > 0$ on a

$$\begin{aligned} I_n(x) &= \sum_{k=0}^{n} (I(\xi|\mathcal{B}_k) - I(\xi|\mathcal{B}_\infty)) \circ T^{n-k}(x) + \sum_{k=0}^{n} I(\xi|\mathcal{B}_\infty) \circ T^{n-k}(x) \\ &\leq \sum_{k=0}^{M-1} g_k \circ T^{n-k}(x) + \sum_{k=M}^{n} G_M \circ T^{n-k}(x) + \sum_{k=0}^{n} I(\xi|\mathcal{B}_\infty) \circ T^{n-k}(x) \\ &\leq \sum_{k=0}^{M-1} g_k \circ T^{n-k}(x) + \sum_{l=0}^{n-M} G_M \circ T^{l}(x) + \sum_{l=0}^{n} I(\xi|\mathcal{B}_\infty) \circ T^{l}(x) \end{aligned}$$

D'après le théorème de Birkhoff (μ est ergodique) on a μ-p.s

$$\begin{aligned} \lim_{n\to\infty} \frac{1}{n} \sum_{l=0}^{n} I(\xi|\mathcal{B}_\infty) \circ T^l(x) &= \int_X I(\xi|\mathcal{B}_\infty) d\mu \\ &= h(T, \xi) \end{aligned}$$

tandis que

$$\lim_{n\to\infty}\frac{1}{n}\sum_{l=0}^{n-M} G_M\circ T^l(x) = \int_X G_M d\mu \leq \epsilon_M.$$

Enfin on a μ-p.p [8]

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{M-1} g_k\circ T^{n-k}(x) = 0.$$

Au total, pour μ-presque tout x et tout M

$$\limsup_{n\to\infty}|\frac{1}{n}I(\xi)(x) - h(T,\xi)| \leq \int_X G_M d\mu \leq \epsilon_M.$$

Comme $\epsilon_M \to 0$ on a bien la conclusion. □

Remarque : Le théorème précédent se reformule de la façon suivante ; si on note $C_{\xi,n}(x)$ l'atome de la partition $\bigvee_{k=0}^{n-1} T^{-k}\xi$ qui contient x, on a pour μ-p.t x

$$\lim_{n\to\infty}\frac{1}{n}\log\mu(C_{\xi,n}(x)) = -h(T,\xi).$$

Il est facile de voir que pour tous $\alpha, \beta > 0$ il existe un n_0 tel que pour tout $n \geq n_0$ il existe $C_n \in X$ tel que : 1) $\mu(C_n) \geq 1_\alpha$; 2) C_n admet une partition mesurable finie dont les atomes sont de μ-mesures comprises entre $e^{-n(h+\beta)}$ et $e^{-n(h-\beta)}$ et dont le nombre d'atome est dans $[e^{n(h-\beta)}, e^{n(h+\beta)}]$.

5.1.5 Entropie d'un facteur, d'un produit et d'une puissance

Théorème 5.1.7 *Si $(Y, \mathcal{B}_Y, S, \nu)$ est un facteur de $(X, \mathcal{B}_X, T, \nu)$ on a*

$$h_\nu(S) \leq h_\mu(T).$$

Démonstration. — Soit η une partition finie de Y et posons $\xi = f^{-1}\eta$ où $f : X \to Y$ est la projection définissant le facteur. Il est facile de vérifier que

$$\frac{1}{n}H_\mu(\xi\vee\ldots\vee T^{-(n-1)}\xi) = \frac{1}{n}H_\nu(\eta\vee\ldots\vee S^{-(n-1)}\eta),$$

et donc

$$h_\nu(S,\eta) = h_\mu(T,\xi) \leq h_\mu(T).$$

Cette identité étant vraie pour toute partition finie η on a la conclusion. □

8. si h est une fonction $L^1(\mu)$, alors $h\circ T^n/n$ converge $\mu - pp$ vers 0 (appliquer p.ex le théorème de Birkhoff à h et $h\circ T$ et faire la différence des sommes ergodiques).

Corollaire 5.1.1 *Si* $(X, \mathcal{B}_X, T, \mu)$ *et* $(Y, \mathcal{B}_Y, S, \nu)$ *sont isomorphes* $h_\mu(T) = h_\nu(S)$.

Théorème 5.1.8 *Si* $(X_i, \mathcal{B}_i, \mu_i, T_i)$, $i = 1, 2$ *sont deux systèmes dynamiques alors le système produit* $(X_1 \times X_2, \mathcal{B}_1 \otimes \mathcal{B}_2, \mu_1 \otimes \mu_2, T_1 \times T_2)$ *vérifie*

$$h_{\mu_1 \otimes \mu_2}(T_1 \times T_2) = h_{\mu_1}(T_1) + h_{\mu_2}(T_2).$$

Démonstration. — Soient $\xi^{(i)} = \{C_k^{(i)}\}$ une partition mesurable finie de X_i, $i = 1, 2$. Il est facile de voir que la partition $\xi := \xi^{(1)} \otimes \xi^{(2)} = \{C_k^{(1)} \times C_l^{(2)}\}$ vérifie

$$H(\bigvee_{s=0}^{n-1} T^{-s}\xi) = H(\bigvee_{s=0}^{n-1} T^{-s}\xi^{(1)}) + H(\bigvee_{s=0}^{n-1} T^{-s}\xi^{(2)}),$$

et que par conséquent

$$h(T_1 \times T_2, \xi^{(1)} \otimes \xi^{(2)}) = h(T_1, \xi^{(1)}) + h(T_2, \xi^{(2)}).$$

Si à présent on choisit deux suites de partitions $\xi_n^{(i)} \uparrow \mathcal{B}_i$ (ce qui est possible car on suppose que les espaces sont de Lebesgue) on aura aussi $\xi_n^{(1)} \otimes \xi_n^{(2)} \uparrow \mathcal{B}_1 \otimes \mathcal{B}_2$. Le théorème 5.1.4 permet de conclure. □

Théorème 5.1.9 *Pour* $m \in \mathbf{Z}$ *(*$m \in \mathbf{N}$ *si* T *n'est pas inversible) on a*

$$h_\mu(T^m) = |m| h_\mu(T).$$

Démonstration. — Supposons $m > 0$. Soit ξ une partition mesurable finie et posons $\eta_m = \xi \vee T^{-1}\xi \vee \cdots \vee T^{m-1}\xi$. On a

$$H(\bigvee_{k=0}^{nm-1} T^{-k}\xi) = H(\bigvee_{l=0}^{n-1} (T^m)^{-l}\eta_m)$$

et divisant par $n|m|$ et faisant $n \to \infty$ on obtient

$$h(T, \xi) = 1mh(T^m, \eta_m).$$

On a donc $|m|h(T, \xi) \leq h(T^m)$ et donc $|m|h(T) \leq h(T^m)$. Démontrons l'inégalité inverse. Pour toute partition ξ

$$\begin{aligned} h(T^m, \xi) &= \lim_{n\to\infty} \frac{1}{n} H(\xi \vee T^{-m}\xi \vee \cdots \vee T^{-mn}\xi) \\ &\leq \lim_{n\to\infty} \frac{m}{mn} H(\xi \vee T^{-1} \vee T^{-2}\xi \cdots \vee T^{-mn}\xi) \\ &\leq mh(T, \xi), \end{aligned}$$

ce qui termine la preuve du théorème. □

Chapitre 6

Appendice

6.1 Théorie de la mesure et intégration

Soit X un ensemble. Nous notons $\mathcal{P}(X)$ l'ensemble de tous les sous-ensembles de X. Un sous-ensemble $\mathcal{A} \subset \mathcal{P}(X)$ est une *algèbre* s'il contient l'ensemble vide, X, s'il est stable par unions finies et ($A_i \in \mathcal{A}$, $i = 1, \ldots, n$ implique $\bigcup_{i=1}^n A_i \in \mathcal{A}$) et stable par passage au complémentaire ($A \in \mathcal{A}$ implique $A^c \in \mathcal{A}$). L'ensemble $\mathcal{A}$ est appelé une *tribu* ou une *σ-algèbre* si c'est une algèbre et s'il est stable par unions dénombrables ($A_i \in \mathcal{A}$, $i \in \mathbb{N}$ implique $\bigcup_{i\in\mathbb{N}} A_i \in \mathcal{A}$) ; bien sûr il est alors stable par intersections dénombrables. Si $\mathcal{A}$ est une algèbre, une *fonction additive d'ensembles* positive (f.a.e.p. , pour faire court) est une application $\mu : \mathcal{A} \to \mathbb{R}_+$ telle que pour tous ensembles disjoints deux à deux $A_1, \ldots, A_n \in \mathcal{A}$ on a $\mu(\bigcup_{i=1}^n A_i) = \sum_{i=1}^n \mu(A_i)$. Si $\mathcal{A}$ est une σ-algèbre, une *mesure positive* μ sur $\mathcal{A}$ est une application $\mu : \mathcal{A} \to \mathbb{R}_+ \cup \{\infty\}$ telle que pour toute famille dénombrable d'ensembles disjoints deux à deux $A_i \in \mathcal{A}$, $i \in \mathbb{N}$ on ait $\mu(\bigcup_{i\in\mathbb{N}} A_i) = \sum_{i\in\mathbb{N}} \mu(A_i)$. Si μ est une faep telle que $\mu(X) = 1$, on dit qu'elle est normalisée et si μ est une mesure positive telle que $\mu(X) = 1$, on dit que c'est une *mesure de probabilité*. Dans ce cas, le triplet $(X, \mathcal{A}, \mu)$ s'appelle un espace probabilisé.

Les exemples les plus simples de σ-algèbres sont $\mathcal{P}(X)$ (la tribu totale) ou $\{\emptyset, X\}$ (la tribu triviale). Pour tout $a \in X$ on peut définir la *mesure de Dirac* en a comme étant la mesure de probabilité qui vérifie $\mu(A) = 1$ ssi $a \in A$ et vaut zéro sinon.

Il y a une façon très simple et commode de construire des σ-algèbres. Soit $\mathcal{S} \subset \mathcal{P}(X)$ une collection de sous-ensembles de X. Nous définissons $\mathcal{T}(\mathcal{S})$, la *tribu engendrée* par $\mathcal{S}$ comme la plus petite σ-algèbre pour l'inclusion contenant $\mathcal{S}$ comme sous-ensemble : c'est l'intersection non vide de toutes les σ-algèbres $\mathcal{C}$ qui contiennent $\mathcal{S}$ comme sous-ensemble. Quand X est un espace topologique, la σ-algèbre engendrée par les ensembles ouverts (ou fermés) est appelée la *tribu borélienne* et ses éléments sont les *boréliens.*

S'il est facile de construire des tribus, il est en revanche plus délicat de construire des mesures de probabilités intéressantes. Avant de voir comment l'on procède, introduisons quelques notions.

Nous disons qu'une suite décroissante (resp. croissante) $A_i \in \mathcal{A}$, $i \in \mathbb{N}$ converge vers $A \in \mathcal{A}$ et nous notons $A_i \downarrow A$ (resp. $A_i \uparrow A$) si $\bigcap_{i\in\mathbb{N}} A_i = A$ (resp. $\bigcup_{i\in\mathbb{N}} A_i = A$). Observons le fait suivant.

Proposition 6.1.1 *Soit $(X, \mathcal{A}, \mu)$ un espace probabilisé. Alors, μ est continue en tout $A \in \mathcal{A}$ dans le sens suivant : pour toute suite décroissante $A_i \downarrow A$, on a*

$$\lim \mu(A_i) = \mu(A).$$

Nous pouvons alors énoncer le théorème fondamental qui permet la construction de mesures de probabilité.

Théorème 6.1.1 (Caratheodory) *Soient $\mathcal{A}$ une algèbre et μ une fonction additive d'ensembles positive et normalisée définie sur $\mathcal{A}$. Supposons que μ soit continue en $\emptyset$. Alors, si nous notons $\hat{\mathcal{A}}$ la tribu engendrée par $\mathcal{A}$, il existe une unique mesure de probabilité $\hat{\mu}$ qui étend μ à $\hat{\mathcal{A}}$: pour tout $A \in \mathcal{A}$ on a $\hat{\mu}(A) = \mu(A)$.*

Donnons un exemple d'application de ce théorème. Soit $\Sigma = \{0,1\}^{\mathbb{N}}$ et notons $\mathcal{B}$ la σ-algèbre engendrée par les cylindres (cf. chapitre 2). On peut décrire $\mathcal{B}$ comme étant la tribu engendrée par l'algèbre $\mathcal{A} = \bigcup \mathcal{A}_n$ où $\mathcal{A}_n$ est l'algèbre des cylindres (centrés) de longueur n ; il est facile de voir que $\mathcal{B} \neq \mathcal{A}$ mais que $\mathcal{B} = \mathcal{T}(\mathcal{A})$. Une faep normalisée sur $\mathcal{A}$ permet d'associer à tout cylindre centré $C(\epsilon_0, \ldots, \epsilon_{n-1})$ de longueur n le nombre $p(\epsilon_0, \ldots, \epsilon_{n-1}) = \mu(C(\epsilon_0, \ldots, \epsilon_{n-1}))$. Observons qu'on a les *conditions de compatibilité de Kolmogorov*

$$p(\epsilon_0, \ldots, \epsilon_{n-1}) = \sum_{\epsilon\in\{0,1\}} p(\epsilon_0, \ldots, \epsilon_{n-1}, \epsilon),$$

entre cylindres centrés de taille n et ceux de taille $n+1$. Réciproquement, il n'est pas difficile de voir (**Exercice**) qu'étant donnés des nombres $p(\epsilon_0, \ldots, \epsilon_{n-1}) \in [0,1]$, $\epsilon_i \in \{0,1\}$, $n \in \mathbb{N}$ vérifiant les relations précédentes de compatibilité, il est possible de construire une unique faep normalisée μ sur $\mathcal{A}$ telle que $p(\epsilon_0, \ldots, \epsilon_{n-1}) = \mu(C(\epsilon_0, \ldots, \epsilon_{n-1}))$. Vérifions que les hypothèses du théorème de Caratheodory sont vérifiées. Soit $A_n \downarrow \emptyset$, $A_n \in \mathcal{A}$. Chaque A_n appartient à un certain $\mathcal{A}_{s_n}$, et est par conséquent une union finie de cylindres centrés de taille s_n. Puisque les cylindres finis sont des ensembles fermés dans Σ et comme Σ est compact, il en est de même des A_n. Mais une union décroissante de compacts non-vides est non-vide, si bien qu'à partir d'un certain rang A_{s_n} est vide et donc $\mu(A_{s_n}) = 0$. Le théorème de Caratheodory permet alors de construire une mesure de probabilité $\hat{\mu}$ sur $\mathcal{B}$ qui prolonge μ. Par exemple si $p(\epsilon_0, \ldots, \epsilon_{n-1}) = p^m(1-p)^m$ $(0 < p < 1)$ où m est le nombre de $0 \le i \le n-1$ pour lesquels $\epsilon_i = 1$, la mesure $\hat{\mu}$ est la mesure de Bernoulli de paramètre p.

Exercice

a) Prouver que la tribu borélienne sur $\mathbb{T}$ est la σ-algebre engendrée par : i) les arcs ; ii) les arcs dyadiques.

b) Soient $(X, \mathcal{B}, \mu)$ un espace probabilisé et $T : X \to X$ une application mesurable (cf. plus bas). Supposons que $\mathcal{B}$ soit engendrée par une algèbre $\mathcal{A}$. Prouver que μ est T-invariante si et seulement si pour tout $A \in \mathcal{A}$, $\mu(T^{-1}(A)) = \mu(A)$.

Si $(X, \mathcal{B}, \mu)$ est un espace probabilisé, il est souvent commode d'y rajouter les ensembles *μ-négilgeables* : un ensemble $X \subset X$ est μ-négligeable s'il existe $\hat{N} \in \mathcal{B}$, tel que $N \subset \hat{N}$ et $\mu(\hat{N}) = 0$. Comme une union dénombrable d'ensembles négligeables est négligeable, l'ensemble $\hat{\mathcal{B}}$ des éléments de la forme $A \cup N$, $A \in \mathcal{B}$ et N μ-négligeable, est une σ-algèbre que l'on appelle la tribu *complétée*. Noter que c'est une notion qui dépend de la mesure de probabilité μ.

6.2 Applications mesurables

Soient X, Y deux ensembles et $\mathcal{B}_X$, $\mathcal{B}_Y$ des tribus sur X et Y respectivement. Nous disons qu'une application $T : X \to Y$ est *mesurable* si pour tout $B \in \mathcal{B}_Y$, sa pré-image par T, $T^{-1}(B)$ est dans $\mathcal{B}_X$.

Si $T : (X, \mathcal{B}_X) \to (Y, \mathcal{B}_Y)$ est mesurable, alors pour toute mesure de probabilité μ sur $(X, \mathcal{B}_X)$ on peut définir la mesure $T_*\mu$ sur $(Y, \mathcal{B}_Y)$ par : $\forall\ B \in \mathcal{B}_Y$, $T_*\mu(B) = \mu(T^{-1}B)$. On dit que $T_*\mu$ est la *mesure image* de μ par T.

6.3 Points de densité

Notons $\mathcal{B}$ la tribu borélienne de $\mathbb{R}^d$ et par m la mesure de Lebesgue d-dimensionnelle (définie par le théorème de Cararheodory par : pour tous intervalles $I_1, \ldots, I_d$, $m(I_1 \times \cdots \times I_d) = |I_1| \cdots |I_d|$). Soit A un ensemble mesurable (c'est-à-dire dans la tribu borélienne complétée) de $\mathbb{R}^d$. On dit que $x \in A$ est un *point de densité* de A si

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

ou nous avons noté $B(x,\epsilon)$ la boule de centre x et de rayon ϵ. Le théorème suivant est dû à Lebesgue :

Théorème 6.3.1 *Si A est un ensemble mesurable de mesure de Lebesgue strictement positive, alors m-presque tout point de A est de densité.*

6.4 Espérance conditionnelle

Soient $(X, \mathcal{B}, \mu)$ un espace probabilisé et $\mathcal{A}$ une tribu telle que $\mathcal{A} \subset \mathcal{B}$. L'espace $L^2(X, \mu, \mathcal{A})$ des fonctions $f : X \to \mathbb{R}$ qui sont $\mathcal{A}$-mesurables et $L^2(\mu)$ est un sous-espace fermé de l'espace de Hilbert $L^2(X, \mu, \mathcal{B})$. La projection orthogonale de $L^2(X, \mu, \mathcal{B})$ sur $L^2(X, \mu, \mathcal{A})$ s'appelle l'espérance conditionnelle sachant la tribu $\mathcal{A}$ et se note $\mathbf{E}(\cdot|\mathcal{A})$.

Il est souvent commode de pouvoir étendre cette espérance conditionnelle à l'espace $L^1(X, \mu)$. On procède de la façon suivante. Si $f \in L^1(X, \mu)$ on peut définir la mesure (bornée) $\mu_f(\cdot) = \int_\cdot f d\mu$. Sa restriction à $\mathcal{A}$ est une mesure qui est absolument continue par rapport à la mesure μ restreinte à $\mathcal{A}$. D'après le théorème de Radon-Nikodym, il existe une unique fonction $g \in L^1(X, \mu, \mathcal{A})$ telle que $\mu_f = g d\mu_{|\mathcal{A}}$. On appelle g l'espérance conditionnelle de f sachant la tribu $\mathcal{A}$; c'est une fonction $L^1(\mu)$ et $\mathcal{A}$-mesurable caractérisée par : pour toute fonction $\varphi : X \to \mathbb{R}$ qui est $\mathcal{A}$-mesurable et $L^\infty(\mu)$:

$$\int_X \varphi \mathbf{E}(f) d\mu = \int_X \varphi f d\mu.$$

Il est facile de vérifier que

$$\|\mathbf{E}(f)\|_{L^1(\mu)} \leq \|f\|_{L^1(\mu)}.$$