UNIVERSITÉ DE PICARDIE JULES VERNE
Laboratoire Amiénois de Mathématique Fondamentale et Appliquée
UMR 6140 du CNRS

THÉORIE ERGODIQUE

M2 AAM de mathématiques
Amiens, 2011-2012

Table des matières

Chapitre 1. Introduction	3
Chapitre 2. Dynamiques topologiques	5
1. Notions de base	5
2. Homéomorphismes de l'intervalle	6
3. Rappels de topologie	8
4. Exemples de dynamiques sur les tores	11
5. Les décalages	13
Chapitre 3. Dynamiques mesurées	15
1. Notions de bases et exemples	15
2. Existence d'une mesure invariante	16
Chapitre 4. Théorèmes ergodiques	23
1. Théorème de récurrence	23
2. Théorème ergodique de Von Neumann	24
3. Théorème ergodique ponctuelle ou Théorème de Birkhoff	27
Chapitre 5. Mélange et ergodicité	31
1. Systèmes ergodiques	31
2. Théorie spectrale	33
3. Systèmes mélangeants.	35
4. Exercices	38
Chapitre 6. Entropie topologique	39
1. Définition à l'aide des recouvrements ouverts	39
2. Définition de Bowen	42
3. Egalité des deux définitions	46
4. Calcul de l'entropie topologique	49
5. Les shifts ou décalages	51
6. Les homéomorphismes du cercle	51
Chapitre 7. Entropie métrique	55
1. Information et entropie	55
2. Calcul d'entropie	58

3. Propriété de l'entropie	59
4. Rappels sur les martingales	63
5. Preuve du théorème de Shanon-Breiman-Mac Millan	65
6. Preuves des théorèmes d'Abramov et de Sinaiï	69
Chapitre 8. Relation entre l'entropie topologique et l'entropie mesurée	73
Bibliographie	79

CHAPITRE 1

Introduction

Les systèmes dynamiques concernent l'étude de l'évolution d'un système.

Par exemple, si l'on considère une population de lapins et de renards. On s'intéressera uniquement au nombre de lapin $L(t) \in \mathbb{R}$ au temps t et au nombre de renards $R(t) \in \mathbb{R}$. Nous savons que plus il y a de renard et moins il y a de lapin et en conséquence il y aura moins de renard. Ainsi les valeurs $L(t)$ et $R(t)$ influencent les futures valeurs des nombres de lapin et de renard. Au prix de quelques hypothèses de nature non mathématique, on peut modéliser cette population par une *évolution discrète*, c'est-à-dire par une loi de la forme

$$\begin{aligned}R(t+1) &= f(L(t), R(t)) \\L(t+1) &= g(L(t), R(t))\end{aligned}$$

où $f, g: \mathbb{R}^2 \rightarrow \mathbb{R}$ sont deux fonctions.

On pourrait également modéliser cette population par une *évolution continue* (en temps) de la forme

$$\begin{cases} \frac{dR(t)}{dt} = f(L(t), R(t)) \\ \frac{dL(t)}{dt} = g(L(t), R(t)). \end{cases}$$

Ainsi si l'on considère le domaine $X \subset \mathbb{R}^2$ de toutes les valeurs possibles des populations de lapins et de renards, la loi d'évolution nous amène à étudier, soit une application $F: X \rightarrow X$ pour une évolution discrète, soit une équation différentielle dans le domaine X .

Il existe de nombreux exemples où l'on ne peut pas résoudre explicitement les équations différentielles ou discrètes. Par exemple celles qui régissent le mouvement de 3 planètes selon la mécanique Newtonnienne. On a donc intérêt, d'après le célèbre scientifique Henri Poincaré, à les étudier qualitativement.

Dans cette optique, on s'intéressera aux comportements asymptotiques (i.e. quand $t \rightarrow +\infty$) qualitatifs des systèmes : vers quoi tend une solution ?

À part pour des exemples très simples, il est difficile de répondre à cette question pour toutes les orbites. Notre objectif sera d'apporter une réponse pour presque toutes les orbites, grâce aux théorèmes ergodiques de Von Neumann et de Birkhoff. Nous regarderons le comportement en moyenne et pour presque toutes les conditions initiales, des observables au cours de l'évolution du système.

Pour nous forger une intuition, nous étudierons tout d'abord quelques exemples de dynamiques sur des espaces simple comme le cercle, le tore et sur l'ensemble des suites. Puis nous nous intéresserons aux systèmes d'un point de vue mesuré. Nous étudierons les phénomènes de récurrence et les théorèmes ergodiques de Birkhoff et de Von Neumann. Finalement nous introduirons un indicateur appelé *entropie* qui nous permettra de mesurer quantitativement à quel point un système dynamique est compliqué.

Pour ce faire, il sera nécessaire de s'aider grâce aux livres suivants :

- ARNOLD, *Ergodic Problems of Classical Mechanics*. Addison Wesley, 1989, 303 pp
- DEVANEY, *An introduction to chaotic dynamical systems*. The Benjamin/Cummings Publishing Co., Inc., Menlo Park, CA, 1986. 320 pp.
- KATOK- HASSELBLATT, *Introduction to the modern theory of dynamical systems*. Encyclopedia of Mathematics and its Applications, 54. Cambridge University Press, Cambridge, 1995. 802 pp.
- PETERSEN, K. *Ergodic theory*. Cambridge Studies in Advanced Mathematics, 2. Cambridge University Press, Cambridge, 1983
- POLLICOTT, M. ; YURI, M. *Dynamical systems and ergodic theory*. London Mathematical Society Student Texts, 40. Cambridge University Press, Cambridge, 1998.
- QUEFFÉLEC, M. *Substitution dynamical systems – spectral analysis*. Lecture Notes in Mathematics, 1294. Springer-Verlag, Berlin, 1987.
- WALTERS, P. *An introduction to ergodic theory*. Graduate Texts in Mathematics, 79. Springer-Verlag, New York-Berlin, 1982.
- ZINSMEISTER, M. *Formalisme thermodynamique et systèmes dynamiques holomorphes*. Panoramas et Synthèses , 4. Société Mathématique de France, Paris, 1996.

Bien entendu cette liste n'est pas exhaustive, Il y aura également d'autres références dans les chapitres. Je vous conseille vivement d'enrichir cette liste par vous même, selon votre point de vue, goûts et intérêts.

CHAPITRE 2

Dynamiques topologiques

1. Notions de base

Un *système dynamique topologique* (s.d.t.) est la donnée (X, f) où X est un espace topologique métrique et $f: X \rightarrow X$ un homéomorphisme (i.e. f est une bijection continue et f^{-1} est également continue).

Définition 1. Soit $x \in X$. On appelle l'orbite du point x la suite

$$\text{orb}(x) = \{f^n(x), n \in \mathbb{Z}\}$$

où $f^n = f \circ \dots \circ f$ désigne l'application f composée n fois.

On s'intéresse aux points d'accumulation de $\text{orb}(x)$.

Ainsi l'ensemble des points d'accumulations pour les itérés positifs est appelé l'ensemble $\omega(x)$ *omega limite*. De façon ensembliste, on a

$$\begin{aligned} \omega(x) &:= \{y, \exists n_i \rightarrow +\infty \text{ tq } f^{n_i}(x) \rightarrow y\} \\ &= \bigcap_{n \geq 0} \overline{\{f^i(x), i \geq n\}}. \end{aligned}$$

De la seconde égalité, on en déduit que cet ensemble est fermé et invariant par f , ie. $f(\omega(x)) = \omega(x)$.

De la même façon, l'ensemble des points d'accumulations pour les itérés négatifs est appelé l'ensemble $\alpha(x)$ *alpha limite*. On a également

$$\begin{aligned} \alpha(x) &:= \{y, \exists n_i \rightarrow -\infty \text{ tq } f^{n_i}(x) \rightarrow y\} \\ &= \bigcap_{n \geq 0} \overline{\{f^{-i}(x), i \geq n\}}. \end{aligned}$$

L'ensemble $\alpha(x)$ est l'ensemble omega-limite pour l'application f^{-1} .

1.1. Quelques comportements. Point fixe. le point x est dit *point fixe* si $f(x) = x$. ainsi $\text{orb}(x) = \{x\}$.

On notera par $\text{Fix}(f)$ l'ensemble des points fixes de f .

Point périodique. Un tel point est tel qu'il existe un entier $n > 0$ tel que $f^n(x) = x$. Dans ce cas

$$\text{orb}(x) = \{x, f(x), \dots, f^n(x)\},$$

où $n = \inf\{n > 0, f^n(x) = x\}$ est la période de x .

Point récurrent. Si $x \in \omega(x)$.

C'est le cas par exemple si x est un point périodique.

1.2. Comparaison de systèmes. Si l'on étudie un système physique avec une loi d'évolution discrète, un changement d'unité, comme la conversion de mètres en miles, conduit à un changement de loi et d'espace. le nouveau système obtenu est alors équivalent au premier selon le sens suivant.

Définition 2. Soient $f: X \rightarrow X$ et $g: Y \rightarrow Y$ deux homéomorphismes. f et g sont conjugués ssi il existe un homéomorphisme $h: X \rightarrow Y$ tel que le diagramme commute.

$$\begin{array}{ccc} X & \xrightarrow{f} & X \\ h \downarrow & & \downarrow h \\ Y & \xrightarrow{g} & Y \end{array}$$

i.e. vérifiant $g \circ h = h \circ f$.

Si l'application h n'est pas inversible mais est surjective, on dira alors que le système (X, f) est *semi-conjugué* à (Y, g) . ou que (Y, g) est un *facteur* de (X, f) .

On remarque aisément que dans le cas de la conjugaison on a

$$g^n = h \circ f^n \circ h^{-1} \quad \forall n \in \mathbb{Z}.$$

Pour $x \in X$, soit $y = h(x) \in Y$, on a

$$\begin{aligned} h(\text{orb}_f(x)) &= \text{orb}_g(y) \\ h(\omega(x, f)) &= \omega(y, g) \\ x \text{ est périodique pour } f &\iff y \text{ est périodique pour } g \\ x \text{ est récurrent pour } f &\iff y \text{ est récurrent pour } g. \end{aligned}$$

La relation de conjugaison définit une relation d'équivalence sur la collection de tous les systèmes dynamiques.

On étudie donc les systèmes à conjugaison près². Parfois grâce à une bonne conjugaison les calculs se simplifient grandement.

2. Homéomorphismes de l'intervalle

Exemple fondamental. Considérons pour un paramètre $a \in [-1, 1]$ l'application

$$f_a(x) = x(ax + 1 - a).$$

On constate qu'on a $f_a(0) = 0$, $f_a(1) = 1$ et $f_a: [0, 1] \rightarrow [0, 1]$ et $f'_a \geq 0$ sur $[0, 1]$.

$f_0 = Id$

Pour $a \in]0, 1]$, on a $f(x) < x \forall x \in]0, 1[$,

² Malheureusement des exemples montrent que l'espace quotient n'est même pas séparé. Ceci limite les résultats très généraux.

et donc $f^n(x) < \dots < f^2(x) < f(x) < x$. Ainsi $\omega(x) = \{0\}$ si $x \neq 1$.

Pour $a \in [-1, 0[$, on a $f(x) > x \forall x \in]0, 1[$

et donc $f^n(x) > \dots > f^2(x) > f(x) > x$. Ainsi $\omega(x) = \{1\}$ si $x \neq 0$.

On constate un changement brutal de comportement des orbites lorsque le paramètre a change de signe. Ceci arrive alors même que les système s dépendent continuellement (et même de façon lisse) de ce paramètre. Ce phénomène est appelé *bifurcation*.

A l'opposé, on constate que ces comportements ne dépendent pas du paramètre si celui-ci ne change pas de signe. nous allons même montrer plus, nous allons montrer que deux tels systèmes sont conjugués.

Considérons deux paramètres $a, b \in [-1, 0[$, (le cas $]0, 1]$ est similaire) et deux points $x, y \in]0, 1[$.

Soit $h_0: [x, f_a(x)] \rightarrow [y, f_b(y)]$ un homéomorphisme croissant comme par exemple une transformation affine.

Proposition 3. h_0 s'étend de façon unique en un homéomorphisme croissant $h: [0, 1] \rightarrow [0, 1]$ tel que

$$f_b = h \circ f_a \circ h^{-1}.$$

DÉMONSTRATION. On note par $I = [x, f_a(x)]$ et $J = [y, f_b(y)]$. On a

$$]0, 1[= \bigcup_{n \in \mathbb{Z}} f_a^n(I) = \bigcup_{n \in \mathbb{Z}} f_b^n(J).$$

On a un découpage de l'intervalle $]0, 1[$, et on va définir h par morceaux sur ces sous-intervalles.

Soit $h_n: f_a^n(I) \rightarrow f_b^n(J)$ défini par

$$h_n := f_b^n \circ h_0 \circ (f_a^n)^{-1}.$$

On définit alors $h:]0, 1[\rightarrow]0, 1[$ par $h = h_n$ sur $f_a^n(I)$.

L'application h est bien définie car les intervalles $f_a^n(I)$ s'intersectent sur leur bords. h est continue et croissante (Exercice). h est un homéomorphisme et on peut le prolonger par continuité en 0 et 1. $\square$

On peut remarquer que cette preuve est très générale et n'utilise pas du tout le fait que les applications f_a sont des polynômes. En suivant cette preuve, on peut montrer le résultat suivant :

Théorème 4. *Soient f, g deux homéomorphismes croissants de $[0, 1]$. Alors les deux affirmations suivantes sont équivalentes :*

- (1) *Il existe $h : [0, 1] \rightarrow [0, 1]$ un homéomorphisme croissant tel que $g = h \circ f \circ h^{-1}$.*
- (2) *Il existe $h_0 : [0, 1] \rightarrow [0, 1]$ un homéomorphisme croissant tel que*
  - $h_0(\text{Fix}(f)) = \text{Fix}(g)$
  - *Pour toute composante connexe c de $[0, 1] \setminus \text{Fix}(f)$, $f - \text{Id}$ sur c et $g - \text{Id}$ sur $h_0(c)$ ont même signe.*

La preuve est laissée à titre d'exercice.

3. Rappels de topologie

3.1. Les tores. Le tore $\mathbb{T}^k$ est le quotient de $\mathbb{R}^k$ par $\mathbb{Z}^k$:

$$\mathbb{T}^k = \mathbb{R}^k / \mathbb{Z}^k.$$

En d'autres termes $\mathbb{T}^k$ est l'ensemble des classes d'équivalence pour la relation $\equiv$ définie par :

$$x \equiv y \Leftrightarrow \exists z \in \mathbb{Z}^k, x = y + z.$$

Si $x \in \mathbb{R}^k$, on note $\pi(x) \in \mathbb{T}^k$ sa classe d'équivalence pour $\equiv$. On dit que $\pi : \mathbb{R}^k \rightarrow \mathbb{T}^k$ est la *projection canonique*. On note $\mathbb{T}$ le tore $\mathbb{T}^1$.

Proposition 5. *Pour tous $x, y \in \mathbb{R}^k$ la classe $\pi(x + y)$ ne dépend que des classes $\pi(x)$ et $\pi(y)$.*

L'ensemble $\mathbb{T}^k$ est un groupe abélien pour l'addition définie par

$$\pi(x) + \pi(y) = \pi(x + y).$$

L'élément neutre est $\pi(0)$ et $\pi(-x)$ est l'inverse de $\pi(x)$.

La preuve est laissée à titre d'exercice.

Comme dans tout groupe abélien noté additivement, on note par 0 l'élément neutre de l'addition sur $\mathbb{T}^k$ et par $-t$ l'inverse pour l'addition de $t \in \mathbb{T}^k$.

Remarquons que si $f : \mathbb{T}^k \rightarrow X$ est une application de $\mathbb{T}^k$ dans l'ensemble X , alors $f \circ \pi$ est une application définie sur $\mathbb{R}^k$ à valeurs dans X qui est $\mathbb{Z}^k$ -périodique :

$$\forall x \in \mathbb{R}^k, \forall z \in \mathbb{Z}^k, f \circ \pi(x + z) = f \circ \pi(x).$$

Exercice 1. *Soit $\theta : \mathbb{R}^k \rightarrow X$ une application $\mathbb{Z}^k$ -périodique à valeurs dans X . Montrer qu'il existe une et une seule application $\bar{\theta} : \mathbb{T}^k \rightarrow X$ telle que $\bar{\theta}\pi = \theta$.*

3.2. La topologie des tores. On munit $\mathbb{T}^k$ de la topologie quotient. Cette topologie peut être définie de la manière suivante. Soit $\|\cdot\|$ une norme sur $\mathbb{R}^k$. On définit sur $\mathbb{T}^k$ une distance par :

$$d(s, t) = \inf\{\|x - y\|; x, y \in \mathbb{R}^k, \pi(x) = s, \pi(y) = t\}.$$

La topologie quotient est alors la topologie définie par la distance d (**exercice**). Par conséquent la topologie définie par d ne dépend pas de la norme choisie.

Lemme 6. *La métrique d sur $\mathbb{T}^k$ est invariante par les translations, c'est-à-dire :*

$$\forall s, t \in \mathbb{T}^k, d(s, t) = d(s - t, 0) = d(0, t - s).$$

De plus on a

$$\forall x, y \in \mathbb{R}^k, \exists z \in \mathbb{Z}^k, d(\pi(x), \pi(y)) = \|x - y + z\|.$$

DÉMONSTRATION. Pour la première partie du lemme, on remarque que l'on a :

$$\{x - y; \pi(x) = s, \pi(y) = t\} = \{w; \pi(w) = s - t\}.$$

La seconde partie résulte du fait que si $x, y \in \mathbb{R}^k$, alors il n'y a qu'un nombre fini de $z \in \mathbb{Z}^k$ tels que $\|x - y + z\| \leq \|x - y\|$. $\square$

Proposition 7. (1) *L'addition sur $\mathbb{T}^k$ et l'application $t \mapsto -t$ sont continues.*

(2) *L'application π est continue et ouverte.*

(3) *$\mathbb{T}^k$ est compact.*

DÉMONSTRATION. Exercice. $\square$

3.3. Les ensembles de Cantor.

Définition 8. *Un espace topologique F est appelé ensemble de Cantor si*

- F est compact métrisable,
- F totalement discontinue, i.e. la composante connexe de tout point est un singleton,
- F n'a pas de point isolé.

Comme les composantes connexes des sous ensembles de $\mathbb{R}$ sont des intervalles et qu'ils engendrent la topologie, nous avons la propriété suivante :

Proposition 9. *Si F est un compact de $\mathbb{R}$. Alors*

$$F \text{ est ensemble de Cantor} \iff \begin{cases} F \text{ est d'intérieur vide} \\ F \text{ n'a pas de point isolé.} \end{cases}$$

Pour construire un exemple d'ensemble de Cantor de $\mathbb{R}$, c'est un peu plus technique, mais très classique. On peut de cette façon construire l'*ensemble triadique de Cantor*.

On note par $K_0 = [0, 1]$ et on définit par récurrence une suite de compacts emboîtés

$$K_{i+1} = K_i \setminus \bigcup_{\text{fini}} I_j,$$

où les I_j désignent des intervalles ouverts disjoints de sorte que chaque composante connexe de K_i contienne en son intérieur au moins un intervalle $\overline{I_j}$.

Il en résulte que l'ensemble $K_\infty = \bigcap_{i \geq 0} K_i$ est un compact non vide d'intérieur vide, et est donc un ensemble de Cantor.

Pour être plus complet sur ce sujet, on peut démontrer, mais c'est plus long, que tous les ensembles de Cantor sont homéomorphes entre eux. En particulier, on a :

Théorème 10. *Un ensemble topologique est un ensemble de Cantor si et seulement s'il est homéomorphe à K_∞ .*

Nous admettrons la preuve de ce théorème.

3.4. Une propriété générale des systèmes dynamiques. Soient X un espace métrique et f un homéomorphisme de X . Un ensemble fermé K est dit *invariant* par f si $f(K) = K$.

Comme pour tout point x de X , son orbite $orb(x)$ est un sous ensemble invariant par f , sa fermeture est un fermé invariant par f .

Définition 11. *Un ensemble fermé invariant K est dit minimal pour f s'il n'existe pas de sous ensemble fermé de K invariant non vide différent de K .*

Par exemple si x est un point périodique de f , son orbite est un ensemble minimal pour f .

On obtient facilement la caractérisation suivante :

Un ensemble K est minimal pour f si et seulement si $\overline{orb(x)} = K, \forall x \in K$.

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

Théorème 12. *Si X est un espace métrique compact et f un homéomorphisme de X . Alors il existe un compact minimal invariant par f .*

DÉMONSTRATION. Il suffit pour cela d'appliquer le lemme de Zorn, à la collection des ensembles compacts non vide invariants par f pour la relation d'ordre donné par l'inclusion. Cette collection n'est pas vide puisque X y appartient. Si l'on considère un chaîne, i.e. une famille dénombrable de compacts non vide emboîtés (pour l'inclusion). Leur intersection est un compact non vide invariant par f , c'est donc un plus petit élément. Par le lemme de Zorn, il existe un plus petit compact non vide invariant par f . $\square$

Ainsi pour un espace compact X , pour un point $x \in X$ quelconque, l'ensemble $\omega(x)$ est un compact invariant qui contient un ensemble minimal invariant.

4. Exemples de dynamiques sur les tores

4.1. Homéomorphismes du cercle. On peut voir le cercle de deux façons différentes. Soit comme l'ensemble des nombres complexes de module 1. Dans cette façon géométrique on le note $\mathbb{S}^1$, il s'agit du cercle de longueur 2π .

$$\mathbb{S}^1 = \{e^{2i\pi t}, t \in [0, 1]\}.$$

Soit de façon algébrique par

$$\mathbb{R}/\mathbb{Z} = \{t \bmod 1, t \in [0, 1]\}.$$

Ces deux versions sont homéomorphes par la bijection $h: t \mapsto e^{2i\pi t}$, en mettant la topologie quotient sur $\mathbb{R}/\mathbb{Z}$.

Ainsi un homéomorphisme sur $\mathbb{R}/\mathbb{Z}$ définit par conjugaison via h un homeomorphisme sur $\mathbb{S}^1$ et inversement.

Dans la suite on se placera dans $\mathbb{S}^1$ ou $\mathbb{R}/\mathbb{Z}$ en fonction de la commodité des écritures.

On appelle le revêtement universel du cercle l'application

$$\begin{aligned} p: \mathbb{R} &\rightarrow \mathbb{S}^1 \\ t &\mapsto e^{2i\pi t}. \end{aligned}$$

L'orientation de $\mathbb{R}$ induit, via cette application une orientation sur le cercle.

On dira qu'un homéomorphisme $f: \mathbb{S}^1 \rightarrow \mathbb{S}^1$ préserve l'orientation s'il préserve celle ci : i.e. si l'image par f d'un segment orienté est un segment avec la même orientation que celle induite par le cercle.

On remarquera que cette proposition est invariante par composition.

Exemple. La rotation d'angle $\alpha \in \mathbb{R}$

$$R_\alpha(t) := t + \alpha \bmod 1 \quad t \in \mathbb{R}/\mathbb{Z}.$$

Ce qui correspond sur $\mathbb{S}^1$ à la rotation géométrique d'angle $2\pi\alpha$ radian :

$$h \circ R_\alpha \circ h^{-1}(z) = ze^{2i\pi\alpha}.$$

Exemple. Si $F: \mathbb{R} \rightarrow \mathbb{R}$ est un homéomorphisme croissant tel que $1 + F(x) = F(x + 1)$ $\forall x \in \mathbb{R}$, alors F induit un homéomorphisme préservant l'orientation de $\mathbb{R}/\mathbb{Z}$.

Par exemple les applications $F(t) = t + \alpha + \beta \sin(2\pi t)$, définies pour $t \in \mathbb{R}$ avec $\alpha \in \mathbb{R}$ et $\beta \in]\frac{-1}{2\pi}, \frac{1}{2\pi}[$.

4.2. Théorème de Poincaré.

Théorème 13 (Classification des homéomorphismes du cercle (Poincaré)). *Soit $f: \mathbb{S}^1 \rightarrow \mathbb{S}^1$ un homéomorphisme du cercle préservant l'orientation. Alors on a l'un de ces trois cas exclusifs :*

- Il existe une orbite finie.
- Le système $(\mathbb{S}^1, f)$ est minimal.
- Il existe un unique ensemble K minimal invariant par f et K est un ensemble de Cantor.

Exercice. Donner un exemple illustrant chacun des cas.

DÉMONSTRATION. Soit K un ensemble compact minimal non vide pour f .

Le bord $\partial K \subset K$ est un sous ensemble fermé invariant par f .

Si $\partial K = \emptyset$, alors K est ouvert et fermé et par connexité du cercle $K = \mathbb{S}^1$. Ainsi le système $(\mathbb{S}^1, f)$ est minimal.

Si $\partial K = K$ et K est d'intérieur vide. Notons par K' l'ensemble des points d'accumulations de K . C'est un ensemble fermé invariant par f . Si $K' = \emptyset$ alors K est discret et f a une orbite finie.

Si $K' = K$ et K n'a pas de points isolés et d'intérieur vide. C'est un ensemble de Cantor.

Il nous reste à montrer, dans le dernier cas, l'unicité de l'ensemble minimal.

Affirmation. Pour tout point $x \in \mathbb{S}^1 \setminus K$, $\omega(x) = K$.

Par la caractérisation des composantes connexes de $\mathbb{R}$, on a $\mathbb{S}^1 \setminus K = \bigsqcup]a_n, b_n[$ avec $a_n, b_n \in K$.

Soient $x \in]a_{n_0}, b_{n_0}[= I_{n_0}$ et $y \in K$. Par minimalité de K , il existe une suite d'entiers $p_i \rightarrow +\infty$ tq $\lim_i f^{p_i}(x) = y$.

Comme $f(K) = K$, on a $f(\mathbb{S}^1 \setminus K) = \mathbb{S}^1 \setminus K$.

Les intervalles $(f^{p_i}(I_{n_0}))_i$ sont deux à deux disjoints : sinon il existe deux entiers $p \neq p'$ tels que $f^p(I_{n_0}) \cap f^{p'}(I_{n_0}) \neq \emptyset$. Ainsi $I_{n_0} \cap f^{p'-p}(I_{n_0}) \neq \emptyset$ et donc $f^{p'-p}(I_{n_0}) = I_{n_0}$ car f préserve les composantes connexes de $\mathbb{S}^1 \setminus K$. Comme f préserve l'orientation, les bords sont envoyés sur les bords et a_{n_0} est un point fixe de $f^{p'-p}$. Ce qui est absurde car K n'est pas fini.

On a donc

$$\sum_i |f^{p_i}(I_{n_0})| = \left| \bigsqcup_i f^{p_i}(I_{n_0}) \right| \leq |\mathbb{S}^1| = 2\pi.$$

Ainsi $|f^{p_i}(I_{n_0})| \rightarrow_{i \rightarrow \infty} 0$ et $f^{p_i}(x) \rightarrow y$. Donc $K \subset \omega(x)$.

---

3. Exercice !

Réciproquement, si $z \in \omega(x) \setminus K$ alors $z \in]a_{n1}, b_{n1}[= I_{n1}$ et pour une infinité d'entiers p , on a $f^p(x) \in I_{n1}$ et donc $f^p(I_{n0}) \cap I_{n1} \neq \emptyset$ ce qui est impossible car sinon les bords seraient fixés par des itérés de f . $\square$

On peut dire beaucoup plus sur les homéomorphismes du cercle. A ce sujet, on consultera GHYS, E. *Groups acting on the circle*. Enseign. Math. (2) 47 (2001), no. 3-4, 329407.

4.3. Homéomorphismes linéaires du tore. Considérons une application linéaire $A : \mathbb{R}^k \rightarrow \mathbb{R}^k$ envoyant $\mathbb{Z}^k$ dans lui-même. L'application A induit une application de $\mathbb{T}^k$ dans lui-même que nous noterons $\bar{A}$. Elle vérifie $\bar{A} \circ \pi = \pi \circ A$.

On remarque que $\bar{A}(s+t) = \bar{A}(s) + \bar{A}(t)$ pour tout s, t dans $\mathbb{T}^k$. Pour cette raison nous dirons que $\bar{A}$ est un homomorphisme linéaire du tore $\mathbb{T}^k$. Si $\bar{A} : \mathbb{T}^k \rightarrow \mathbb{T}^k$ est inversible, nous dirons que c'est un automorphisme linéaire du tore $\mathbb{T}^k$.

Proposition 14. *Soit $A : \mathbb{R}^k \rightarrow \mathbb{R}^k$ une application linéaire telle que $A(\mathbb{Z}^k) \subset \mathbb{Z}^k$. L'application $\bar{A}$ est surjective si et seulement si $\det A \neq 0$. Elle est injective si et seulement si $|\det A| = 1$; Dans ce cas elle est bijective.*

DÉMONSTRATION. Pour la première partie il suffit de montrer que $\bar{A}$ est surjective si et seulement si A est surjective. Soit V le sous-espace vectoriel $A(\mathbb{R}^k)$. Si $\bar{A}$ est surjective, on obtient que $\mathbb{R}^k = \cup_{z \in \mathbb{Z}^k} V + z$. Ceci implique que $V = \mathbb{R}^k$. En effet, si $V \neq \mathbb{R}^k$ alors il existerait une forme linéaire non nulle $l : \mathbb{R}^k \rightarrow \mathbb{R}$ telle que $l|_V = 0$. Ce qui donne $\mathbb{R} = l(\mathbb{R}^k) = l(\cup_{z \in \mathbb{Z}^k} V + z) = \{l(z); z \in \mathbb{Z}^k\}$. C'est absurde car le membre de droite est dénombrable.

Réciproquement, il est évident que si A est surjective alors $\bar{A}$ l'est également.

Si $\bar{A}$ est injective, on a nécessairement $A^{-1}(\mathbb{Z}^k) \subset \mathbb{Z}^k$ car $\bar{A} \circ \pi(A^{-1}(\mathbb{Z}^k)) = \pi \circ A(A^{-1}(\mathbb{Z}^k)) = \pi(\mathbb{Z}^k) = [0]$; En particulier, l'application A est injective puisque son noyau est un sous-espace vectoriel inclus dans $\mathbb{Z}^k$, c'est-à-dire $\{0\}$. Dans ce cas A^{-1} est bien définie et induit une application de $\mathbb{T}^k$ qui est l'inverse de $\bar{A}$. De $A^{-1}(\mathbb{Z}^k) \subset \mathbb{Z}^k$ on a également que $\det A = 1$. $\square$

Ainsi l'application $S : \mathbb{R}^2/\mathbb{Z}^2 \rightarrow \mathbb{R}^2/\mathbb{Z}^2$ définie par

$$S \begin{pmatrix} x \\ y \end{pmatrix} = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \begin{pmatrix} x \\ y \end{pmatrix},$$

est un homéomorphisme. Nous verrons sa dynamique en exercice.

5. Les décalages

Soit $K = \mathbb{N}$ ou $\mathbb{Z}$. Munissons $X = \{x = (x_n)_{n \in K}; x_n \in \{0, 1\}\} = \{0, 1\}^K$ de la topologie produit infini des topologies discrètes. Alors, X est un espace compact d'après le théorème de Tychonoff. De plus, il est métrique. En effet cette topologie est engendrée⁴ par la distance définie pour tout $(x, y) \in X^2$ par

4. Exercice

$$d(x, y) = \sum_{n \in K} \frac{\delta(x_n, y_n)}{2^{|n|}}$$

où $\delta(x_n, y_n) = 0$ si $x_n = y_n$ et 1 sinon.

Exercice 2. Montrer, dans le cas où $K = \mathbb{N}$, que la topologie de X est engendrée par la famille des ensembles de la forme

$$[u_0 u_1 \dots u_N] = \{(x_n)_{n \in \mathbb{N}}; x_i = u_i, 0 \leq i \leq N\}.$$

Ces ensembles sont appelés **cylindres**. Par analogie, donner une famille d'ensembles engendrant la topologie de X pour $K = \mathbb{Z}$.

Remarquer que ces cylindres sont ouverts et fermés. En déduire que X est **totale-ment discontinu** i.e. la composante connexe de tout point x est trivial $\{x\}$.

Montrer que contrairement aux ensembles discrets, aucun point n'est isolé

Un ensemble métrique compact, sans point isolé, totalement discontinu est appelé ensemble de *Cantor*. On peut montrer que ces ensembles sont tous homéomorphes à $\{0, 1\}^K$.

Pour tout $x \in K$, définissons T par

$$T(x) = (x_{n+1})_{n \in K}.$$

Le couple (X, T) est un s.d.t. appelé le *fullshift* (le *décalage complet*). De plus, T est bijective si et seulement si $K = \mathbb{Z}$. Si $K = \mathbb{N}$, T est surjective.

Plus généralement, nous appelons *alphabet* un ensemble fini dont les éléments sont appelés *lettres*. Soit A un alphabet. Un *mot* sur A est une suite finie d'éléments de A . Soit $x = x_0 x_1 \dots x_{n-1}$ un mot, sa *longueur* est n et se note $|x|$. Si $J = [i, j]$ est un intervalle de $\mathbb{N}$ alors x_J est le mot $x_i x_{i+1} \dots x_j$. Le mot x_J est appelé *préfixe* de x si $i = 0$. Une *occurrence* d'un mot u dans x est un entier i tel que $x_{[i, i+|u|-1]} = u$.

Pour $x \in X$ nous définissons de façon analogue la notion d'occurrence. L'ensemble des mots ayant une occurrence dans x est le *langage de x* et se note $L(x)$. Le langage de X est $L(X) = \cup_{x \in X} L(x)$. La fonction $p_X : \mathbb{N} \rightarrow \mathbb{N}$ définie par $p_X(n) = \#L(X) \cap A^n$ est appelée *fonction de complexité de X* (ou *complexité*) et sera utile lors des calculs d'entropie.

Soit Y le sous-ensemble de X constitué de suites n'ayant aucune occurrence du mot 11 : $Y = \{x \in X : 11 \notin L(x)\}$. Muni de la topologie induite Y est un espace compact et $(Y, T|_Y)$ est un s.d.t. (Exercice).

Dynamiques mesurées

1. Notions de bases et exemples

Un *système dynamique mesuré (s.d.m.)* est la donnée $(X, \mathcal{B}, \mu, T)$ où

- $(X, \mathcal{B}, \mu)$ est un espace de probabilité ;
- $T : X \rightarrow X$ est une application mesurable ;
- μ est une mesure T -invariante : Pour tout $B \in \mathcal{B}$ on a

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

Dans la plupart des situations nous supposons que T est bijective et que T^{-1} est mesurable et continue respectivement.

Plus tard nous verrons que tout s.d.t. admet une mesure de probabilité T -invariante définie sur les boréliens de X . Par conséquent on pourra voir ce s.d.t. comme un s.d.m..

De la même façon que pour les s.d.t., pour un s.d.m. et un point $x \in X$, l'*orbite* de x est l'ensemble des points $orb_x := \{T^n x, n \in \mathbb{Z}\}$ où T^n désigne l'application T composée n fois $T \circ \dots \circ T$.

Exemple. Le tore $\mathbb{T} = \mathbb{R}/\mathbb{Z}$ peut se voir comme le segment $[0, 1]$ dont on a identifié les extrémités. Il peut donc se voir comme le cercle de périmètre 1. Munissons $\mathbb{T}$ de la mesure de Lebesgue λ restreinte à $[0, 1]$, définie sur la tribu $\mathcal{B}$ des boréliens de $\mathbb{T}$. Considérons $\alpha \in \mathbb{T}$, puis définissons $R_\alpha : \mathbb{T} \rightarrow \mathbb{T}$ par $R_\alpha(x) = x + \alpha$. Le couple $(\mathbb{T}, R_\alpha)$ est un s.d.t. et le quadruplet $(\mathbb{T}, \mathcal{B}, m, R_\alpha)$ est un s.d.m..

1.1. Les isomorphismes. De même que pour les s.d.t., nous dirons que deux s.d.m. sont équivalents s'ils sont conjugués aux ensembles mesurable près. Plus précisément, nous dirons que $(Y, \mathcal{A}, \nu, S)$ est un *facteur mesuré (ou en mesure)* de $(X, \mathcal{B}, \mu, T)$ s'il existe une application mesurable et surjective $\pi : \tilde{X} \subset X \rightarrow \tilde{Y}$ telle que

$$\begin{aligned} \mu(\tilde{X}) &= 1 = \nu(\tilde{Y}), \\ \pi\mu &= \nu \text{ et} \\ S \circ \pi(x) &= \pi \circ T(x) \text{ pour tout } x \in \tilde{X}. \end{aligned}$$

Si, de plus, π est une bijection et π^{-1} est mesurable, nous dirons que $(X, \mathcal{B}, \mu, T)$ et $(Y, \mathcal{A}, \nu, S)$ sont *conjugués en mesure*.

Exercice. Considérons le full shift $(\{0, 1\}^{\mathbb{N}}, T)$ défini en section 5 du chapitre précédent. Soit $a \in [0, 1]$. Posons $m(0) = 1 - a$ et $m(1) = a$. ce ci définit une mesure sur $\{0, 1\}$. Notons μ la mesure produit définie par m sur l'ensemble des boréliens $\mathcal{B}$ de X . Montrer que cette mesure est T invariante et le quadruplet $(X, \mathcal{B}, \mu, T)$ est un s.d.m. Soit $S : [0, 1[\rightarrow [0, 1[$ définie par $S(x) = 2x \bmod 1$. Montrer que les s.d.m. $([0, 1[, \mathcal{A}, \lambda, S)$ et $(X, \mathcal{B}, \mu, T)$ sont conjugués en mesure pour $a = 1/2$.

1.2. Mesures de Haar sur les tores.

Théorème 15. *Il y a une et une seule mesure borélienne sur le tore $\mathbb{T}^k$ qui soit invariante par les translations et de masse totale 1. On note cette mesure m et on l'appelle la **mesure de Haar sur $\mathbb{T}^k$** . Si l'on note λ la mesure de Lebesgue sur $\mathbb{R}^k$, pour toute partie borélienne A de $\mathbb{T}^k$, on a $m(A) = \lambda(\pi^{-1}(A) \cap [0, 1]^k)$.*

DÉMONSTRATION. La mesure m sur $\mathbb{T}^k$ définie par $m(A) = \lambda(\pi^{-1}(A) \cap [0, 1]^k)$, pour toute partie borélienne A de $\mathbb{T}^k$, est invariante par les translations (**Exercice**). Montrons l'unicité. Considérons la famille $\mathcal{S}$ des parties de $\mathbb{T}^k$ de la forme

$$\pi([p_1/2^n, (p_1 + 1)/2^n] \times \cdots \times [p_k/2^n, (p_k + 1)/2^n]),$$

où $n \in \mathbb{N}$ et $(p_1, \dots, p_k) \in \mathbb{N}^k$ avec $0 \leq p_i \leq 2^n - 1$. La famille $\mathcal{A}$ formée par les réunions disjointes finies d'éléments de $\mathcal{S}$ est une algèbre qui engendre la σ -algèbre des boréliens de $\mathbb{T}^k$ (**Exercice**). Soit μ une mesure sur $\mathbb{T}^k$ invariante par translation et de masse totale 1. Pour chaque entier $n \in \mathbb{N}$ et chaque $(p_1, \dots, p_k) \in \mathbb{N}^k$ avec $0 \leq p_i \leq 2^n - 1$, considérons le cube :

$$C_{p_1, \dots, p_k}^n = [p_1/2^n, (p_1 + 1)/2^n] \times \cdots \times [p_k/2^n, (p_k + 1)/2^n].$$

A n fixé, la famille des $\tilde{C}_{p_1, \dots, p_k}^n = \pi(C_{p_1, \dots, p_k}^n)$ forme une partition de $\mathbb{T}^k$ par 2^{kn} ensembles boréliens deux à deux disjoints qui se déduisent l'un de l'autre par translation et de masse totale 1. On en déduit que $\mu(\tilde{C}_{p_1, \dots, p_k}^n) = 2^{-kn}$. Il en résulte que μ coïncide avec m sur $\mathcal{S}$ et donc aussi sur l'algèbre $\mathcal{A}$. Ainsi, d'après le Théorème de Carathéodory, μ et m sont égales. $\square$

Exercice. Le couple $(\mathbb{R}^2/\mathbb{Z}^2, S)$ défini section 4.2 est un s.d.t.. A-t-il une mesure invariante ?

2. Existence d'une mesure invariante

Dans cette section (X, d) est un espace métrique compact. Nous noterons $\mathcal{M}(X)$ l'ensemble des mesures de probabilité boréliennes définies sur X et $\mathcal{M}(X, T)$ le sous-ensemble de $\mathcal{M}(X)$ constitué des mesures T -invariantes. Remarquons que $\mathcal{M}(X)$ et $\mathcal{M}(X, T)$ sont convexes (**Exercice**).

Théorème 16 (Krylov-Bogolioubov). *Soit X un espace métrique compact. Tout s.d.t. (X, T) admet une mesure T -invariante, i.e. $\mathcal{M}(X, T) \neq \emptyset$.*

2.1. Rappels. Commençons par rappeler un résultat de topologie et un autre de théorie de la mesure.

Théorème 17. *Soit X un espace compact séparé. Il y a équivalence entre les assertions suivantes.*

- (1) X est métrisable.
- (2) X a une base de sa topologie qui est dénombrable.
- (3) $C(X)$ (l'espace des fonctions continues sur X à valeurs réelles) contient un sous-ensemble dénombrable dense.

Théorème 18. *Soit X un espace métrique compact. Si m est une mesure de probabilité définie sur les boréliens de X alors elle est régulière : Pour tout borélien B et tout $\varepsilon > 0$ il existe un ensemble ouvert U et un compact C tels que*

$$C \subset B \subset U \text{ et } m(U \setminus C) < \varepsilon.$$

Grâce à ce théorème obtenir le résultat suivant très utile pour la suite.

Théorème 19. *Soient m et μ deux mesures de probabilité boréliennes sur X . Alors, $m = \mu$ si et seulement si $\int_X f dm = \int_X f d\mu$ pour tout $f \in C(X)$.*

DÉMONSTRATION. D'après le Théorème 18, il suffit de montrer que nous avons $m(C) = \mu(C)$ pour tous les compacts $C \subset X$. Soient C un compact et $\varepsilon > 0$. D'après le Théorème 18 il existe un ouvert U tel que $C \subset U$ et $m(U \setminus C) < \varepsilon$. Définissons $f : X \rightarrow \mathbb{R}$ par $f(x) = 0$ si $x \notin U$ et

$$f(x) = \frac{d(x, X \setminus U)}{d(x, X \setminus U) + d(x, C)} \text{ sinon.}$$

Le dénominateur est toujours non-nulle, f est donc bien définie. De plus, f est continue, vaut 0 sur $X \setminus U$, 1 sur C , et $0 \leq f(x) \leq 1$ pour tout $x \in X$. Par conséquent,

$$\mu(C) \leq \int_X f d\mu = \int_X f dm \leq m(U) < m(C) + \varepsilon.$$

D'où $\mu(C) < m(C) + \varepsilon$ pour tout $\varepsilon > 0$, puis $\mu(C) \leq m(C)$. Par symétrie nous avons également $m(C) \leq \mu(C)$. $\square$

Rappelons le théorème de représentation intégrale de Riesz.

Théorème 20. *Soit X un espace métrique compact et $J : C(X) \rightarrow \mathbb{R}$ une application linéaire continue telle que si $f \geq 0$ alors $J(f) \geq 0$. Alors, il existe $\mu \in \mathcal{M}(X)$ telle que $J(f) = \int_X f d\mu$ pour tout $f \in C(X)$.*

2.2. Topologie faible.

Définition 21. La **topologie faible*** sur $M(X)$ est la moins fine des topologies rendant continues les fonctions $\mu \rightarrow \int_X f d\mu$ ($f \in C(X)$).

Une base de la topologie faible est donnée par la collection des ensembles de la forme

$$V_\mu(f_1, \dots, f_k; \varepsilon) = \left\{ m \in M(X); \left| \int f_i dm - \int f_i d\mu \right| < \varepsilon, 1 \leq i \leq k \right\}$$

où $\mu \in M(X)$, $k \geq 1$, $f_i \in C(X)$ et $\varepsilon > 0$.

Dans le théorème suivant nous allons prendre un sous-ensemble $\{f_i; i \in \mathbb{N}\}$ dense de $C(X)$. Ceci est toujours possible car $C(X)$ est séparable d'après le Théorème 17.

Théorème 22. L'ensemble $\mathcal{M}(X)$ est métrisable pour la topologie faible*. Si $\{f_i; i \in \mathbb{N}\}$ est un sous-ensemble dense de $C(X)$ alors

$$D(m, \mu) = \sum_{n=1}^{\infty} \frac{\left| \int f_n dm - \int f_n d\mu \right|}{2^n \|f_n\|}$$

est une distance sur $\mathcal{M}(X)$ définissant la topologie faible*.

DÉMONSTRATION. On laisse le soin au lecteur de vérifier que D est une distance sur $\mathcal{M}(X)$. Considérons l'espace métrique $(\mathcal{M}(X), D)$. Pour chaque i l'application $\mu \mapsto \int f_i d\mu$ est continue sur $(\mathcal{M}(X), D)$ car $\left| \int f_i dm - \int f_i d\mu \right| \leq 2^i \|f_i\| D(m, \mu)$. Or $\{f_i; i \in \mathbb{N}\}$ est dense dans $C(X)$ donc pour chaque $f \in C(X)$ l'application $\mu \mapsto \int f d\mu$ est continue sur $(\mathcal{M}(X), D)$. Par conséquent les ouverts pour la topologie faible* sont des ouverts pour D (par définition de la topologie faible*).

Pour montrer la réciproque il suffit de montrer que les boules $B(\mu, \varepsilon) = \{m \in \mathcal{M}(X); D(m, \mu) < \varepsilon\}$ contiennent un ensemble $V_\mu(f_1, \dots, f_k; \delta)$ où $k \geq 1$, $f_i \in C(X)$ et $\delta > 0$. Fixons $\mu \in \mathcal{M}(X)$ et $\varepsilon > 0$. Choisissons N tel que

$$\sum_{n=N+1}^{\infty} \frac{2}{2^n} < \frac{\varepsilon}{2}.$$

Posons

$$\delta = \frac{\varepsilon}{2} \left(\sum_{n=1}^N \frac{1}{2^n \|f_n\|} \right)^{-1}$$

Ainsi $V_\mu(f_1, \dots, f_k; \delta) \subset B(\mu, \varepsilon)$. □

Le lemme suivant, bien qu'évident, est très utile.

Lemme 23. Soit μ une mesure de probabilité borélienne sur X . L'application de $\mathcal{C}^0(X, \mathbb{R})$ dans $\mathbb{R}$ qui à ϕ associe $\int \phi d\mu$ est linéaire et continue de norme 1.

Faisons quelques remarques que nous utiliserons dans les preuves suivantes sans les rappeler.

(1) $\mu_n \rightarrow \mu$ dans $\mathcal{M}(X)$ si et seulement si $\int f d\mu_n \rightarrow \int f d\mu$ pour tout $f \in C(X)$ (**Exercice**).

(2) L'application de X dans $\mathcal{M}(X)$ qui à x associe la mesure de Dirac δ_x est continue.

(3) Soient $\mu_n, \mu \in M(X)$, $n \geq 1$. Les énoncés suivants sont équivalents :

- (1) $\mu_n \rightarrow \mu$ dans la topologie faible.
- (2) Pour tout sous-ensemble fermé $F \subset X$, $\limsup_{n \rightarrow \infty} \mu_n(F) \leq \mu(F)$
- (3) Pour tout sous-ensemble ouvert $U \subset X$, $\liminf_{n \rightarrow \infty} \mu_n(U) \geq \mu(U)$.
- (4) Pour chaque $A \in \mathcal{B}$ avec $\mu(\partial A) = 0$, $\mu_n(A) \rightarrow \mu(A)$.

Montrons que (1) implique (2). Soit F un fermé de X . Pour tout $k \geq 1$, définissons $U_k = \{x \in X; d(x, F) < 1/k\}$. Les ensembles U_k sont des ouverts décroissent vers F . Par conséquent $\mu(U_k) \rightarrow \mu(F)$. Procédons comme dans la preuve du Théorème 19 nous pouvons construire des fonctions $f_k \in C(X)$ telles que $0 \leq f_k \leq 1$, $f_k = 1$ sur F et $f_k = 0$ sur $X \setminus U_k$. Ainsi

$$\limsup_{n \rightarrow \infty} \mu_n(F) \leq \limsup_{n \rightarrow \infty} \int f_k d\mu_n = \int f_k d\mu \leq \mu(U_k)$$

d'où $\lim_{n \rightarrow \infty} \sup \mu_n(F) \leq \mu(F)$.

Supposons (2). Soit U un ouvert de X . Alors, $\limsup_{n \rightarrow \infty} \mu_n(X|U) \leq \mu(X|U)$ puis $\lim_{n \rightarrow \infty} \inf \mu_n(U) \geq \mu(U)$. D'où (3).

Supposons (3). Soit A tel que $\mu(\partial A) = 0$. Alors $\mu(\text{int}(A)) = \mu(A) = \mu(\bar{A})$, $\limsup_{n \rightarrow \infty} \mu_n(\bar{A}) \leq \mu(\bar{A}) = \mu(A)$ et $\liminf_{n \rightarrow \infty} \mu_n(\text{int}(A)) \geq \mu(\text{int}(A)) = \mu(A)$. Par conséquent $\mu_n(A) \rightarrow \mu(A)$. D'où (4).

Pour montrer que (4) donne (1) il suffit d'utiliser la densité des fonctions étagées dans l'ensemble des fonctions intégrables (**Exercice**).

Théorème 24. *L'ensemble $\mathcal{M}(X)$ est compact pour la topologie faible*.*

DÉMONSTRATION. Nous écrirons $\mu(f)$ à la place de $\int f d\mu$. Soit (μ_n) une suite dans $\mathcal{M}(X)$. D'après le Théorème 22 il suffit d'en extraire une sous-suite convergente.

D'après les théorèmes 22 et 17 il existe un ensemble $\{f_i; i \in \mathbb{N}\} \subset C(X)$ qui est dense dans $C(X)$. Considérons la suite de réels $(\mu_n(f))$. Elle est bornée par $\|f_1\|$. Donc on peut en extraire une sous-suite convergente $(\mu_n^{(1)}(f_1))$. La suite $(\mu_n^{(1)}(f_2))$ est également bornée, on en extrait une sous-suite convergente $(\mu_n^{(2)}(f_2))$. Remarquons que $(\mu_n^{(2)}(f_1))$ converge aussi. En procédant de cette façon nous construisons, pour chaque $i \geq 1$, une sous-suite $(\mu_n^{(i)})$ de $(\mu_n^{(i-1)})$ telle que $(\mu_n^{(i)}(f))$ converge pour $f \in \{f_1, \dots, f_i\}$. Pour chaque i , la suite $(\mu_n^{(n)}(f_i))$ converge. Par conséquent $(\mu_n^{(n)}(f))$ converge pour tout $f \in C(X)$.

Posons $J(f) = \lim_{n \rightarrow \infty} \mu_n^{(n)}(f)$. On vérifie sans difficulté que $J : C(X) \rightarrow \mathbb{R}$ est linéaire et continue, puisque $|J(f)| \leq \|f\|$. De plus $J(1) = 1$ et $J(f) \geq 0$ si $f \geq 0$. D'après le Théorème 20, il existe une mesure de probabilité définie sur les boréliens de X telle que $J(f) = \int_X f d\mu$ pour tout $f \in C(X)$, i.e

$$\int_X f d\mu_n^{(n)} \longrightarrow \int_X f d\mu.$$

Ceci achève la preuve. □

Par conséquent $\mathcal{M}(X)$ est un espace convexe, compact et métrique.

2.3. Mesures invariantes pour les transformations continues. Notons $\tilde{T}$ l'application de $\mathcal{M}(X)$ dans $\mathcal{M}(X)$ définie par

$$\tilde{T}(\mu)(B) = \mu(T^{-1}(B))$$

pour tout $B \in \mathcal{B}(X)$ et $\mu \in \mathcal{M}(X)$.

Lemme 25. *Pour tout $f \in C(X)$*

$$\int f d(\tilde{T}\mu) = \int f \circ T d\mu.$$

DÉMONSTRATION. Soit $f \in C(X)$. Par définition de $\tilde{T}$ nous avons bien sûr $\int 1_B d(\tilde{T}\mu) = \int 1_B \circ T d\mu$ pour tout $B \in \mathcal{B}(X)$. Donc nous aurons l'égalité pour toutes les fonctions étagées. Nous étendons l'égalité aux fonctions continues par densité et limite croissante. □

Théorème 26. *L'application $\tilde{T}$ est continue.*

DÉMONSTRATION. **Exercice.** □

Remarquons que $\mathcal{M}(T) = \{\mu \in \mathcal{M}(X); \tilde{T}\mu = \mu\}$.

Théorème 27. *Si T est continue et $\mu \in \mathcal{M}(X)$ alors, $\mu \in \mathcal{M}(X, T)$ si et seulement si $\int f \circ T d\mu = \int f d\mu$ pour tout $f \in C(X)$.*

DÉMONSTRATION. **Exercice.** □

Le théorème suivant a pour corollaire immédiat le Théorème de Krylov-Bogolioubov.

Théorème 28. *Soient $T : X \rightarrow X$ continue et (σ_n) une suite de $\mathcal{M}(X)$. Posons $\mu_n = (1/n) \sum_{i=0}^{n-1} \tilde{T}^i \sigma_n$ alors toute valeur d'adhérence μ de (μ_n) est un élément de $\mathcal{M}(X, T)$. (De telle μ existe par compacité de $\mathcal{M}(X)$.)*

DÉMONSTRATION. Soit (μ_{n_j}) une sous-suite convergent vers μ . Soit $f \in C(X)$. Alors,

$$\begin{aligned} \left| \int f \circ T d\mu - \int f d\mu \right| &= \lim_{j \rightarrow \infty} \left| \int f \circ T d\mu_{n_j} - \int f d\mu_{n_j} \right| \\ &= \lim_{j \rightarrow \infty} \left| \frac{1}{n_j} \int \sum_{i=0}^{n_j-1} (f \circ T^{i+1} - f \circ T^i) d\mu_{n_j} \right| \\ &= \lim_{j \rightarrow \infty} \left| \frac{1}{n_j} \int (f \circ T^{n_j} - f) d\mu_{n_j} \right| \\ &\leq \lim_{j \rightarrow \infty} \frac{2\|f\|}{n_j} = 0. \end{aligned}$$

Par conséquent $\mu \in \mathcal{M}(X, T)$. □

Exemples, exercices : soit (X, T) un système dynamique ayant un point $x \in X$ d'orbite finie (de cardinal N). Montrer que la mesure

$$\frac{1}{N} \sum_{i=1}^N \delta_{T^i x},$$

est une mesure invariante pour T .

Donner un exemple de transformation de l'intervalle qui n'admet pas de mesure invariante.

CHAPITRE 4

Théorèmes ergodiques

Commençons par quelques questions sous forme d'exercice.

Exercice 3. Soient $\alpha \in [0, 1[\setminus \mathbb{Q}$ et B un Borélien de $[0, 1[$. Que vaut

$$\lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{k=1}^n \mathbf{1}_B(x + k\alpha) ?$$

Exercice 4. Pensez-vous que pour presque tout $x \in [0, 1[$ la fréquence de 1 dans l'écriture de x en base 2 est $1/2$?

Exercice 5. Quelle est la fréquence du chiffre 7 dans la suite des premiers chiffres de 2^n , $n \geq 0$, en base 10 ?

Question. Quand a-t-on

$$\lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{n \geq 0} f(T^n(x)) = \int_X f d\mu ?$$

1. Théorème de récurrence

Théorème 29 (Théorème de récurrence de Poincaré). Soient $(X, \mathcal{B}, T, \mu)$ et $E \in \mathcal{B}$ avec $\mu(E) > 0$. Alors, presque tout point de E retourne infiniment souvent dans E sous les itérations positives de T : Il existe $F \subset E$ avec $\mu(F) = \mu(E)$ tel que pour tout $x \in F$ il existe des entiers naturels $n_0 < n_1 < n_2 < \dots$ vérifiant $T^{n_i}(x) \in F$ pour tout $i \in \mathbb{N}$.

DÉMONSTRATION. Pour tout $N \geq 0$ posons $E_N = \cup_{n=N}^{\infty} T^{-n}E$. Alors, $F = E \cap (\cap_{N=0}^{\infty} E_N)$ est l'ensemble des points de E qui entrent dans E infiniment souvent sous les itérations positive de T . En d'autres termes, $x \in E$ appartient à F si et seulement s'il existe une suite $0 < n_1 < n_2 < \dots$ d'entiers naturels tel que $T^{n_i}(x) \in E$ pour tout $i \in \mathbb{N}$. Pour tout $i \in \mathbb{N}$ nous avons $T^{n_i}(x) \in F$ car $T^{n_j - n_i}(T^{n_i}(x)) \in E$ pour tout $j \in \mathbb{N}$. Il reste à montrer que $\mu(F) = \mu(E)$.

Nous avons $T^{-1}E_N = E_{N+1}$, ainsi $\mu(E_N) = \mu(E_{N+1})$ puis $\mu(E_N) = \mu(E_0)$ pour tout N . De plus, puisque $\dots \subset E_2 \subset E_1 \subset E_0$, nous avons $\mu(F) = \mu(E \cap E_0)$. Ainsi, $\mu(F) = \mu(E)$ car $E \subset E_0$. $\square$

Corollaire 30. Soient $(X, \mathcal{B}, T, \mu)$ et $B \in \mathcal{B}$ avec $\mu(B) > 0$. Alors, pour tout $m \in \mathbb{N}$ il existe $n \geq m$ tel que

$$\mu(T^{-n}(B) \cap B) > 0.$$

Une petite question. Le théorème de récurrence de Poincaré est-il vrai lorsque la mesure est infinie ?

2. Théorème ergodique de Von Neumann

Théorème 31 (Théorème ergodique de Von Neumann). *Soient H un espace de Hilbert et $U : H \rightarrow H$ un opérateur contractant ($\|U(f)\| \leq \|f\|$ pour tout $f \in H$ où $\|\cdot\|$ est la norme de H). Soient $M = \{f \in H; Uf = f\}$ et $P : H \rightarrow M$ la projection orthogonale sur M . Alors pour tout $f \in H$ nous avons*

$$\frac{1}{N} \sum_{n=0}^{N-1} U^n(f) \xrightarrow{N \rightarrow \infty} P(f).$$

DÉMONSTRATION. Remarquons tout d'abord que U^* est également une contraction :

$$\begin{aligned} \|U^*f\|^2 &= \langle U^*h, U^*f \rangle \\ &= \langle UU^*h, f \rangle \\ &\leq \|UU^*h\| \cdot \|h\| \\ &\leq \|U^*h\| \cdot \|h\|, \end{aligned}$$

pour tout $h \in H$. Par conséquent nous avons bien $\|U^*h\| \leq \|h\|$.

Ensuite, nous avons l'inégalité suivante pour tout $h \in H$:

$$\begin{aligned} \|Uh - h\|^2 &= \langle Uh - h, Uh - h \rangle \\ &= \langle Uh, Uh \rangle - \langle Uh, h \rangle - \langle h, Uh \rangle + \langle h, h \rangle \\ &\leq \|h\|^2 - \langle h, U^*h \rangle - \langle U^*h, h \rangle + \|h\|^2. \end{aligned}$$

Par ces deux remarques nous concluons que

$$M = \{h : Uh = h\} = \{h : U^*h = h\}.$$

Observons à présent que $H = M \oplus N$ où $N = \overline{\{g - Ug : g \in H\}}$.

En effet, pour $h \in M$ et $g \in H$, nous avons $\langle h, g - Ug \rangle = \langle h, g \rangle - \langle h, Ug \rangle = \langle h, g \rangle - \langle U^*h, g \rangle = \langle h, g \rangle - \langle h, g \rangle = 0$. Puis par continuité du produit scalaire, nous avons que $N \subset M^\perp$.

Pour $h \in N^\perp$ et pour tout $g \in H$, nous avons $\langle h, g - Ug \rangle = 0$. Nous en déduisons que $\langle h, g \rangle = \langle U^*h, g \rangle$ pour tout $g \in H$, et donc que $h = U^*h$ et par la remarque précédente $h \in M$. Ainsi $H = M \oplus N$.

Prouvons le théorème pour tout $f = g - Ug$, $g \in H$. Remarquons que $P(f) = 0$ car $N^\perp = M$.

Nous avons

$$\begin{aligned} \left\| \frac{1}{N} \sum_{n=0}^{N-1} U^n f \right\| &= \left\| \frac{1}{N} \sum_{n=0}^{N-1} (U^n g - U^{n+1} g) \right\| \\ &= \frac{1}{N} \left\| (U^N g - g) \right\| \\ &\leq \frac{2\|g\|}{N} \end{aligned}$$

qui tend vers 0 quand N tend vers l'infini. Ceci prouve le théorème pour ce f .
Si $f \in N$ alors par continuité on en déduit que :

$$\left\| \frac{1}{N} \sum_{n=0}^{N-1} U^n f \right\| \rightarrow_{N \rightarrow \infty} 0.$$

Finalement, si $f \in H$ alors $f = P(f) + f_0$, $f_0 \in N$ et donc

$$\begin{aligned} \left\| \frac{1}{N} \sum_{n=0}^{N-1} U^n f - P f \right\| &= \left\| \frac{1}{N} \sum_{n=0}^{N-1} U^n (f - P f) \right\| \\ &= \left\| \frac{1}{N} \sum_{n=0}^{N-1} U^n f_0 \right\|, \end{aligned}$$

ce qui tend vers 0 puisque $f_0 \in N$. □

Avant de donner une interprétation en termes de s.d.m. du théorème précédent, faisons quelques rappels sur l'espérance conditionnelle. Pour plus de détails voir [Wi].

Théorème 32. Soient $(X, \mathcal{B}, \mu)$ un espace de probabilité, $f \in L^1(X, \mathcal{B}, \mu)$ et $\mathcal{A}$ une sous- σ -algèbre de $\mathcal{B}$. Alors, il existe une fonction g vérifiant :

- (1) g est $\mathcal{A}$ -mesurable,
- (2) $g \in L^1(X, \mathcal{B}, \mu)$ et
- (3) pour tout $A \in \mathcal{A}$ nous avons

$$\int_A g d\mu = \int_A f d\mu \quad \text{pour tout } A \in \mathcal{A}.$$

De plus, si h est une autre fonction ayant ces propriétés alors $g = h$ μ -p.p.. Nous dirons que g est une version de l'espérance conditionnelle de f par rapport à $\mathcal{A}$ ou plus simplement, lorsqu'aucune ambiguïté n'est possible, que c'est l'espérance conditionnelle de f par rapport à $\mathcal{A}$. On la note $\mathbb{E}(f|\mathcal{A})$.

Proposition 33. Soient $(X, \mathcal{B}, \mu)$ un espace de probabilité, $f \in L^1(X, \mathcal{B}, \mu)$ et $\mathcal{A}$ une sous- σ -algèbre de $\mathcal{B}$

- (1) $\int_A \mathbb{E}(f|\mathcal{A}) d\mu = \int_A f d\mu$ pour tout $A \in \mathcal{A}$.
- (2) $\mathbb{E}(f|\{\emptyset, X\}) = \int_X f d\mu$.
- (3) $\int_X \mathbb{E}(f|\mathcal{A}) d\mu = \int_X f d\mu$.
- (4) Si f est $\mathcal{A}$ -mesurable alors $\mathbb{E}(f|\mathcal{A}) = f$ μ -p.p..
- (5) $f \mapsto \mathbb{E}(f|\mathcal{A})$ est linéaire.
- (6) Si $f \geq 0$ alors $\mathbb{E}(f|\mathcal{A}) \geq 0$.
- (7) On a les analogues du théorème de convergence monotone, du lemme de Fatou et du théorème de la convergence dominée pour l'espérance conditionnelle.

(8) Si $c : \mathbb{R} \rightarrow \mathbb{R}$ est convexe et $\int_X |c \circ f| d\mu < \infty$ alors

$$\mathbb{E}(c \circ f | \mathcal{A}) \geq c(\mathbb{E}(f | \mathcal{A})).$$

(9) Si $\mathcal{C}$ est une sous- σ -algèbre de $\mathcal{A}$ alors

$$\mathbb{E}[\mathbb{E}(f | \mathcal{A}) | \mathcal{C}] = \mathbb{E}(f | \mathcal{C}).$$

(10) Si g est $\mathcal{A}$ -mesurable et bornée alors

$$\mathbb{E}(fg | \mathcal{A}) = g\mathbb{E}(f | \mathcal{A}).$$

Théorème 34. Soient $(X, \mathcal{B}, \mu, T)$ un s.d.m., $f \in L^2(X, \mathcal{B}, \mu)$ et

$$\mathcal{I} = \{B \in \mathcal{B} : \mu(T^{-1}(B) \Delta B) = 0\}$$

(qui est une σ -algèbre d'ensemble T -invariant). Alors,

$$\left\| \frac{1}{N} \sum_{n=0}^{N-1} f \circ T^n - \mathbb{E}(f | \mathcal{I}) \right\|_2 \xrightarrow{n \rightarrow \infty} 0$$

DÉMONSTRATION. Traduisons le théorème précédent :

1) $H = L^2(X, \mathcal{B}, \mu)$

2) $U : L^2(X, \mathcal{B}, \mu) \rightarrow L^2(X, \mathcal{B}, \mu)$, défini par $U(f) = f \circ T$. De plus, μ étant T -invariante nous avons

$$\int_X |f|^2 d\mu = \int_X |f \circ T|^2 d\mu.$$

3) $\|Uf\|_2^2 = \int_X |f \circ T|^2 d\mu = \int_X |f|^2 d\mu = \|f\|_2^2$, et donc U est une contraction.

Soit $M = \{f \in L^2(X, \mathcal{B}, \mu) : Uf = f \circ T = f\}$. Montrons que $P = \mathbb{E}(\cdot | \mathcal{I})$.

La fonction $P(f)$ est T invariante ($P(f) \circ T = P(f)$ presque partout). Elle est donc $\mathcal{I}$ -mesurable.

Il suffit donc montrer que pour tout $f \in L^2(X, \mathcal{B}, \mu)$ et tout $B \in \mathcal{I}$ on a

$$\int_B Pf d\mu = \int_B f d\mu.$$

Soit $f = P(f) + f_0$ où $f_0 \in N$ (nous utilisons les notations du théorème précédent). Nous avons

$$\int_B P(f) d\mu = \int_B f d\mu - \int_B f_0 d\mu = \int_B f d\mu - \int_X f_0 \cdot 1_B d\mu = \int_B f d\mu,$$

où $1_B \in M$ (exercice), $f_0 \in N$ et $N^\perp = M$. □

3. Théorème ergodique ponctuelle ou Théorème de Birkhoff

Théorème 35 (Théorème ergodique de Birkhoff). *Soient $(X, \mathcal{B}, \mu, T)$ un s.d.m. et $f \in L^1(X, \mathcal{B}, \mu)$. Alors,*

- (1) $\lim_{N \rightarrow +\infty} \frac{1}{N} \sum_{n=0}^{N-1} f \circ T^n = \bar{f}$ existe μ -presque partout ;
- (2) $\bar{f} \circ T = \bar{f}$ μ -presque partout ;
- (3) $\bar{f} \in L^1(X, \mathcal{B}, \mu)$ et $\|\bar{f}\|_1 \leq \|f\|_1$;
- (4) $\frac{1}{N} \sum_{n=0}^{N-1} f \circ T^n \xrightarrow{N \rightarrow +\infty} \bar{f}$ dans $L^1(X, \mathcal{B}, \mu)$;
- (5) Si $\mathcal{I} = \{A \in \mathcal{B}; \mu(A \Delta T^{-1}(A)) = 0\}$, alors $\bar{f} = \mathbb{E}(f|\mathcal{I})$.

DÉMONSTRATION. Nous supposons $L^1(X, \mathcal{B}, \mu)$ réel.

Pour commencer, nous aurons besoin du théorème technique suivant :

Théorème 36 (Théorème Maximal). *Soit $(X, \mathcal{B}, \mu, T)$ un s.d.m., où $f \in L^1(X, \mu)$. Pour tout $x \in X$ et $n \in \mathbb{N}$, notons*

$$s_n(x) = f(x) + f(Tx) + \dots + f(T^n x),$$

et $A = \{x : \sup_n s_n(x) > 0\}$. Alors $\int_A f d\mu \geq 0$.

DÉMONSTRATION. Notons $S_n(x) = \max\{s_0(x), \dots, s_n(x)\}$ et $S_n^+ = \max\{S_n(x), 0\}$. Nous avons $S_n^+(Tx) + f(x) = \max\{0, s_0(Tx), \dots, s_n(Tx)\} + f(x) = \max\{f(x), s_1(x), \dots, s_{n+1}(x)\} = S_{n+1}(x)$.

Soit $A_n = \{x : S_n(x) > 0\}$, nous avons

$$\begin{aligned} \int_{A_n} f d\mu &= \int_{A_n} S_{n+1} - S_n^+ \circ T d\mu \\ &\geq \int_{A_n} S_n - S_n^+ \circ T d\mu && \text{car la suite } S_n \text{ est croissante} \\ &= \int_{A_n} S_n^+ - S_n^+ \circ T d\mu && \text{car } S_n = S_n^+ \text{ sur } A_n \\ &= \int_X S_n^+ d\mu - \int_{A_n} S_n^+ \circ T d\mu && \text{car le support de } S_n^+ \text{ est } A_n \\ &\geq \int_X S_n^+ - S_n^+ \circ T d\mu && \text{car } S_n^+ \geq 0 \\ &= 0 && \text{car la mesure } \mu \text{ est } T \text{ invariante.} \end{aligned}$$

Puisque $A = \bigcup_n A_n$, on obtient le théorème. □

Corollaire 37. *Soit $\alpha \in \mathbb{R}$. Alors*

$$\int_{\{x \in X; \sup_n \frac{s_n(x)}{n+1} > \alpha\}} f d\mu \geq \alpha \cdot \mu\{x \in X; \sup_n \frac{s_n(x)}{n+1} > \alpha\}.$$

DÉMONSTRATION. Il suffit de prendre $g = f - \alpha$ et de montrer que $\{x \in X : \sup_n (s_n(x) - (n+1)\alpha) > 0\} = \{x \in X; \sup_n \frac{s_n(x)}{n+1} > \alpha\}$. $\square$

Le théorème ergodique. Notons

$$a_n(x) = \frac{s_{n-1}(x)}{n} = \frac{1}{n} \sum_{i=0}^{n-1} f(T^i x).$$

Supposons au contraire que $a_n(x)$ ne converge pas presque partout. Il existe donc un ensemble E de mesure strictement positive tel que

$E = \{x; \limsup a_n(x) > \alpha > \beta > \liminf a_n(x)\}$, pour deux réels α, β (**Exercice**).

L'ensemble E est invariant par T ($T(E) \subset E$). En appliquant le corollaire précédent à f restreint à E , nous en déduisons :

$$\int_E f d\mu \geq \alpha \mu(E).$$

Par ailleurs, en appliquant ce résultat à $-f$, et comme $E = \{x; \limsup -a_n(x) > -\beta > -\alpha > -\liminf -a_n(x)\}$, nous obtenons également

$$\int_E f d\mu \leq \beta \mu(E).$$

Par conséquent $\alpha \mu(E) \leq \beta \mu(E)$ avec $\alpha > \beta$, ce qui n'est possible que lorsque $\mu(E) = 0$. Contradiction.

(2) Exercice.

(3) D'après le lemme de Fatou nous avons

$$\begin{aligned} \int_X \left| \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} f \circ T^n(x) \right| d\mu(x) &\leq \int_X \liminf_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} |f| \circ T^n(x) d\mu(x) \\ &\leq \liminf_{N \rightarrow \infty} \int_X \frac{1}{N} \sum_{n=0}^{N-1} |f| \circ T^n(x) d\mu(x) \\ &= \int_X |f| d\mu(x) < +\infty. \end{aligned}$$

(4) Soient $f \geq 0$ et g mesurables et bornées telles que $0 \leq g \leq f$. Alors,

$$\begin{aligned}
\left\| \frac{1}{N} \sum_{n=0}^{N-1} f \circ T^n - \bar{f} \right\|_1 &\leq \left\| \frac{1}{N} \sum_{n=0}^{N-1} f \circ T^n - \frac{1}{N} \sum_{n=0}^{N-1} g \circ T^n \right\|_1 \\
&+ \left\| \frac{1}{N} \sum_{n=0}^{N-1} g \circ T^n - \bar{g} \right\|_1 + \|\bar{f} - \bar{g}\|_1 \\
&\leq 2\|f - g\|_1 + \left\| \frac{1}{N} \sum_{n=0}^{N-1} g \circ T^n - \bar{g} \right\|_1.
\end{aligned}$$

On conclut en prenant $\|f - g\|_1$ petit et en utilisant le théorème de convergence dominée pour le deuxième terme :

$$\left\| \frac{1}{N} \sum_{n=0}^{N-1} f \circ T^n - \bar{f} \right\|_1 \xrightarrow{N \rightarrow \infty} 0.$$

(5) Par la relation (2) la fonction $\bar{f}$ est $\mathcal{I}$ -mesurable. Il suffit alors de montrer que $\int_A f d\mu = \int_A \bar{f} d\mu$ pour tout $A \in \mathcal{I}$.

Appliquons (1) à la fonction $f \cdot \mathbf{1}_A$: la suite $\frac{1}{N} \sum_{n=0}^{N-1} f \circ T^n \cdot \mathbf{1}_A \circ T^n$ converge presque sûrement vers une fonction intégrable $\overline{f \cdot \mathbf{1}_A}$.

Puisque $A \in \mathcal{I}$, $\mathbf{1}_A \circ T = \mathbf{1}_A$ μ -presque partout. Ainsi presque sûrement nous avons $\overline{f \cdot \mathbf{1}_A} = \bar{f} \cdot \mathbf{1}_A$.

Par le point (1) et le théorème de convergence dominée, nous en déduisons que

$$\begin{aligned}
\int_A \bar{f} d\mu &= \int \overline{f \cdot \mathbf{1}_A} d\mu = \lim_N \frac{1}{N} \sum_{n=0}^{N-1} \int f \circ T^n \cdot \mathbf{1}_A \circ T^n d\mu \\
&= \lim_N \frac{1}{N} \sum_{n=0}^{N-1} \int f \cdot \mathbf{1}_A d\mu \text{ puisque } T \text{ préserve la mesure} \\
&= \int_A f d\mu.
\end{aligned}$$

Pour finir, si f est à valeurs dans $\mathbb{C}$, il suffit de décomposer f en $f = f_1 + if_2$ où f_1 et f_2 sont à valeurs réelles. $\square$

Faisons deux remarques importantes pour la suite.

Si $\mathcal{I} = \{X, \phi\}$ alors $\mathbb{E}(f | \mathcal{I}) = \int_X f d\mu$.

Il est clair que dans le théorème précédent nous avons que (1) implique (2). Supposons que (2) implique (1). Prenons $A \in \mathcal{I}$, alors

$$\frac{1}{N} \sum_{n=0}^{N-1} 1_A(T^n x) = \frac{1}{N} \sum_{n=0}^{N-1} 1_{T^{-n}A}(x) = 1_A(x) \xrightarrow{N \rightarrow \infty} \mu(A) \mu - p.s.$$

ce qui implique que $\mu(A) = 0$ ou $\mu(A) = 1$. Ceci signifie, dans ce cas, que les ensembles T -invariants sont de mesure 0 ou 1 ...

Mélange et ergodicité

1. Systèmes ergodiques

Définition 38. Nous dirons que s.d.m. $(X, \mathcal{B}, \mu, T)$ est **ergodique** si

$$\mathcal{I} = \{A \in \mathcal{B} : \mu(T^{-1}(A) \Delta A) = 0\}$$

est trivial, c'est-à-dire : si $A \in \mathcal{I}$ alors $\mu(A) = 0$ ou $\mu(A) = 1$.

Lemme 39. Les d.m. $(X, \mathcal{B}, \mu, T)$ est ergodique si et seulement si $\mathbb{E}(f | \mathcal{I}) = \int_X f d\mu$ pour toute $f \in L^1$.

DÉMONSTRATION. Exercice. □

Théorème 40. Soit $(X, \mathcal{B}, \mu, T)$ un s.d.m.. Les assertions suivantes sont équivalentes :

- (1) $(X, \mathcal{B}, \mu, T)$ est ergodique ;
- (2) Toute fonction mesurable $f : X \rightarrow \mathbb{C}$ telle que $f = f \circ T$ μ -p.p. est constante μ -p.p. ;
- (3) Pour toute $f \in L^1(X, \mathcal{B}, \mu)$ nous avons

$$\frac{1}{N} \sum_{n=0}^{N-1} f(T^n(x)) \xrightarrow{N \rightarrow \infty} \int_X f d\mu.$$

DÉMONSTRATION. Posons $\mathcal{I} = \{A \in \mathcal{B} ; \mu(A \Delta T^{-1}(A)) = 0\}$. Montrons que (1) implique (2). Soient f une fonction mesurable telle que $f = f \circ T$ et $\bar{f} = \mathbb{E}(f | \mathcal{I})$.

D'après le théorème de Birkhoff nous avons $f = \bar{f}$ μ -p.p.. De plus, le s.d.m. étant ergodique nous avons $\bar{f} = \mathbb{E}(f | \mathcal{I}) = \int_X f d\mu$ d'après le lemme précédent.

Montrons que (2) implique (3). Soit $f \in L^1(X, \mathcal{B}, \mu)$. D'après le théorème ergodique de Birkhoff nous avons $\frac{1}{N} \sum_{n=0}^{N-1} f(T^n(x)) \xrightarrow{N \rightarrow \infty} \bar{f} = \mathbb{E}(f | \mathcal{I})$ et $\bar{f}$ est T -invariante. Donc $\bar{f}$ est constante et $\int_X \bar{f} d\mu = \bar{f}$. Or nous avons $\int_X \mathbb{E}(f | \mathcal{I}) d\mu = \int_X f d\mu$ (propriété de l'espérance conditionnelle).

Montrons que (3) implique (1). Soit $A \in \mathcal{I}$. Soit $f = 1_A$. Ainsi, d'après le théorème de Birkhoff et l'hypothèse nous avons $\mu(A) = \mathbb{E}(f | \mathcal{I})$ puis

$$\mu(A) = \int_A f d\mu = \int_A \mathbb{E}(f | \mathcal{I}) d\mu = \mu(A)^2.$$

Ainsi $\mu(A) = 0$ ou 1 et le s.d.m. est ergodique. □

Le corollaire suivant est une conséquence du théorème de Birkhoff et donne une version quantitative du théorème de récurrence de Poincaré.

Corollaire 41. *Soit $(X, \mathcal{B}, \mu, T)$ un s.d.m. ergodique.*

La proportion de temps passée par presque tous les points dans un ensemble $B \in \mathcal{B}$ est donné par la mesure $\mu(B)$, i.e.

$$\lim_{N \rightarrow \infty} \frac{1}{N} \text{Card}\{0 \leq n \leq N-1 : T^n x \in B\} = \mu(B)$$

pour μ presque tous les points $x \in X$.

Ainsi si $\mu(B) > 0$ presque tous les points de X passent une infinité de fois dans B .

DÉMONSTRATION. La limite s'obtient en appliquant le théorème de Birkhoff à la fonction indicatrice de B .

La seconde partie découle de la première : la limite du terme de gauche pour un point dont l'orbite passe un nombre fini de fois dans B est nulle. Ceci est faux pour μ presque tous les points de X si $\mu(B) > 0$. $\square$

Question. Existe-t-il toujours des mesures invariantes ergodiques ?

Lemme 42. *Soient X un espace métrique compact, $\mathcal{B}$ sa tribu borélienne, $T : X \rightarrow X$ une application mesurable et $\mathcal{M}$ l'ensemble des mesures de probabilité invariantes par T . Alors $\mathcal{M}$ est un ensemble convexe, compact dont les points extrémaux sont ergodiques (i.e. $\mu \in \mathcal{M}$ est ergodique si : pour tout $\mu_1, \mu_2 \in \mathcal{M}$, et $\alpha \in]0, 1[$, si $\mu = \alpha\mu_1 + (1-\alpha)\mu_2$ alors $\mu_1 = \mu_2$).*

La réciproque est également vraie, mais nous n'en avons pas besoin.

DÉMONSTRATION. Si $\mu \in \mathcal{M}$ n'est pas ergodique, il existe $B \in \mathcal{B}$ tel que $T^{-1}B = B$ et $0 < \mu(B) < 1$. Tout ensemble $A \in \mathcal{B}$ s'écrit $A = (A \cap B) \cup (A \cap B^c)$ et ainsi

$$\begin{aligned} \mu(A) &= \mu(A \cap B) + \mu(A \cap B^c) \\ &= \mu(B) \frac{\mu(A \cap B)}{\mu(B)} + \mu(B^c) \frac{\mu(A \cap B^c)}{\mu(B^c)} \\ &= \alpha \mu_1(A) + (1-\alpha) \mu_2(A), \end{aligned}$$

où $\alpha = \mu(B)$ et $\mu_1(A) = \frac{\mu(A \cap B)}{\mu(B)}$, $\mu_2(A) = \frac{\mu(A \cap B^c)}{\mu(B^c)}$.

Ceci montre que $\mu = \alpha\mu_1 + (1-\alpha)\mu_2$. $\square$

Nous allons montrer à présent l'existence de points extrémaux.

Proposition 43. *(Existence de mesures ergodiques) Soient X un espace métrique compact, $\mathcal{B}$ sa tribu borélienne, $T : X \rightarrow X$ une application mesurable. Alors il existe une mesure de probabilité ergodique T -invariante.*

DÉMONSTRATION. Soit $\{\varphi_k\}_k$ une famille de fonctions continues, dense dans $C(X)$. Puisque l'application $\mu \mapsto \int \varphi_0 d\mu$ est continue pour la topologie faible*, par compacité de $\mathcal{M}$, il existe $\nu \in \mathcal{M}$ telle que $\int \varphi_0 d\nu = \sup_{\mu \in \mathcal{M}} \int \varphi_0 d\mu$. Notons

$$\mathcal{M}_0 = \left\{ \nu \in \mathcal{M} ; \int \varphi_0 d\nu = \sup_{\mu \in \mathcal{M}} \int \varphi_0 d\mu \right\}.$$

L'ensemble $\mathcal{M}_0$ est non vide fermé. Par induction et de façon similaire, définissons pour tout entier k

$$\mathcal{M}_{k+1} = \left\{ \nu \in \mathcal{M} ; \int \varphi_0 d\nu = \sup_{\mu \in \mathcal{M}_k} \int \varphi_k d\mu \right\}.$$

Les ensembles $\mathcal{M}_k \subset \dots \subset \mathcal{M}_1 \subset \mathcal{M}_0 \subset \mathcal{M}$ sont des ensembles non vides compacts et emboîtés. Leur intersection est donc non vide, et soit $\mu \in \bigcap_{k \geq 0} \mathcal{M}_k$. Montrons à présent que μ est un point extrémal de $\mathcal{M}$.

Soient μ_1, μ_2 et $\alpha \in]0, 1[$ tels que $\mu = \alpha\mu_1 + (1 - \alpha)\mu_2$. Alors pour toute fonction continue φ , $\int \varphi d\mu = \alpha \int \varphi d\mu_1 + (1 - \alpha) \int \varphi d\mu_2$. Puisque $\mu \in \mathcal{M}_0$, $\alpha \int \varphi_0 d\mu_1 + (1 - \alpha) \int \varphi_0 d\mu_2 \geq \max\{\int \varphi_0 d\mu_1, \int \varphi_0 d\mu_2\}$. Nous en déduisons que $\int \varphi_0 d\mu_1 = \int \varphi_0 d\mu_2$.

Par récurrence sur k , nous montrons que $\int \varphi_k d\mu_1 = \int \varphi_k d\mu_2$ pour tout $k \geq 0$. Ainsi par densité des fonctions $\{\varphi_k\}_k$, nous avons $\int \varphi d\mu_1 = \int \varphi d\mu_2$ pour tout $\varphi \in C(X)$. D'après l'unicité du théorème de représentation de Riesz, nous obtenons $\mu_1 = \mu_2$. $\square$

2. Théorie spectrale

Dans ce qui suit $U_T : L^2(X, \mathcal{B}, \mu) \rightarrow L^2(X, \mathcal{B}, \mu)$ est l'opérateur défini par $U_T f = f \circ T$.

Théorème 44. *Soit $(X, \mathcal{B}, \mu, T)$ un s.d.m. ergodique.*

- (1) *Si $U_f = \lambda f$, $f \in L^2$ non-nulle, alors $|\lambda| = 1$ et $|f|$ est constante μ -p.p. ;*
- (2) *Les fonctions propres correspondant à des valeurs propres différentes sont orthogonales ;*
- (3) *Si f et g sont deux fonctions propres associées à la même valeur propre, alors il existe $c \in \mathbb{R}$ tel que $f = cg$ μ -p.p.*
- (4) *L'ensemble des valeurs propres de U_T forment un sous-groupe dénombrable de $\mathbb{T} = \{z \in \mathbb{C} : |z| = 1\}$.*
- (5) *Si $G \subseteq \mathbb{C}$ est un sous-groupe dénombrable de $\mathbb{T}$ alors il existe un s.d.m. ergodique $(X, \mathcal{B}, \mu, T)$ tel que G est son sous-groupe de valeurs propres.*

DÉMONSTRATION. Posons $L^2 = L^2(X, \mathcal{B}, \mu)$.

1) Soit $f \in L^2$ non-nulle telle que $U_f = \lambda f$. Nous avons $\|f \circ T\| = \|f\|$ car μ est T -invariante. Par conséquent $\|f\| = |\lambda| \cdot \|f\|$ et $|\lambda| = 1$. Ainsi, $|f \circ T| = |U_T(f)| = |\lambda| |f| = |f|$ et l'ergodicité implique que $|f|$ est constante μ -p.p..

2) Supposons $\lambda \neq \gamma$ non nulles, $U_T(f) = \lambda f$ et $U_T(g) = \gamma g$. Alors, μ étant T -invariante, nous avons

$$\langle f, g \rangle = \langle U_T(f), U_T(g) \rangle = \langle \lambda f, \gamma g \rangle = \lambda \bar{\gamma} \langle f, g \rangle$$

d'où $\langle f, g \rangle = 0$.

3) Soient λ une valeur propre de U_T et $f, g \in L^2$ deux fonctions propres non-nulles associées à λ . La fonction $|g|$ est non-nulle et constante μ -p.p. d'après 1) donc $g \neq 0$ μ -p.p.. Ainsi f/g est une fonction T -invariante et donc constante μ -p.p..

4) Exercice.

5) Admis. □

Corollaire 45. *Un s.d.m. $(X, \mathcal{B}, \mu, T)$ est ergodique si et seulement si toutes les valeurs propres de l'opérateur $U_T : L^2(X, \mathcal{B}, \mu) \rightarrow L^2(X, \mathcal{B}, \mu)$ défini par $U_T f = f \circ T$ sont simples.*

Proposition 46. *Un s.d.m. $(X, \mathcal{B}, \mu, T)$ est ergodique si et seulement si pour toutes $f, g \in L^2(X, \mathcal{B}, \mu)$ nous avons*

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=0}^{N-1} \langle U_T^k f, g \rangle = \langle f, 1 \rangle \cdot \langle 1, g \rangle.$$

DÉMONSTRATION. Posons $L^2 = L^2(X, \mathcal{B}, \mu)$.

Supposons le s.d.m. ergodique. Soient f et g dans L^2 . D'après le théorème ergodique de von Neumann nous avons

$$\frac{1}{N} \sum_{n=0}^{N-1} \langle U_T^n f, g \rangle = \left\langle \frac{1}{N} \sum_{n=0}^{N-1} U_T^n f, g \right\rangle \xrightarrow{N \rightarrow \infty} \langle \mathbb{E}(f | \mathcal{I}), g \rangle.$$

Or, le s.d.m. étant ergodique, nous avons $\mathbb{E}(f | \mathcal{I}) = \langle f, 1 \rangle$. Ainsi,

$$\langle \mathbb{E}(f | \mathcal{I}), g \rangle = \langle \langle f, 1 \rangle, g \rangle = \langle f, 1 \rangle \cdot \langle 1, g \rangle.$$

Réciproquement, soit $A \in \mathcal{I}$. Prenons f et g dans L^2 . D'après le théorème ergodique de von Neumann la suite $(\frac{1}{N} \sum_{n=0}^{N-1} U_T^n f; N \in \mathbb{N})$ converge vers $\mathbb{E}(f | \mathcal{I})$ dans L^2 . Donc $\left\langle \frac{1}{N} \sum_{n=0}^{N-1} U_T^n f, g \right\rangle$ converge vers $\langle \mathbb{E}(f | \mathcal{I}), g \rangle$ dans L^2 pour tout $g \in L^2$. En prenant $g = 1_A$ nous obtenons $\int_A \mathbb{E}(f | \mathcal{I}) d\mu = \int_A \left(\int_X f d\mu \right) d\mu$. Puis en prenant $f = 1_A$ on obtient $\mu(A) = \mu(A)^2$. □

Proposition 47. *Un s.d.m. $(X, \mathcal{B}, \mu, T)$ est ergodique si et seulement si pour tout $A, B \in \mathcal{B}$ nous avons*

$$\lim_{N \rightarrow +\infty} \frac{1}{N} \sum_{n=0}^{N-1} \mu(T^{-n} A \cap B) = \mu(A) \cdot \mu(B).$$

DÉMONSTRATION. Exercice. □

3. Systèmes mélangeants.

Soit $(X, \mathcal{B}, \mu, T)$ un s.d.m.. Nous avons vu que l'ergodicité est équivalente à ce que pour tout $A, B \in \mathcal{B}$ nous ayons

$$\frac{1}{N} \sum_{n=0}^{N-1} \mu(T^{-n}A \cap B) \xrightarrow{N \rightarrow \infty} \mu(A) \mu(B).$$

C'est-à-dire, $\left| \frac{1}{N} \sum_{n=0}^{N-1} \mu(T^{-n}A \cap B) - \mu(A) \cdot \mu(B) \right| \xrightarrow{N \rightarrow \infty} 0$. Or, nous avons

$$\begin{aligned} & \left| \frac{1}{N} \sum_{n=0}^{N-1} \mu(T^{-n}A \cap B) - \mu(A) \mu(B) \right| \\ & \leq \frac{1}{N} \sum_{n=0}^{N-1} \left| \mu(T^{-n}A \cap B) - \mu(A) \mu(B) \right|. \end{aligned}$$

Donc, si $|\mu(T^{-n}A \cap B) - \mu(A) \cdot \mu(B)|$ converge vers 0 alors le système est ergodique. Ceci motive les deux définitions suivantes.

Définition 48. Soit $(X, \mathcal{B}, \mu, T)$ un s.d.m..

- (1) $(X, \mathcal{B}, \mu, T)$ est **faiblement mélangeant** si pour tout $A, B \in \mathcal{B}$ nous avons

$$\frac{1}{N} \sum_{n=0}^{N-1} \left| \mu(T^{-n}A \cap B) - \mu(A) \cdot \mu(B) \right| \xrightarrow{N \rightarrow +\infty} 0.$$

- (2) $(X, \mathcal{B}, \mu, T)$ est **fortement mélangeant** si pour tout $A, B \in \mathcal{B}$ nous avons

$$\mu(T^{-n}A \cap B) \xrightarrow{n \rightarrow +\infty} \mu(A) \mu(B).$$

Évidemment, le mélange fort implique le mélange faible qui lui implique l'ergodicité. Il existe des exemples sophistiqué de s.d. qui montrent que ces notions sont toutes distinctes.

Théorème 49. Soit $(X, \mathcal{B}, \mu, T)$ un s.d.m.. Les assertions suivantes sont équivalentes.

- (1) $(X, \mathcal{B}, \mu, T)$ est faiblement mélangeant ;
- (2) $(X \times X, \mathcal{B} \times \mathcal{B}, \mu \times \mu, T \times T)$ est ergodique ;
- (3) Pour tout $A, B \in \mathcal{B}$ il existe $E \subseteq \mathbb{N}$ de densité nulle tel que

$$\lim_{n \rightarrow +\infty, n \notin E} \mu(T^{-n}A \cap B) = \mu(A) \mu(B)$$

où

$$d(E) = \lim_{n \rightarrow +\infty} \frac{|E \cap \{0 \leq j \leq n\}|}{n}$$

est la densité de E ;

(4) Pour tout $f, g \in L^2(X, \mathcal{B}, \mu)$ nous avons

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} |\langle U_T^n f, g \rangle - \langle f, 1 \rangle \langle 1, g \rangle| = 0;$$

(5) Les uniques fonctions propres de U_T sont les constantes (et ainsi 1 est l'unique valeur propre de U).

Afin de prouver ce théorème nous avons besoin du lemme suivant.

Lemme 50. Si $f : \mathbb{R} \rightarrow \mathbb{R}$ est une fonction positive ou nulle et bornée alors les assertions suivantes sont équivalentes.

(1) $\frac{1}{N} \sum_{n=0}^{N-1} f(n) \xrightarrow{N \rightarrow \infty} 0$;

(2) Il existe $E \subseteq \mathbb{N}$ de densité nulle tel que $\lim_{n \rightarrow +\infty, n \in E^c} f(n) = 0$.

DÉMONSTRATION. Il existe $K \in \mathbb{R}$ tel que $\sup_{x \in \mathbb{R}} f(x) \leq K$.
Supposons qu'il existe $E \subseteq \mathbb{N}$ de densité nulle tel que $\lim_{n \rightarrow +\infty, n \in E^c} f(n) = 0$. Alors,

$$\begin{aligned} \frac{1}{N} \sum_{n=0}^{N-1} f(n) &= \frac{1}{N} \sum_{n \leq N-1, n \in E} f(n) + \frac{1}{N} \sum_{n \leq N-1, n \in E^c} f(n) \\ &\leq K \frac{|E \cap \{n \leq N-1\}|}{N} + \frac{1}{N} \sum_{\substack{n \leq N-1 \\ n \in E^c}} f(n). \end{aligned}$$

D'où le résultat.

Supposons que nous avons (1). Posons $E_m = \{n \in \mathbb{N} : f(n) > \frac{1}{m}\}$. Ainsi,

$$\frac{|E_m \cap \{n \leq N-1\}|}{N} = \frac{1}{N} \sum_{k=0}^{N-1} 1_{E_m}(k) \leq \frac{m}{N} \sum_{k=0}^{N-1} f(k) \xrightarrow{N \rightarrow \infty} 0.$$

Par conséquent E_m est de densité nulle pour tout m .

Soit $i_0 = 0 < i_1 < i_2 < \dots$ une suite d'entiers tel que pour tout $m \geq 1$:

$$\frac{1}{N} \sum_{n=0}^{N-1} 1_{E_m}(n) \leq \frac{1}{m} \text{ pour tout } N \geq i_{m-1}.$$

Posons $E = \bigcup_{m \geq 1} E_m \cap [i_{m-1}, i_m]$ et montrons que E est de densité nulle. Soit $i_{m-1} \leq N \leq i_m$, ainsi

$$\begin{aligned}
\frac{1}{N} \sum_{n=0}^{N-1} 1_E(n) &= \frac{1}{N} \sum_{n=0}^{i_{m-1}-1} 1_E(n) + \frac{1}{N} \sum_{n=i_{m-1}}^{N-1} 1_E(n) \\
&\leq \frac{1}{N} \sum_{n=0}^{i_{m-1}-1} 1_{E_{m-1}}(n) + \frac{1}{N} \sum_{n=0}^{N-1} 1_{E_m}(n) \\
&\leq \frac{1}{N} \sum_{n=0}^{i_{m-1}-1} 1_{E_{m-1}}(n) + \frac{1}{m} \\
&\leq \frac{1}{m-1} + \frac{1}{m}.
\end{aligned}$$

Nous en déduisons que $d(E) = 0$. Remarquons que $E^c = \bigcap_{m \geq 1} E_m^c \cap [i_{m-1}, i_m]$. Ainsi, si $n \in E^c \cap [i_{m-1}, i_m]$ alors $f(n) \leq \frac{1}{m}$. Ce qui achève la preuve. $\square$

PREUVE DU THÉORÈME 49. Montrons que (1) implique (2). Il suffit de montrer que $(X \times X, \mathcal{B} \times \mathcal{B}, \mu \times \mu, T \times T)$ est faiblement mélangeant. Soient $A, B, C, D, \in \mathcal{B}$. On a

$$\begin{aligned}
&\frac{1}{N} \sum_{n=0}^{N-1} \left| \mu \times \mu \left((T \times T)^{-n} (A \times B) \cap (C \times D) \right) - \mu(A) \mu(B) \mu(C) \mu(D) \right| \\
&= \frac{1}{N} \sum_{n=0}^{N-1} \left| \mu(T^{-n}(A) \cap C) \mu(T^{-n}(B) \cap D) - \mu(A) \mu(B) \mu(C) \mu(D) \right| = (*).
\end{aligned}$$

D'après le lemme précédent, il existe $E_1, E_2 \subseteq \mathbb{N}$ de densités nulles tels que

$$\lim_{n \rightarrow +\infty, n \in E_1^c} \left| \mu(T^{-n}A \cap C) - \mu(A) \mu(C) \right| = 0$$

et

$$\lim_{n \rightarrow +\infty, n \in E_2^c} \left| \mu(T^{-n}B \cap D) - \mu(B) \mu(D) \right| = 0.$$

L'ensemble $E = E_1 \cup E_2$ est de densité nulle et

$$\mu(T^{-n}(A) \cap C) \mu(T^{-n}(B) \cap D) \xrightarrow{n \in E^c, n \rightarrow \infty} \mu(A) \mu(B) \mu(C) \mu(D).$$

Par conséquent, (*) converge vers 0. Or, $\{A \times B : A, B \in \mathcal{B}\}$ est une semi-algèbre qui engendre $\mathcal{B}$ donc $(X \times X, \mathcal{B} \times \mathcal{B}, \mu \times \mu, T \times T)$ est faiblement mélangeant (exercice).

Montrons que (2) implique (3). Soient $A, B \in \mathcal{B}$. D'après le lemme précédent, il suffit de montrer que

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

Or, $T \times T$ étant ergodique nous avons

$$\begin{aligned}
\frac{1}{N} \sum_{n=0}^{N-1} \mu(T^{-n}A \cap B) &= \frac{1}{N} \sum_{n=0}^{N-1} \mu \times \mu ((T \times T)^{-n}(A \times X) \cap (B \times X)) \\
&\xrightarrow{n \rightarrow +\infty} (\mu \times \mu)(A \times X)(\mu \times \mu)(B \times X) \\
&= \mu(A)\mu(B).
\end{aligned}$$

Montrons que (3) implique (4). D'après le lemme précédent, (4) est vrai pour les fonctions indicatrices d'ensembles mesurables. On montre aisément que c'est également vrai pour les combinaisons linéaires de fonctions indicatrices. En passant à la limite on montre que c'est vrai pour toutes les f et g dans L^2 .

Montrons que (4) implique (1). Facile.

Montrons que (1) implique (5). Soit $f \in L^2(X, \mathcal{B}, \mu)$, $f \neq 0$, $\alpha \in \mathbb{C}$ tel que $f \circ T = \alpha f$ et posons $g(x, y) = f(x)\overline{f(y)}$. Alors, $g(Tx, Ty) = \alpha f(x)\overline{\alpha f(y)} = |\alpha|^2 g(x, y)$. Or, $|\alpha| = 1$, donc $g \circ T \times T = g$. Comme $T \times T$ est ergodique, g est constante presque partout. D'où f est constante presque partout et $\alpha = 1$.

On admettra que (5) implique (1). □

4. Exercices

Exercice 6. ([KH], p. 141) Soient $(X, \mathcal{B}, \mu, T)$ un s.d.m. où $\mathcal{B}$ est la tribu borélienne associée à l'espace métrique compact X . Soit Ω un ensemble de fonctions continues dense dans $\mathcal{C}(X)$. Montrer que si pour tout f dans Ω , $(1/n) \sum_{k=0}^{n-1} f(T^k x)$ converge uniformément vers une constante, alors μ est la seule mesure invariante de $(X, \mathcal{B}, \mu, T)$.

Exercice 7. Soit $T(z) = az$ une rotation du cercle unité $X = \{e^{it}; t \in [0, 1]\}$.

1) Construire sur X une mesure λ T -invariante.

2) $(X, \mathcal{B}, \mu, T)$ est-il (fortement, faiblement) mélangeant ?

3) Montrer que $(X, \mathcal{B}, \mu, T)$ est ergodique si et seulement si a n'est pas une racine de l'unité ([Wa], p. 29).

4) Lorsque a n'est pas une racine de l'unité, montrer que T est uniquement ergodique ([KH], p. 146).

Entropie topologique

1. Définition à l'aide des recouvrements ouverts

Dans ce qui suit X est un espace topologique compact. Nous nous intéresserons aux recouvrements ouverts de X . Nous les noterons $\alpha, \beta \dots$. La remarque suivante est utile pour ce qui suit : Si $T : X \rightarrow X$ est une application continue alors $T^{-1}\alpha = \{T^{-1}A; A \in \alpha\}$ est aussi un recouvrement ouvert de X .

Définition 51. *Nous dirons que β est plus fin que α , ce que l'on écrit $\alpha \preceq \beta$, si chaque élément de β est un sous-ensemble d'un élément de α .*

Si β est un sous recouvrement de α alors $\alpha \preceq \beta$

Lorsque α et β sont deux familles de sous-ensembles de X , nous poserons

$$\alpha \vee \beta = \{A \cap B : A \in \alpha, B \in \beta\}$$

que nous appellerons **raffinement** de α et β . Ainsi $\alpha \preceq \alpha \vee \beta$.

De même nous pouvons définir le raffinement $\bigvee_{i=0}^{n-1} \alpha_i$ de n'importe quelle collection finie de familles de sous-ensembles de X .

Notons que si α et β sont deux recouvrements de X (resp. partitions de X) alors $\alpha \vee \beta$ est également un recouvrement de X (resp. une partition de X).

Définition 52. *L'entropie topologique d'un recouvrement α est le réel $H(\alpha) = \log N(\alpha)$ où $N(\alpha)$ est la plus petite cardinalité d'un sous recouvrement fini de α .*

Faisons quelques remarques.

- (1) $H(\alpha) \geq 0$.
- (2) $H(\alpha) = 0$ si et seulement si $N(\alpha) = 1$.
- (3) Si $\alpha \preceq \beta$ alors $H(\alpha) \leq H(\beta)$.
- (4) $H(\alpha \vee \beta) \leq H(\alpha) + H(\beta)$.

Preuve. Soient $\{A_1, \dots, A_{N(\alpha)}\}$ et $\{B_1, \dots, B_{N(\beta)}\}$ des sous recouvrements de, respectivement, α et β . Alors $\{A_i \cap B_j; 1 \leq i \leq N(\alpha), 1 \leq j \leq N(\beta)\}$ est un sous-recouvrement de $\alpha \vee \beta$. Ainsi $N(\alpha \vee \beta) \leq N(\alpha)N(\beta)$. $\square$

- (5) Si $T : X_1 \rightarrow X_2$ est une fonction continue alors $H(T^{-1}\alpha) \leq H(\alpha)$. Si T est aussi surjective alors $H(T^{-1}\alpha) = H(\alpha)$.

Lemme 53. Si $\{a_n\}_{n \geq 1}$ est une suite de nombres réels telle que $a_{n+p} \leq a_n + a_p$ pour tout n et p alors $\lim_{n \rightarrow \infty} a_n/n$ existe et égale à $\inf_n a_n/n$.

DÉMONSTRATION. Fixons $p > 0$. Chaque $n > 0$ peut s'écrire $n = kp + i$ avec $k \in \mathbb{Z}$ et $0 \leq i < p$. Ainsi

$$\frac{a_n}{n} = \frac{a_{i+kp}}{i+kp} \leq \frac{a_i}{kp} + \frac{a_{kp}}{kp} \leq \frac{a_i}{kp} + \frac{ka_p}{kp} = \frac{a_i}{kp} + \frac{a_p}{p}.$$

Lorsque n tend vers ∞ , k tend également vers ∞ . Ainsi

$$\limsup \frac{a_n}{n} \leq \frac{a_p}{p}$$

et par conséquent

$$\limsup \frac{a_n}{n} \leq \inf \frac{a_p}{p}.$$

Mais

$$\inf \frac{a_p}{p} \leq \liminf \frac{a_n}{n}$$

de sorte que la $\lim a_n/n$ existe et est égale à $\inf a_n/n$. $\square$

Théorème 54. Si α est un recouvrement ouvert de X et $T : X \rightarrow X$ est continue alors :

$$\lim_{n \rightarrow +\infty} \frac{1}{n} H \left(\bigvee_{i=0}^{n-1} T^{-i} \alpha \right) \text{ existe.}$$

DÉMONSTRATION. Soit $a_n = H \left(\bigvee_{i=0}^{n-1} T^{-i} \alpha \right)$. D'après le Lemme précédent, il suffit de montrer : $a_{n+k} \leq a_n + a_k$ pour $k, n \geq 1$. Or,

$$\begin{aligned} a_{n+k} &= H \left(\bigvee_{i=0}^{n+k-1} T^{-i} \alpha \right) \\ &\leq H \left(\bigvee_{i=0}^{n-1} T^{-i} \alpha \right) + H \left(\bigvee_{i=n}^{n+k-1} T^{-i} \alpha \right) \text{ (Remarque (4))} \\ &= a_n + H \left(T^{-n} \bigvee_{j=0}^{k-1} T^{-j} \alpha \right) \\ &\leq a_n + H \left(\bigvee_{j=0}^{k-1} T^{-j} \alpha \right) \text{ (Remarque (5))} \\ &= a_n + a_k. \end{aligned}$$

$\square$

Définition 55. Soit $T : X \rightarrow X$ est une application continue. L'entropie de T relativement à α est donnée par :

$$h(T, \alpha) = \lim_{n \rightarrow +\infty} \frac{1}{n} H \left(\bigvee_{i=0}^{n-1} T^{-i} \alpha \right).$$

Remarques.

(6) $h(T, \alpha) \geq 0$.

(7) $\alpha < \beta$ alors $h(T, \alpha) \leq h(T, \beta)$.

Notons que si β est un sous recouvrement fini extrait de α alors $\alpha < \beta$ et donc $h(T, \alpha) \leq h(T, \beta)$.

(8) $h(T, \alpha) \leq H(\alpha)$.

Définition 56. Si $T : X \rightarrow X$ est continue, l'entropie topologique de T est donnée par :

$$h(T) = \sup_{\alpha} h(T, \alpha).$$

où α parcourt tous les recouvrements ouverts de X .

Remarques

(9) $h(T) \geq 0$.

(10) Dans la définition de $h(T)$, on peut prendre la borne supérieure sur tous les recouvrements ouverts finis de X (d'après (7)).

(11) $h(I) = 0$ où I est la fonction identité de X .

(12) Si Y est un sous ensemble fermé de X et $TY = Y$ alors $h(T|_Y) \leq h(T)$.

Théorème 57. Soient X_1, X_2 des espaces compacts, $T_i : X_i \rightarrow X_i$, $i \in \{1, 2\}$, des applications continues, et $\phi : X_1 \rightarrow X_2$ une application continue avec $\phi(X_1) = X_2$ et $\phi \circ T_1 = T_2 \circ \phi$. Alors $h(T_1) \geq h(T_2)$.

De plus, si ϕ est un homéomorphisme alors $h(T_1) = h(T_2)$.

DÉMONSTRATION. Soit α un recouvrement ouvert de X_2 . Pour $n \geq 1$, grâce à la Remarque (5), puisque ϕ est surjective, nous savons que :

$$\begin{aligned} H \left(\bigvee_{i=0}^{n-1} T_2^{-i} \alpha \right) &= H \left(\phi^{-1} \bigvee_{i=0}^{n-1} T_2^{-i} \alpha \right) \\ &= H \left(\bigvee_{i=0}^{n-1} (\phi^{-1} T_2^{-i} \alpha) \right) = H \left(\bigvee_{i=0}^{n-1} T_1^{-i} (\phi^{-1} \alpha) \right) \end{aligned}$$

En particulier

$$h(T_2, \alpha) = \lim_n \frac{1}{n} H \left(\bigvee_{i=0}^{n-1} T_2^{-i} \alpha \right) = \lim_n \frac{1}{n} H \left(\bigvee_{i=0}^{n-1} T_1^{-i} (\phi^{-1} \alpha) \right) = h(T_1, \phi^{-1} \alpha)$$

Ce qui donne ainsi

$$h(T_2) = \sup_{\alpha} h(T_2, \alpha) = \sup_{\alpha} h(T_1, \phi^{-1}\alpha) \leq \sup_{\beta} h(T_1, \beta) = h(T_1).$$

D'où $h(T_2) \leq h(T_1)$.

Si ϕ est un homéomorphisme alors $\phi^{-1}T_2 = T_1\phi^{-1}$. Ainsi, par ce qui précède $h(T_1) \leq h(T_2)$.
D'où $h(T_2) = h(T_1)$. $\square$

Dans la prochaine partie, nous donnerons une définition de $h(T)$ qui n'exige pas de X d'être compact et nous prouverons des propriétés de $h(T)$ dans un contexte plus général. Cependant, un résultat qui est faux quand X n'est pas compact est le suivant :

Théorème 58. *Si $T : X \rightarrow X$ est un homéomorphisme d'un espace compact X alors $h(T) = h(T^{-1})$.*

DÉMONSTRATION. Grâce à la Remarque (5) nous savons que :

$$H\left(\bigvee_{i=0}^{n-1} T^i \alpha\right) = H\left(T^{-(n-1)} \bigvee_{i=0}^{n-1} T^i \alpha\right) = H\left(\bigvee_{j=0}^{n-1} T^{-j} \alpha\right).$$

En particulier :

$$h(T, \alpha) = \lim_n \frac{1}{n} H\left(\bigvee_{j=0}^{n-1} T^{-j} \alpha\right) = \lim_n \frac{1}{n} H\left(\bigvee_{i=0}^{n-1} T^i \alpha\right) = h(T^{-1}, \alpha).$$

D'où $h(T) = h(T^{-1})$. $\square$

2. Définition de Bowen

Dans cette partie, nous définissons l'entropie pour des espaces non nécessairement compacts, mais métriques.

Dans ce paragraphe, (X, d) est un espace métrique (pas nécessairement compact). La boule ouverte de centre x et de rayon r sera notée $B(x, r)$ et la boule fermée par $\bar{B}(x, r)$. Nous définirons l'entropie topologique pour les fonctions uniformément continues $T : X \rightarrow X$. Nous noterons $UC(X, d)$ l'espace des fonctions uniformément continues de X dans $\mathbb{R}$. Dans toute cette section T sera un élément de $UC(X, d)$.

Si n est un nombre naturel, nous pouvons définir une nouvelle métrique d_n sur X par : $d_n(x, y) = \max_{0 \leq i \leq n-1} d(T^i(x), T^i(y))$ (cette notation ne montre pas la dépendance par rapport à T). La boule ouverte de centre x et de rayon r dans la métrique d_n est : $B_{d_n}(x, r) = \bigcap_{i=0}^{n-1} T^{-i} B(T^i x, r)$.

Définition 59. *Soient $n \in \mathbb{N}$, $\varepsilon > 0$ et K un sous-ensemble compact de X . On dit qu'un sous-ensemble F de X est (n, ε) -engendre K par T , si pour tout $x \in K$, il existe $y \in F$ tel*

que $d_n(x, y) \leq \varepsilon$, autrement dit :

$$K \subset \bigcup_{y \in F} \overline{B}_{d_n}(y, \varepsilon) = \bigcap_{i=0}^{n-1} T^{-i} \overline{B}(T^i y, \varepsilon).$$

Définition 60. Soient $n \in \mathbb{N}$, $\varepsilon \geq 0$ et K un sous-ensemble compact de X . Nous noterons

$$r_n(\varepsilon, K) = \inf\{\#F; F \text{ } (n, \varepsilon)\text{-engendre } K\}.$$

Remarques.

(1) Evidemment $r_n(\varepsilon, K) < \infty$ car K est compact.

(2) Si $\varepsilon_1 < \varepsilon_2$ alors $r_n(\varepsilon_1, K) \geq r_n(\varepsilon_2, K)$.

Définition 61. Soient $\varepsilon > 0$ et K un sous-ensemble compact de X . Nous notons

$$r(\varepsilon, K, T) = \limsup_{n \rightarrow \infty} \frac{1}{n} \log r_n(\varepsilon, K).$$

Nous écrivons $r(\varepsilon, K, T, d)$ si nous souhaitons souligner la distance d .

Remarque.

(3) Si $\varepsilon_1 < \varepsilon_2$ alors $r(\varepsilon_1, K, T) \geq r(\varepsilon_2, K, T)$ (Remarque (2)).

Définition 62. Lorsque K est un sous ensemble compact de X , définissons $h_d(T, K) = \lim_{\varepsilon \rightarrow 0} r(\varepsilon, K, T)$. L'entropie topologique de T est

$$h_d(T) = \sup_K h_d(T, K)$$

où K parcourt tous les sous ensembles compacts de X .

Définition 63. Soient $n \in \mathbb{N}$, $\varepsilon > 0$ et K est un sous-ensemble compact de X . Un sous-ensemble E de K est (n, ε) -séparé par T si pour tout $x, y \in E$, $x \neq y$, nous avons $d_n(x, y) > \varepsilon$. Autrement dit, pour tout $x \in E$, l'ensemble $\bigcap_{i=0}^{n-1} T^{-i} B(T^i x; \varepsilon)$ ne contient aucun autre point de E .

Définition 64. Soient $n \in \mathbb{N}$ est un nombre naturel, $\varepsilon > 0$ et K un sous-ensemble compact de X , définissons

$$s_n(\varepsilon, K) = \max\{\#E; E \subset K \text{ est } (n, \varepsilon)\text{-séparé par } T\}.$$

(Nous écrivons $s_n(\varepsilon, K, T)$ pour souligner la dépendance avec T si nous en avons besoin).

Remarques.

(4) Nous avons $r_n(\varepsilon, K) \leq s_n(\varepsilon, K) \leq r_n(\varepsilon/2, K)$ et d'où $s_n(\varepsilon, K) < \infty$.

Preuve. Si $E \subset K$ est un sous-ensemble (n, ε) -séparé de cardinalité maximale, alors E est un ensemble qui (n, ε) -engendre K (Exercice). Par conséquent $r_n(\varepsilon, K) \leq s_n(\varepsilon, K)$.

Pour montrer l'autre inégalité, supposons que $E \subset K$ est (n, ε) -séparé pour K et que F est un ensemble qui $(n, \varepsilon/2)$ -engendre K . Définissons $\phi : E \rightarrow F$, en choisissant, pour chaque $x \in E$, un certain point $\phi(x) \in F$ avec $d_n(x, \phi(x)) \leq \varepsilon/2$. Alors ϕ est injective et

par conséquent la cardinalité de E n'est pas plus grande que celle de F . D'où $s_n(\varepsilon, K) \leq r_n(\varepsilon/2, K)$. $\square$

(5) Si $\varepsilon_1 < \varepsilon_2$ alors $s_n(\varepsilon_1, K) \geq s_n(\varepsilon_2, K)$.

Définition 65. Lorsque $\varepsilon > 0$ et K est un sous ensemble compact de X , nous notons :

$$s(\varepsilon, K, T) = \lim_{n \rightarrow \infty} \sup \frac{1}{n} \log s_n(\varepsilon, K).$$

Nous écrirons parfois $s(\varepsilon, K, T, d)$ pour souligner la métrique d .

Remarques.

(6) Nous avons $r(\varepsilon, K, T) \leq s(\varepsilon, K, T) \leq r(\varepsilon/2, K, T)$ selon la remarque (4).

(7) Si $\varepsilon_1 < \varepsilon_2$ alors $s(\varepsilon_1, K, T) \geq s(\varepsilon_2, K, T)$.

(8) Nous avons $h_d(T, K) = \lim_{\varepsilon \rightarrow 0} s(\varepsilon, K, T)$, (Remarque (6)), de sorte que :

$$h_d(T) = \sup_K \lim_{\varepsilon \rightarrow 0} s(\varepsilon, K, T).$$

Ainsi $h_d(T) = \sup_K \lim_{\varepsilon \rightarrow 0} r(\varepsilon, K, T) = \sup_K \lim_{\varepsilon \rightarrow 0} s(\varepsilon, K, T)$ peut-être définie en utilisant les ensembles qui engendrent ou séparent.

(9) Si T est une isométrie de (X, d) , alors $d_n = d$ pour tout $n \in \mathbb{N}$ de sorte que $s_n(\varepsilon, K) = s_1(\varepsilon, K)$ et $h_d(T) = 0$. Par exemple, $h_d(R_\alpha) = 0$ pour toute rotation R_α sur le tore $\mathbb{T}^1$.

Définition 66. Deux métriques d et d' sur X sont **uniformément équivalentes** si l'identité de (X, d) dans (X, d') l'identité de (X, d') dans (X, d) sont toutes les deux uniformément continues.

Dans ce cas-ci, $T \in UC(X, d)$ si et seulement si $T \in UC(X, d')$.

Théorème 67. Si d et d' sont uniformément équivalentes et $T \in UC(X, d)$ alors $h_d(T) = h_{d'}(T)$.

DÉMONSTRATION. Soit $\varepsilon_1 > 0$. Choisissons $\varepsilon_2 > 0$ tel que $d'(x, y) < \varepsilon_2$ implique $d(x, y) < \varepsilon_1$, puis choisissons $\varepsilon_3 > 0$ tel que $d(x, y) < \varepsilon_3$ implique $d'(x, y) < \varepsilon_2$. Soit K un compact. Alors (Exercice),

$$r_n(\varepsilon_1, K, d) \leq r_n(\varepsilon_2, K, d') \text{ et } r_n(\varepsilon_2, K, d') \leq r_n(\varepsilon_3, K, d).$$

D'où $r(\varepsilon_1, K, T, d) \leq r(\varepsilon_2, K, T, d') \leq r(\varepsilon_3, K, T, d)$. Lorsque $\varepsilon_1, \varepsilon_2$ et ε_3 tendent vers 0, nous obtenons $h_d(T, K) = h_{d'}(T, K)$. $\square$

Remarques.

(12) Présentons un exemple de deux distances équivalentes, mais pas uniformément équivalentes, qui vont donner différentes valeurs d'entropie pour une certaine transformation.

Soit $X =]0, \infty[$. Définissons $T :]0, \infty[\rightarrow]0, \infty[$ par $T(x) = 2x$. Considérons d la métrique euclidienne sur $]0, \infty[$. Alors $T \in UC(x, d)$ et on peut montrer facilement que $h_d(T) \geq \log(2)$ en estimant la valeur de $r_n(\varepsilon, [1, 2])$ (Exercice).

Soit d' la distance qui coïncide avec d sur $[1, 2]$ mais qui fait de T une isométrie pour d' (utiliser les intervalles $]2^{n-1}, 2^n]$). Alors $h_{d'} = 0$ (Remarque (10)).

(13) Si X est compact et si d et d' sont des métriques équivalentes alors elles sont uniformément équivalentes. En outre, chaque fonction continue $T : X \rightarrow X$ est uniformément continue. Par conséquent, si X est un espace compact métrisable, l'entropie de T ne dépend pas de la métrique choisie sur X (à condition que la métrique induise la topologie de X).

Ce qui suit sera utile plus tard et il nous permettra de simplifier la définition de $h_d(T)$ quand X est compact.

Théorème 68. *Soient (X, d) un espace métrique et $T \in UC(X, d)$. Si $K \subset K_1 \cup \dots \cup K_m$, où les K_i sont des sous-ensembles compacts de X , alors $h_d(T, K) \leq \max_{1 \leq i \leq m} h_d(T, K_i)$.*

DÉMONSTRATION. Remarquons que nous avons $s_n(\varepsilon, K) \leq s_n(\varepsilon, K_1) + \dots + s_n(\varepsilon, K_m)$ (Exercice). Pour chaque n , choisissons un compact $K_{i(n, \varepsilon)}$ tel que $s_n(\varepsilon, K_{i(n, \varepsilon)}) = \max_j s_n(\varepsilon, K_j)$. Ainsi, $s_n(\varepsilon, K) \leq m s_n(\varepsilon, K_{i(n, \varepsilon)})$, et ainsi, $\log s_n(\varepsilon, K) \leq \log m + \log s_n(\varepsilon, K_{i(n, \varepsilon)})$.

Choisissons $n_j \rightarrow \infty$ tel que $\frac{1}{n_j} \log s_{n_j}(\varepsilon, K) \rightarrow \limsup_{n \rightarrow \infty} \frac{1}{n} \log s_n(\varepsilon, K)$. Comme il n'y a qu'un nombre fini de compacts K_i , quitte à considérer une sous-suite, on peut considérer que $K_{i(n_j, \varepsilon)}$ ne dépend pas de j (i.e $K_{i(n_j, \varepsilon)} = K_{i(\varepsilon)}$ pour tout j). Par conséquent $s(\varepsilon, K, T) \leq s(\varepsilon, K_{i(\varepsilon)}, T)$.

Choisissons $\varepsilon_q \rightarrow 0$ tel que $K_{i(\varepsilon_q)}$ est constant ($= K_{i_0}$). Ainsi, $h_d(T, K) \leq h_d(T, K_{i_0}) \leq \max_j h_d(T, K_j)$. $\square$

Corollaire 69. *Soient (X, d) un espace métrique et $T \in UC(X, d)$. Soit $\delta > 0$. Alors,*

$$h_d(T) = \sup_K h_d(T, K)$$

où K parcourt les sous-ensembles compacts de diamètre inférieur à δ .

DÉMONSTRATION. Si K est compact, il peut être couvert par un nombre fini de boules $B_1, \dots, B_m$ de diamètre $\delta/2$. Ainsi, $h_d(T, K) \leq h_d(T, K \cap \bar{B}_i) \leq h_d(T)$. $\square$

Corollaire 70. *Si (X, d) est un espace compact métrique et d' est une distance équivalente à d alors*

$$h_d(T) = h_{d'}(T) = h_d(T, X).$$

DÉMONSTRATION. La première égalité correspond à la Remarque (13). La seconde de l'inégalité $h_d(T, K) \leq h_d(T, X)$ pour tout compact K . $\square$

Pour résumer, lorsque X est compact, nous pouvons employer le Corollaire 70 pour simplifier la définition de $h(T)$. Prenons n'importe quelle métrique d engendrant la topologie de X . Alors

$$h(T) = \lim_{\varepsilon \rightarrow 0} \limsup_{n \rightarrow \infty} \frac{1}{n} \log r_n(\varepsilon, X) = \lim_{\varepsilon \rightarrow 0} \limsup_{n \rightarrow \infty} \frac{1}{n} \log s_n(\varepsilon, X).$$

3. Egalité des deux définitions

Nous allons montrer que les définitions des entropies données dans les deux sections précédentes coïncident, lorsque T est une fonction continue d'un espace compact métrisable. A cette fin, notons $h^*(T)$ et $h^*(T, \alpha)$ les nombres apparaissant dans la définition de l'entropie topologique à l'aide des recouvrements ouverts.

Dans un espace métrique (X, d) , nous définissons le diamètre d'un recouvrement par $\text{diam}(\alpha) = \sup_{A \in \alpha} \text{diam}(A)$, où $\text{diam}(A)$ détermine le diamètre de l'ensemble A .

Pour commencer, rappelons ce qu'est un nombre de Lebesgue.

Théorème 71. *Soient X un espace métrique compact et α un recouvrement ouvert de X . Il existe $\rho > 0$ tel que pour tout $x \in X$, il existe $A \in \alpha$ vérifiant $B(x, \rho) \subset A$. De tels nombres ρ sont appelés **nombre de Lebesgue**.*

Si α et γ sont des recouvrements ouverts de X et $\text{diam}(\alpha)$ est inférieur à un nombre de Lebesgue pour γ alors $\gamma \preceq \alpha$.

Le résultat suivant est souvent utile pour calculer $h^*(T)$.

Théorème 72. *Soient (X, d) un espace métrique compact et $(\alpha_n)_{1 \leq n < \infty}$ une suite de recouvrements ouverts de X telle que $\text{diam}(\alpha_n) \rightarrow 0$.*

Si $h^*(T) < \infty$ alors $\lim_{n \rightarrow \infty} h^*(T, \alpha_n)$ existe et est égal à $h^*(T)$.

Si $h^*(T) = \infty$ alors $\lim_{n \rightarrow \infty} h^*(T, \alpha_n) = \infty$.

DÉMONSTRATION. Supposons $h^*(T) < \infty$. Soit $\varepsilon > 0$. Choisissons un recouvrement ouvert γ tel que $h^*(T, \gamma) > h^*(T) - \varepsilon$. Soit δ un nombre de Lebesgue pour γ .

Choisissons N de sorte que $n \geq N$ implique $\text{diam}(\alpha_n) < \delta$. Alors $\gamma \preceq \alpha_n$, et ainsi $h^*(T, \gamma) < h^*(T, \alpha_n)$ quand $n \geq N$. Par conséquent, lorsque $n \geq N$ nous avons $h^*(T) \geq h^*(T, \alpha_n) > h^*(T) - \varepsilon$. Finalement, $\lim_{n \rightarrow \infty} h^*(T, \alpha_n) = h^*(T)$.

Supposons $h^*(T) = \infty$. Soient $a > 0$ et γ un recouvrement ouvert γ tel que $h^*(T, \gamma) > a$. En procédant comme ci-dessus on obtient $\lim_{n \rightarrow \infty} h^*(T, \alpha_n) = \infty$. $\square$

Corollaire 73. *Nous avons $h^*(T) = \lim_{\delta \rightarrow 0} \sup\{h^*(T, \alpha); \text{diam}(\alpha) < \delta\}$.*

Le prochain résultat donne les relations de base entre les deux manières de définir l'entropie topologique.

Théorème 74. *Soit $T : X \rightarrow X$ une fonction continue d'un espace métrique compact (X, d) .*

(1) *Si α est un recouvrement ouvert de X avec le nombre de Lebesgue δ alors :*

$$N \left(\bigvee_{i=0}^{n-1} T^{-i} \alpha \right) \leq r_n(\delta/2, X) \leq s_n(\delta/2, X).$$

(2) *Si $\varepsilon > 0$ et γ est un recouvrement ouvert tels que $\text{diam}(\gamma) \leq \varepsilon$ alors :*

$$r_n(\varepsilon, X) \leq s_n(\varepsilon, X) \leq N \left(\bigvee_{i=0}^{n-1} T^{-i} \gamma \right).$$

DÉMONSTRATION. D'après la Remarque (4) de la section précédente $r_n(\varepsilon, K) \leq s_n(\varepsilon, K)$ pour tout $\varepsilon > 0$.

Montrons (1). Soit F un ensemble qui $(n, \delta/2)$ -engendre X de cardinalité $r_n(\delta/2, X)$. Ainsi,

$$X = \cup_{x \in F} \cap_{i=0}^{n-1} T^{-i} \bar{B}(T^i x; \delta/2)$$

et puisque pour chaque i , $\bar{B}(T^i x; \delta/2)$ est inclus dans un élément de α , nous avons $N\left(\bigvee_{i=0}^{n-1} T^{-i} \alpha\right) \leq r_n(\delta/2, X)$.

Montrons (2). Soit E un ensemble (n, ε) -séparé de cardinalité $s_n(\varepsilon, X)$. Tous les éléments du recouvrement $\bigvee_{i=0}^{n-1} T^{-i} \gamma$ contiennent au plus un élément de E . Par conséquent, $s_n(\varepsilon, X) \leq N\left(\bigvee_{i=0}^{n-1} T^{-i} \gamma\right)$. $\square$

Corollaire 75. *Soit $T : X \rightarrow X$ une fonction continue d'un espace métrique compact (X, d) . Soit $\varepsilon > 0$. Considérons α_ε le recouvrement de X par toutes les boules ouvertes de rayon 2ε et γ_ε n'importe quel recouvrement de X par des boules ouvertes de rayon $\varepsilon/2$. Alors*

$$N\left(\bigvee_{i=0}^{n-1} T^{-i} \alpha_\varepsilon\right) \leq r_n(\varepsilon, X) \leq s_n(\varepsilon, X) \leq N\left(\bigvee_{i=0}^{n-1} T^{-i} \gamma_\varepsilon\right).$$

Théorème 76. *Si $T : X \rightarrow X$ est une fonction continue d'un espace métrique compact (X, d) , alors $h(T) = h^*(T)$ i.e les deux définitions de l'entropie topologique coïncident.*

DÉMONSTRATION. Soient $\varepsilon > 0$ et $\alpha_\varepsilon, \gamma_\varepsilon$ comme dans le Corollaire 75. Alors, $h^*(T, \alpha_\varepsilon) \leq r(\varepsilon, X, T) \leq s(\varepsilon, X, T) \leq h^*(T, \gamma_\varepsilon)$. En faisant tendre ε vers 0, le Théorème 72 implique que les deux termes extrémaux convergent vers $h^*(T)$, et par ailleurs les termes centraux tendent vers $h(T)$. Ainsi, $h^*(T) = h(T)$. $\square$

Le théorème suivant montre que l'on peut remplacer $\limsup$ par $\liminf$ dans la définition de l'entropie.

Théorème 77. *Si $T : X \rightarrow X$ est une fonction continue d'un espace métrique compact (X, d) alors*

$$h(T) = \lim_{\varepsilon \rightarrow 0} \liminf_{n \rightarrow \infty} \frac{1}{n} \log r_n(\varepsilon, X) = \lim_{\varepsilon \rightarrow 0} \liminf_{n \rightarrow \infty} \frac{1}{n} \log s_n(\varepsilon, X).$$

DÉMONSTRATION. Le Corollaire 75 donne

$$h^*(T, \alpha_\varepsilon) \leq \liminf_{n \rightarrow \infty} \frac{1}{n} \log r_n(\varepsilon, X) \leq \liminf_{n \rightarrow \infty} \frac{1}{n} \log s_n(\varepsilon, X) \leq h^*(T, \gamma_\varepsilon)$$

En faisant tendre ε vers 0, puis en utilisant le Théorème 76. $\square$

Théorème 78. (1) *Si (X, d) est un espace métrique, $T \in UC(X, d)$ et $m > 0$ alors $h_d(T^m) = m h_d(T)$.*

- (2) Soient (X_i, d_i) , $i = 1, 2$, deux espaces métriques et $T_i \in UC(X_i, d_i)$. Définissons une distance sur $X_1 \times X_2$ par $d((x_1, x_2), (y_1, y_2)) = \max\{d_1(x_1, y_1), d_2(x_2, y_2)\}$. Alors, $T_1 \times T_2 \in UC(X_1 \times X_2, d)$ et $h_d(T_1 \times T_2) \leq h_{d_1}(T_1) + h_{d_2}(T_2)$. De plus, si X_1 ou X_2 est compact alors $h_d(T_1 \times T_2) = h_{d_1}(T_1) + h_{d_2}(T_2)$.

DÉMONSTRATION. Montrons (1). Pour commencer remarquons que nous avons $r_n(\varepsilon, K, T^m) \leq r_{mn}(\varepsilon, K, T)$. Ainsi,

$$\frac{1}{n} \log r_n(\varepsilon, K, T^m) \leq \frac{m}{mn} \log r_{mn}(\varepsilon, K, T)$$

et par conséquent $h_d(T^m) \leq mh_d(T)$.

Puisque T est uniformément continue, pour tout $\varepsilon > 0$, il existe $\delta > 0$ tel que $d(x, y) < \delta$ implique $d_n(x, y) < \varepsilon$.

Ainsi, un ensemble qui (n, δ) -engendre K par T^m est également un ensemble qui (mn, ε) -engendre K par T . Par conséquent, $r(\delta, K, T^m) \geq mr(\varepsilon, K, T)$. Et finalement, $mh_d(T, K) \leq h_d(T^m, K)$, d'où le résultat.

Montrons (2) Soient $K_i \subseteq X_i$, $i = 1, 2$ deux compacts. Si F_i (n, ε) -engendre K_i par T_i , alors $F_1 \times F_2$ (n, ε) -engendre $K_1 \times K_2$ par $T_1 \times T_2$. D'où,

$$r_n(\varepsilon, K_1 \times K_2, T_1 \times T_2) \leq r_n(\varepsilon, K_1, T_1) \cdot r_n(\varepsilon, K_2, T_2).$$

Ce qui implique

$$r(\varepsilon, K_1 \times K_2, T_1 \times T_2) \leq r(\varepsilon, K_1, T_1) + r(\varepsilon, K_2, T_2).$$

Par conséquent

$$h_d(T_1 \times T_2, K_1 \times K_2) \leq h_{d_1}(T_1, K_1) + h_{d_2}(T_2, K_2).$$

Soient $\pi_i : X_1 \times X_2 \rightarrow X_i$, $i = 1, 2$, les projections canoniques.

Si $K \subseteq X_1 \times X_2$ est compact, alors $K_1 = \pi_1(K)$ et $K_2 = \pi_2(K)$ sont compacts et $K \subseteq K_1 \times K_2$. D'où

$$h_d(T_1 \times T_2, K) \leq h_d(T_1 \times T_2, K_1 \times K_2).$$

Par conséquent,

$$\begin{aligned} h_d(T_1 \times T_2) &= \sup_{\substack{K \subseteq X_1 \times X_2 \\ \text{compact}}} h_d(T_1 \times T_2, K) \\ &= \sup_{\substack{K_1 \subseteq X_1 \\ K_2 \subseteq X_2 \\ \text{compact}}} h_d(T_1 \times T_2, K_1 \times K_2) \\ &\leq \sup_{\substack{K_1 \subseteq X_1 \\ \text{compact}}} h_{d_1}(T_1, K_1) + \sup_{\substack{K_2 \subseteq X_2 \\ \text{compact}}} h_{d_2}(T_2, K_2) \\ &= h_{d_1}(T_1) + h_{d_2}(T_2). \end{aligned}$$

Nous supposons maintenant que X_1 est compact. Tout sous-ensemble compact de $X_1 \times X_2$ est un sous-ensemble de $X_1 \times K_2$ pour un certain sous-ensemble compact $K_2 \subset X_2$. Par conséquent, $h_d(T_1 \times T_2) = \sup\{h_d(T_1 \times T_2, X_1 \times K_2); K_2 \text{ compact}\}$.

Soit K_2 un sous-ensemble compact de X_2 . Soit $\delta > 0$. Soient E_1 un sous-ensemble qui (n, δ) -sépare X_1 et E_2 un sous-ensemble qui (n, δ) -sépare K_2 . Alors, $E_1 \times E_2$ est un sous-ensemble qui (n, δ) -sépare $X_1 \times K_2$. Ainsi,

$$s_n(\delta, X_1 \times K_2, T_1 \times T_2) \geq s_n(\delta, X_1, T_1) \cdot s_n(\delta, K_2, T_2)$$

et ainsi,

$$\begin{aligned} s(\delta, X_1 \times K_2, T_1 \times T_2) &\geq \limsup_{n \rightarrow \infty} \frac{1}{n} [\log s_n(\delta, X_1, T_1) + \log s_n(\delta, K_2, T_2)] \\ &\geq \liminf_{n \rightarrow \infty} \frac{1}{n} \log s_n(\delta, X_1, T_1) + \limsup_{n \rightarrow \infty} \frac{1}{n} \log s_n(\delta, K_2, T_2). \end{aligned}$$

En faisant tendre δ vers 0 et en utilisant le Théorème 77 on obtient

$$h_d(T_1 \times T_2, X_1 \times K_2) \geq h_{d_1}(T_1) + h_{d_2}(T_2, K_2)$$

puis

$$h_d(T_1 \times T_2) \geq \sup_{K_2} h_d(T_1 \times T_2, X_1 \times K_2) \geq h_{d_1}(T_1) + h_{d_2}(T_2).$$

□

4. Calcul de l'entropie topologique

Dans cette section nous allons faire des calculs d'entropie explicites pour des systèmes classiques.

4.1. Générateurs et systèmes expansifs. Nous allons énoncer et prouver un analogue du Théorème de Kolmogorov-Sinaï. Ceci permettra de calculer l'entropie de certains systèmes expansifs.

Définition 79. *Un homéomorphisme T d'un espace métrique compact est **expansif** s'il existe $\delta > 0$ tel que, si $x \neq y$ alors il existe $n \in \mathbb{Z}$ avec $d(T^n x, T^n y) > \delta$. Nous dirons que δ est une constante d'expansivité pour T .*

Exemple : le décalage est expansif (Exercice).

Définition 80. *Soient X un espace métrique compact et $T : X \rightarrow X$ un homéomorphisme. Un recouvrement fini α d'ouverts de X est un **générateur** pour T si pour toute suite $(A_n)_{n \in \mathbb{Z}}$ d'éléments de α , l'ensemble*

$$\bigcap_{n \in \mathbb{Z}} T^{-n} \bar{A}_n$$

contient au plus un point.

Proposition 81. *Soient T un homéomorphisme d'un espace métrique compact (X, d) et α un générateur pour T . Alors, pour tout $\varepsilon > 0$, il existe $N > 0$ tel que chaque élément de $\bigvee_{-N}^N T^{-n}\alpha$ a un diamètre inférieur à ε . Réciproquement, pour tout $N > 0$, il existe $\varepsilon > 0$ tel que $d(x, y) < \varepsilon$ implique $x, y \in \bigcap_{-N}^N T^{-n}A_n$ pour certains $A_{-N}, \dots, A_N \in \alpha$.*

DÉMONSTRATION. Procédons par l'absurde pour le premier énoncé : Il existe $\varepsilon > 0$ tel que pour tout $j > 0$, il existe x_j, y_j avec $d(x_j, y_j) > \varepsilon$ et il existe $A_{j,i} \in \alpha$, $-j \leq i \leq j$, avec $x_j, y_j \in \bigcap_{i=-j}^j T^{-i}A_{j,i}$. Par compacité, il existe une sous-suite $(j_k)_k$ d'entiers positifs telle que $x_{j_k} \rightarrow x$ et $y_{j_k} \rightarrow y$. Nous avons évidemment $x \neq y$. Considérons les ensembles $A_{j_k,0}$. Une infinité d'entre eux sont égaux puisque α est fini. Ainsi, $x_{j_k}, y_{j_k} \in A_0$, pour une infinité de k . D'où $x, y \in \bar{A}_0$. De la même façon nous montrons que $x, y \in T^{-n}\bar{A}_n$ pour tout $n \in \mathbb{Z}$.

Ainsi, $x, y \in \bigcap_{-\infty}^{\infty} T^{-n}\bar{A}_n$. Ce qui contredit le fait que α est un générateur.

Montrons le second énoncé. Soit $N > 0$. Soit $\delta > 0$ un nombre de Lebesgue pour α . Choisissons $\varepsilon > 0$ tel que $d(x, y) < \varepsilon$ implique que $d(T^i x, T^i y) < \delta$ pour $-N \leq i \leq N$. Ainsi, si $d(x, y) < \varepsilon$ et $|i| \leq N$ alors $T^i x, T^i y \in A_i$ pour un certain $A_i \in \alpha$. Ce qui donne $x, y \in \bigcap_{-N \leq i \leq N} T^{-i}A_i$. $\square$

Théorème 82. *Soit $T : X \rightarrow X$ un homéomorphisme expansif d'un espace métrique compact (X, d) .*

- (1) *Si α est un générateur pour T alors $h(T) = h(T, \alpha)$.*
- (2) *Si δ est une constante d'expansivité pour T , alors $h(T) = r(\delta_0, T) = s(\delta_0, T)$ pour tout $\delta_0 < \delta/4$.*

DÉMONSTRATION. Montrons (1). Soient β un recouvrement ouvert et δ un nombre de Lebesgue pour β .

D'après la Proposition 81, il existe $N > 0$ tel que chaque élément de $\bigvee_{n=-N}^N T^{-n}\alpha$ a un diamètre inférieur à δ . D'où $\beta \preceq \bigvee_{n=-N}^N T^{-n}\alpha$, et ainsi

$$\begin{aligned} h(T, \beta) &\leq h\left(T, \bigvee_{n=-N}^N T^{-n}\alpha\right) = \lim_{k \rightarrow \infty} \frac{1}{k} H\left(\bigvee_{i=0}^{k-1} T^{-i}\left(\bigvee_{n=-N}^N T^{-n}\alpha\right)\right) \\ &= \lim_{k \rightarrow \infty} \frac{1}{k} H\left(\bigvee_{n=-N}^{N+k-1} T^{-n}\alpha\right) = \lim_{k \rightarrow \infty} \frac{1}{k} H\left(\bigvee_{n=0}^{2N+k-1} T^{-n}\alpha\right) \\ &= \lim_{k \rightarrow \infty} \frac{2N+k-1}{k} \frac{1}{2N+k-1} H\left(\bigvee_{n=0}^{2N+k-1} T^{-n}\alpha\right) = h(T, \alpha). \end{aligned}$$

Par conséquent $h(T, \beta) \leq h(T, \alpha)$ pour tout recouvrement ouvert β , d'où le résultat.

Montrons (2). Soit $\delta_0 < \delta/4$. Choisissons des points $x_1, \dots, x_k$ tels que $X = \bigcup_{i=1}^k B(x_i, (\delta/2) - 2\delta_0)$. Il est facile de vérifier que le recouvrement $\alpha = \{B(x_i, \delta/2); 1 \leq i \leq k\}$ est un générateur pour t et qu'il a $2\delta_0$ pour nombre de Lebesgue. Ainsi par le Théorème 74 nous

avons $h(T, \alpha) \leq r(\delta_0, X) \leq s(\delta_0, X) \leq h(T)$. Or α est un générateur car la constante d'expansivité de T est δ . On conclut en utilisant (1). $\square$

5. Les shifts ou décalages

Avant de calculer l'entropie de certains shifts, voici un aspect de leur intérêt.

Théorème 83. *Soit T un homéomorphisme expansif d'un espace métrique compact. Alors, il existe un entier $k \geq 1$, un fermé Ω de $X_k = \{0, 1, \dots, k-1\}^{\mathbb{Z}}$, tel que $S\Omega = \Omega$, où S est le décalage de X_k , et une surjection continue $\phi : \Omega \rightarrow X$ tel que $\phi \circ S = T \circ \phi$.*

Théorème 84. *Le décalage bilatère sur $X = Y^{\mathbb{Z}}$, où $Y = \{0, 1, \dots, k-1\}$, a une entropie topologique $\log(k)$.*

DÉMONSTRATION. Soit $\alpha = \{A_0, \dots, A_{k-1}\}$ où $A_j = \{(x_n)_{n \in \mathbb{Z}} | x_0 = j\}$, c'est un générateur pour T (**Exercice**). Alors, par le Théorème 82,

$$h(T) = h(T, \alpha) = \lim_{n \rightarrow \infty} \frac{1}{n} \log N \left(\bigvee_{i=0}^{n-1} T^{-i} \alpha \right) = \lim_{n \rightarrow \infty} \frac{1}{n} \log(k^n) = \log(k).$$

$\square$

On a le même résultat pour le décalage sur $Y^{\mathbb{N}}$.

Théorème 85. *Soit $T : X \rightarrow X$ le décalage bilatère sur $X = Y^{\mathbb{Z}}$ où $Y = \{0, 1, \dots, k-1\}$. Si X_1 est un sous ensemble fermé de X avec $TX_1 = X_1$, alors $h(T|_{X_1}) = \lim_{n \rightarrow \infty} \frac{1}{n} \log p_{X_1}(n)$.*

DÉMONSTRATION. Soit α le générateur pour $T : X \rightarrow X$ de la preuve du Théorème 84. Evidemment, α est un générateur pour $T|_{X_1}$ et $p_{X_1}(n) = N(\bigvee_{i=0}^{n-1} T^{-i} \alpha)$. Le résultat découle maintenant du Théorème 82. $\square$

Exercice 8. *Montrer que si (X, T) est un décalage tel qu'il existe n vérifiant $p_X(n) \leq n$ alors X est fini et toutes les orbites sont ultimement périodiques.*

6. Les homéomorphismes du cercle

Nous savons déjà que les rotations de groupes métriques compacts sont d'entropie nulle car il existe une métrique qui fait de la rotation une isométrie. En fait, nous allons maintenant montrer que tous les homéomorphismes sur le cercle sont d'entropie nulle.

Théorème 86. *Si $T : \mathbb{T} \rightarrow \mathbb{T}$ est un homéomorphisme alors $h(T) = 0$.*

DÉMONSTRATION. Nous savons que T envoie les intervalles sur les intervalles car les intervalles sont les sous-ensembles connexes de $\mathbb{T}$. Choisissons $\varepsilon > 0$ tel que :

$$d_n(x, y) \leq \varepsilon \text{ implique } d(T^{-1}x, T^{-1}y) \leq 1/4.$$

Nous avons $r_1(\varepsilon, \mathbb{T}) \leq [1/\varepsilon] + 1$. Nous allons montrer que $r_n(\varepsilon, \mathbb{T}) \leq n([1/\varepsilon] + 1)$. Soit F un ensemble qui $(n-1, \varepsilon)$ -engendre $\mathbb{T}$ par T et de cardinalité $r_{n-1}(\varepsilon, \mathbb{T})$. Considérons les

points de $T^{n-1}F$ et les intervalles qu'ils déterminent. Ajoutons des points à ces ensembles de tel sorte que les intervalles aient une longueur inférieure à ε . Nous devons ajouter au plus $[1/\varepsilon] + 1$ points. Si E est l'ensemble des nouveaux points, nous posons

$$F' = F \cup T^{-(n-1)}E.$$

Montrons que F' est un ensemble qui (n, ε) -engendre $\mathbb{T}$ par T . Soit $x \in \mathbb{T} \setminus F$. Il existe $y \in F$ tel que

$$\max_{0 \leq i \leq n-2} d(T^i x, T^i y) \leq \varepsilon.$$

Il suffit de montrer que $d(T^{n-1}x, T^{n-1}y) \leq \varepsilon$. Nous allons en fait construire (à partir de y) un $z \in F'$ qui vérifie $d(T^{n-1}x, T^{n-1}z) \leq \varepsilon$, puis, par induction décroissante $\max_{0 \leq i \leq n-2} d(T^i x, T^i z) \leq \varepsilon$.

Il existe deux intervalles fermés dont les points extrémaux sont $T^{n-1}(x)$ et $T^{n-1}(y)$. L'image par T^{-1} de ces intervalles sont les deux intervalles dont les points extrémaux sont $T^{n-2}(x)$ et $T^{n-2}(y)$. L'un d'entre eux est de longueur inférieure ou égale à ε . Notons le I' et posons $I = TI'$. Si I est de diamètre inférieur à ε , alors $d(T^{n-1}x, T^{n-1}y) \leq \varepsilon$. Ce qui montre que F' (n, ε) -engendre $\mathbb{T}$ par T .

Supposons que le diamètre de I est supérieur à ε . Ainsi, nous pouvons choisir $z' \in I \cap T^{n-1}F'$ tel que $d(T^{n-1}x, z') \leq \varepsilon$. Posons $z = T^{-n+1}z'$. Ainsi $T^{n-1}z \in I$, $z \in F'$ et $d(T^{n-1}x, T^{n-1}z) \leq \varepsilon$. Par conséquent, $T^{n-2}(z) \in I'$ et donc $d(T^{n-2}x, T^{n-2}z) \leq \varepsilon$.

D'après l'hypothèse sur ε , l'intervalle I' est envoyé par T^{-1} sur l'intervalle I'' dont les points extrémaux sont $T^{n-3}(x)$ et $T^{n-3}(y)$ et qui est de longueur inférieure à $1/4$. I'' est donc le plus petit intervalle délimité par ces deux points. Par conséquent, la longueur de I'' est inférieure ou égale à ε . Puisque $T^{n-3}(z) \in I''$ nous avons $d(T^{n-3}x, T^{n-3}z) \leq \varepsilon$. Par induction

$$d(T^i x, T^i z) \leq \varepsilon \text{ pour } 0 \leq i \leq n-1.$$

Ainsi F' (n, ε) -engendre $\mathbb{T}$. Nous en déduisons

$$r_n(\varepsilon, \mathbb{T}) \leq r_{n-1}(\varepsilon, \mathbb{T}) + [1/\varepsilon] + 1$$

d'où $r_n(\varepsilon, \mathbb{T}) \leq n([1/\varepsilon] + 1)$. Par conséquent $r(\varepsilon, \mathbb{T}) = 0$ et $h(T) = 0$. $\square$

Corollaire 87. *Tout homéomorphisme T sur $[0, 1]$ a une entropie nulle.*

DÉMONSTRATION. L'application $T : [0, 1] \rightarrow [0, 1]$ est telle que $T(0) = 0$ et $T(1) = 1$, ou $T(0) = 1$ et $T(1) = 0$. Dans les deux cas T^2 fixe 0 et 1. Soit S un homéomorphisme de $[0, 1]$ ayant pour points fixes 0 et 1. Soit $\phi : [0, 1] \rightarrow \mathbb{T}$ la fonction continue définie par $\phi(t) = e^{2\pi it}$. La fonction ϕ est injective sur $(0, 1)$ et donc on peut vérifier que $\phi S \phi^{-1}$ est un homéomorphisme sur $\mathbb{T}$ qui fixe $0 \in \mathbb{T}$. Soit α_n un recouvrement ouvert de $[0, 1]$ par les intervalles $[0, 1/n)$, $(1 - 1/n, 1]$ et $(k/2n, (k+2)/2n)$, $1 \leq k \leq 2n-3$. Les arcs ouverts

$\phi((k/2n, (k+2)/2n)), 1 \leq k \leq 2n-3$, et $\phi([0, 1/n) \cup (1-1/n, 1])$ forment un recouvrement ouvert β_n de $\mathbb{T}$.

Prenons un atome de $\bigvee_{j=0}^{p-1} (\phi S \phi^{-1})^j \beta_n : \cap_{j=0}^{p-1} (\phi S \phi^{-1})^j A_j$ où les A_j sont des atomes de β_n . Si A_j n'est pas l'atome $\phi([0, 1/n) \cup (1-1/n, 1])$ alors on lui associe $B_j = \phi^{-1} A_j$, sinon on lui associe $[0, 1/n)$ ou $(1-1/n, 1]$. Cela fait au plus 2^p possibilités. Grâce à cette remarque nous obtenons

$$N \left(\bigvee_{j=0}^{p-1} S^j \alpha_n \right) \leq 2^p N \left(\bigvee_{j=0}^{p-1} (\phi S \phi^{-1})^j \beta_n \right).$$

Par conséquent $h(S, \alpha_n) \leq \log 2 + h(\phi S \phi^{-1}, \beta_n)$. Puis, grâce au Théorème 72, $h(S) \leq \log 2 + h(\phi S \phi^{-1}) = \log 2$, la dernière égalité s'obtenant par le Théorème 86.

Cette formule vaut pour tout homéomorphisme S de $[0, 1]$ qui fixe 0 et 1. Posons $S = T^{2q}$ pour obtenir $2qh(T) \leq \log 2$. D'où $h(T) = 0$. $\square$

Entropie métrique

1. Information et entropie

Soit $(X, \mathcal{B}, \mu)$ un espace de probabilité et $\alpha = (A_i; i \in I)$ une partition mesurable (finie ou dénombrable) de X , c'est-à-dire : $\alpha \subset \mathcal{B}$ et

- (1) $\mu(\cup_{i \in I} A_i) = 1$,
- (2) $\forall i, j \in I, i \neq j, \mu(A_i \cap A_j) = 0$.

Nous supposons implicitement que pour tout $i \in I, \mu(A_i) > 0$. Par commodité nous utiliserons la notation $\alpha = (A)$ pour dire que nous appellerons A les atomes de α .

Étant données deux partitions α et β , nous dirons que α est plus fine que β , ce que nous noterons $\alpha \succeq \beta$, si pour tout atome $A \in \alpha$ il existe un atome $B \in \beta$ tel que $A \subseteq B$. Nous dirons également que β est plus grossière que α .

Nous dirons que α et β sont indépendantes si pour tout $A \in \alpha$ et tout $B \in \beta$ on a $\mu(A \cap B) = \mu(A)\mu(B)$.

Lorsque α et β sont deux familles de sous-ensembles de X , nous poserons

$$\alpha \vee \beta = \{A \cap B : A \in \alpha, B \in \beta\}$$

que nous appellerons **raffinement** de α et β . Notons que si α et β sont deux partitions de X alors $\alpha \vee \beta$ est également une partition de X .

Définition 88. Soit $\alpha = (A)$ une partition mesurable et dénombrable. La fonction d'information associée à α est

$$I_\alpha(x) = - \sum_{A \in \alpha} 1_A(x) \log(\mu(A)),$$

i.e. $I_\alpha(x) = -\log \mu(A_i)$ si $x \in A_i$.

L'entropie associée à α est la quantité d'information moyenne :

$$H_\mu(\alpha) = \mathbb{E}(I_\alpha) = \int_X I_\alpha d\mu = - \sum_{A \in \alpha} \mu(A) \cdot \log(\mu(A)).$$

Pour la partition trivial $\mathcal{N} = \{X, \emptyset\}$ nous avons $I_{\mathcal{N}} = 0$ et $H_\mu(\mathcal{N}) = 0$.

Dans ce qui suit nous allons définir l'information et l'entropie conditionnelle. Nous identifierons souvent la partition α avec la σ -algèbre qu'elle engendre, c'est-à-dire :

$$\sigma(\alpha) = \left\{ \bigcup_{i \in J} A_i : J \subseteq I \right\}.$$

Définition 89. Soient $\alpha = (A)$ et $\beta = (B)$ deux partitions mesurables et dénombrables. L'information conditionnelle de α par rapport à β est

$$I_{\alpha|\beta} = - \sum_{B \in \beta} 1_B \left(\sum_{A \in \alpha} 1_A \log(\mu(A | B)) \right) = - \sum_{B \in \beta} \sum_{A \in \alpha} 1_{A \cap B} \log(\mu(A | B)),$$

où $\mu(A | B) = \frac{\mu(A \cap B)}{\mu(B)}$. L'entropie conditionnelle est

$$H(\alpha | \beta) = \int_X I_{\alpha|\beta} d\mu = - \sum_{B \in \beta} \sum_{A \in \alpha} \mu(A \cap B) \cdot \log(\mu(A | B)).$$

Suposons que nous connaissons la position du point x par rapport à une partition β , la valeur $I_{\alpha|\beta}(x)$ est un indicateur de la "quantité" d'information que nous avons en plus en connaissant la position de x par rapport à α .

Les propriétés suivantes découlent directement des définitions.

Lemme 90. Pour deux partitions β et $\alpha = (A_i)$.

- (1) Lorsque $\beta = \{X\}$, $\mu(A_i | \{\emptyset, X\}) = \mu(A_i)$ et $I_{\alpha|\beta}(x) = I_\alpha(x)$ pour tout $x \in X$.
- (2) $I_{\alpha|\beta}(x) \geq 0$ pour tout $x \in X$.
- (3) Si $T : X \rightarrow X$ préserve la mesure μ , alors $I_{\alpha|\beta}(Tx) = I_{T^{-1}\alpha|T^{-1}\beta}(x)$.
- (4) Si $\alpha \subset \sigma(\beta)$, alors $I_{\alpha|\beta} = 0$ μ p.p.
- (5) Réciproquement, si $H(\alpha | \beta) = 0$, alors $\alpha \subset \sigma(\beta)$.

DÉMONSTRATION. Seul le point 5 n'est pas évident.

Si $H(\alpha | \beta) = 0$, alors $I_{\alpha|\beta} = 0$ presque partout. Pour tout $A \in \alpha$ et tout $B \in \beta$, nous avons $\mu(A \cap B) = \mu(B)$ dès que $\mu(A \cap B) > 0$. Comme β est une partition, nous obtenons que $\mu(A) = \sum_{B \in \beta} \mu(A \cap B)$ et donc que $A = \cup_i B_i \cup E$ pour une certaine famille $\{B_i\} \subset \beta$ et E un ensemble μ négligeable. Remarquons que $\mathbb{E}(1_A | \beta) = \sum_{B \in \beta} 1_B \cdot \mu(A | B)$, par conséquent $\mathbb{E}(1_A | \beta) = 1_A$ μ p.p. Nous en concluons que $A \in \sigma(\beta)$. $\square$

Nous en déduisons le lemme suivant.

Lemme 91. Pour deux partitions β et $\alpha = (A_i)$.

- (1) Lorsque $\beta = \{X\}$, $H(\alpha | \{\emptyset, X\}) = H(\alpha)$.
- (2) Si $T : X \rightarrow X$ préserve la mesure μ , alors $H(\alpha | \beta) = H(T^{-1}\alpha | T^{-1}\beta)$ et $H(T^{-1}\alpha) = H(\alpha)$.
- (3) Si $\alpha \subset \sigma(\beta)$, alors $H(\alpha | \beta) = 0$.

(4) Nous avons $H(\alpha | \beta) \leq H(\alpha)$.

DÉMONSTRATION. Les points (1) à (3) s'obtiennent à partir du lemme précédent par simple intégration.

Pour le point (4) nous avons

$$\begin{aligned} H(\alpha | \beta) &= - \sum_{A \in \alpha, B \in \beta} \mu(A \cap B) \log \left(\frac{\mu(A \cap B)}{\mu(B)} \right) \\ &\leq - \sum_{A \in \alpha, B \in \beta} \mu(A \cap B) \log(\mu(A \cap B)) \\ &\leq - \sum_{A \in \alpha} \mu(A) \log(\mu(A)), \end{aligned}$$

puisque, par concavité de la fonction $t \mapsto -t \log t$, pour un $A \in \alpha$ fixé, nous avons

$$- \sum_{B \in \beta} \mu(A \cap B) \log(\mu(A \cap B)) \leq -\mu(A) \log(\mu(A)).$$

□

La propriété suivante se révélera très utile tout au long du chapitre.

Proposition 92 (Propriété de base). *Soient α, β, γ des partitions mesurables dénombrables. Alors,*

$$I_{\alpha \vee \beta | \gamma} = I_{\alpha | \gamma} + I_{\beta | \alpha \vee \gamma},$$

presque partout. De même, nous avons

$$H(\alpha \vee \beta | \gamma) = H(\alpha | \gamma) + H(\beta | \alpha \vee \gamma).$$

DÉMONSTRATION. Soient $\alpha = (A)$, $\beta = (B)$ et $\gamma = (C)$ des partitions mesurables dénombrables. Soit $x \in A \cap B \cap C$ où $A \in \alpha$, $B \in \beta$ et $C \in \gamma$. Alors,

$$\begin{aligned} I_{\alpha \vee \beta | \gamma}(x) &= -\log(\mu(A \cap B | C)) = -\log\left(\frac{\mu(A \cap B \cap C)}{\mu(C)}\right), \\ I_{\alpha | \gamma}(x) &= -\log(\mu(A | C)) = -\log\left(\frac{\mu(A \cap C)}{\mu(C)}\right) \text{ et} \\ I_{\beta | \alpha \vee \gamma}(x) &= -\log(\mu(B | A \cap C)) = -\log\left(\frac{\mu(B \cap A \cap C)}{\mu(A \cap C)}\right). \end{aligned}$$

Ainsi, $I_{\alpha | \gamma}(x) + I_{\beta | \alpha \vee \gamma}(x) = -\log \frac{\mu(B \cap A \cap C)}{\mu(C)} = I_{\alpha \vee \beta | \gamma}(x)$.

La seconde relation se déduit de la première. □

Corollaire 93. *Soient α, β, γ des partitions mesurables de X . Si $\alpha \succeq \beta$ alors $I_{\alpha | \gamma} \geq I_{\beta | \gamma}$ et $H(\alpha | \gamma) \geq H(\beta | \gamma)$. En particulier,*

$$H(\alpha) \geq H(\beta).$$

DÉMONSTRATION. Si $\alpha \succeq \beta$, la propriété précédente implique que $I_{\alpha|\gamma} = I_{\alpha \vee \beta|\gamma} = I_{\beta|\gamma} + I_{\alpha|\beta \vee \gamma}$, d'où nous concluons que $I_{\alpha|\gamma} \geq I_{\beta|\gamma}$. □

Supposons que $T : X \rightarrow X$ préserve la mesure μ . Pour une partition mesurable α , nous notons

$$\bigvee_{i=0}^{n-1} T^{-i}\alpha = \{A_{r_0} \cap T^{-1}A_{r_1} \cap \dots \cap T^{-(n-1)}A_{r_{n-1}} ; A_{r_i} \in \alpha, i = 0 \dots n-1\}.$$

Pour $n \geq 1$, notons $H_n(\alpha) = H(\bigvee_{i=0}^{n-1} T^{-i}\alpha)$. D'après la proposition précédente, nous avons pour tout $n, m \in \mathbb{N}$

$$\begin{aligned} H_{n+m}(\alpha) &= H\left(\bigvee_{i=0}^{n+m-1} T^{-i}\alpha\right) \\ &= H\left(\bigvee_{i=0}^{n-1} T^{-i}\alpha\right) + H\left(\bigvee_{i=n}^{n+m-1} T^{-i}\alpha \mid \bigvee_{i=0}^{n-1} T^{-i}\alpha\right) \\ &\leq H\left(\bigvee_{i=0}^{n-1} T^{-i}\alpha\right) + H\left(\bigvee_{i=n}^{n+m-1} T^{-i}\alpha\right) \text{ Par le point (4) du Lemme 91} \\ &= H_n(\alpha) + H_m(\alpha), \end{aligned}$$

puisque $\bigvee_n^{n+m-1} T^{-i}\alpha = T^{-n} \bigvee_0^{m-1} T^{-i}\alpha$ et H est T -invariant.

La suite $(H_n)_n$ est donc une suite sous-additive. Par le Lemme 53, la limite suivante est bien définie.

Définition 94. *L'entropie d'une partition α relativement à une transformation $T : X \rightarrow X$ est la limite*

$$h_\mu(T, \alpha) = \lim_{n \rightarrow +\infty} \frac{H(\bigvee_{i=0}^{n-1} T^{-i}\alpha)}{n}.$$

Définition 95. *L'entropie du s.d.m. $(X, \mathcal{B}, \mu, T)$ est*

$$h_\mu(T) = \sup \{h_\mu(T, \alpha) ; \alpha \text{ partition mesurable et finie}\}.$$

Nous verrons que cette constante est un invariant de s.d.m.

2. Calcul d'entropie

Comme nous pouvons nous l'imaginer l'entropie métrique est difficile à déterminer en pratique. Nous voudrions donner une méthode effective pour la calculer.

Nous commençons tout d'abord par un résultat de Sinaï montrant que, sous de faible hypothèses, le supremum est atteint pour une certaine partition.

Définition 96. Une partition α telle que $H_\mu(\alpha) < +\infty$ est un générateur fort pour l'espace $(X, \mathcal{B}, \mu)$ si $\bigvee_{i=0}^{\infty} T^{-i}\alpha = \mathcal{B} \pmod{\mu}$.

Si T est inversible, alors la partition α telle que $H_\mu(\alpha) < +\infty$ est un générateur pour l'espace $(X, \mathcal{B}, \mu)$ si $\bigvee_{i=-\infty}^{\infty} T^{-i}\alpha = \mathcal{B} \pmod{\mu}$.

Théorème 97 (Sinaï). Soit $(X, \mathcal{B}, \mu, T)$ un s.d.m.. Soit α une partition d'entropie finie. Alors,

- (1) si α est un générateur fort, alors $h_\mu(T) = h_\mu(T, \alpha)$;
- (2) Si T est inversible et α est un générateur alors $h_\mu(T) = h_\mu(T, \alpha)$.

Ce théorème sera démontré dans la section 6.

Exemple. Pour le décalage σ sur $X = \{0, 1\}^{\mathbb{Z}}$ munit de la mesure de Bernouilli μ de paramètre $1/2$, la partition $\alpha = \{[0], [1]\}$ est un générateur (Exercice). Les atomes de la partition $\alpha_n = \bigvee_{i=0}^{n-1} \sigma^{-i}\alpha$ sont les cylindres associés aux mots de longueur n . La mesure de chacun de ces atomes est donc de 2^{-n} . Par conséquent

$$H_\mu\left(\bigvee_{i=0}^{n-1} \sigma^{-i}\alpha\right) = \sum_{C \in \alpha_n} -\mu(C) \log(\mu(C)) = -\log(2^{-n}).$$

Ainsi, $h_\mu(\sigma, \alpha) = \log(2)$, et par le théorème de Sinaï, $h_\mu(\sigma) = \log(2)$.

Nous allons donner à présent une application de l'entropie pour décrire la mesure asymptotique des éléments d'une partition.

Corollaire 98 (du théorème de Shanon-Mac Millan -Breiman). Soit $(X, \mathcal{B}, \mu, T)$ un s.d.m. ergodique et α une partition mesurable et dénombrable telle que $H_\mu(\alpha) < \infty$. Notons $B_n(x)$ l'atome de $\bigvee_{i=0}^{n-1} T^{-i}\alpha$ contenant $x \in X$. Alors,

$$\lim_{N \rightarrow \infty} -\frac{1}{N} \log \mu(B_N(x)) = h_\mu(\alpha, T)$$

μ -presque partout et dans $L^1(X, \mathcal{B}, \mu)$.

Ce résultat découle d'un théorème plus général de Shanon-Mac Millan-Breiman, présenté avec les preuves, dans la section 5.

3. Propriété de l'entropie

L'entropie est un invariant de s.d.m.

Proposition 99.

Si $(X_2, \mathcal{B}_2, \mu_2, T_2)$ est un facteur de $(X_1, \mathcal{B}_1, \mu_1, T_1)$ alors $h_{\mu_2}(T_2) \leq h_{\mu_1}(T_1)$.

Si $(X_1, \mathcal{B}_1, \mu_1, T_1)$ et $(X_2, \mathcal{B}_2, \mu_2, T_2)$ sont conjugués alors $h_{\mu_1}(T_1) = h_{\mu_2}(T_2)$.

DÉMONSTRATION. Soit α une partition mesurable et dénombrable de X_2 telle que $H_{\mu_2}(\alpha) < \infty$. Si $\Phi : X_1 \rightarrow X_2$ est un facteur entre les s.d.m. $(X_1, \mathcal{B}_1, \mu_1, T_1)$ et $(X_2, \mathcal{B}_2, \mu_2, T_2)$ alors $\Phi^{-1}\alpha = (\Phi^{-1}A; A \in \alpha)$ est une partition de X_1 . Elle est d'entropie finie car

$$\begin{aligned} H_{\mu_1}(\Phi^{-1}\alpha) &= - \sum_{A \in \alpha} \mu_1(\Phi^{-1}A) \log \mu_1(\Phi^{-1}A) \\ &= - \sum_{A \in \alpha} \mu_2(A) \log \mu_2(A) = H_{\mu_2}(\alpha) < \infty. \end{aligned}$$

$$\begin{aligned} h_{\mu_1}(T_1, \Phi^{-1}\alpha) &= \lim_{N \rightarrow \infty} \frac{1}{N} H_{\mu_1} \left(\bigvee_{n=0}^{N-1} T_1^{-n}(\Phi^{-1}\alpha) \right) \\ &= \lim_{N \rightarrow \infty} \frac{1}{N} H_{\mu_1} \left(\bigvee_{n=0}^{N-1} \Phi^{-1}(T_2^{-n}\alpha) \right) \\ &= \lim_{N \rightarrow \infty} \frac{1}{N} H_{\mu_1} \left(\Phi^{-1} \left(\bigvee_{n=0}^{N-1} T_2^{-n}\alpha \right) \right) \\ &= \lim_{N \rightarrow \infty} \frac{1}{N} H_{\mu_2} \left(\bigvee_{n=0}^{N-1} T_2^{-n}\alpha \right) = h_{\mu_2}(T_2, \alpha). \end{aligned}$$

Ainsi, $h_{\mu_2}(T_2, \alpha) = h_{\mu_1}(T_1, \Phi^{-1}\alpha) \leq h_{\mu_1}(T_1)$, puis $h_{\mu_2}(T_2) \leq h_{\mu_1}(T_1)$. $\square$

Remarquer le résultat suivant, affirmant que, dans le cas des systèmes définis par des shifts, l'entropie est un invariant complet.

Théorème 100 (Ornstein). *Deux shifts de Bernoulli sont conjugués en mesure si et seulement s'ils ont la même entropie.*

Proposition 101. *Soit $(X, \mathcal{B}, \mu, T)$ un s.d.m..*

- (1) *Pour tout $n \geq 0$ nous avons $h_{\mu}(T^n) = nh_{\mu}(T)$.*
- (2) *Si T est inversible, alors $h_{\mu}(T^n) = |n|h_{\mu}(T)$ pour tout $n \in \mathbb{Z}$.*

DÉMONSTRATION. Montrons (1). Si $n = 0$, alors $h_{\mu}(\alpha, T^0) = 0$. Soit $n > 0$ et α une partition telle que $H_{\mu}(\alpha) < \infty$. Alors, par le Corollaire 93

$$\begin{aligned} h_{\mu}(\alpha, T^n) &\leq h_{\mu} \left(\bigvee_{i=0}^{n-1} T^{-i}\alpha, T^n \right) = \lim_{N \rightarrow \infty} \frac{1}{N} H_{\mu} \left(\bigvee_{l=0}^{N-1} T^{-nl} \left(\bigvee_{i=0}^{n-1} T^{-i}\alpha \right) \right) \\ &= \lim_{N \rightarrow \infty} \frac{1}{N} H_{\mu} \left(\bigvee_{i=0}^{nN-1} T^{-i}\alpha \right) = \lim_{N \rightarrow \infty} \frac{n}{nN} H_{\mu} \left(\bigvee_{i=0}^{nN-1} T^{-i}\alpha \right) \\ &= nh_{\mu}(\alpha, T). \end{aligned}$$

De plus, en utilisant les égalités précédentes nous obtenons

$$nh_\mu(\alpha, T) = h_\mu\left(\bigvee_{i=0}^{n-1} T^{-i}\alpha, T^n\right) \leq h_\mu(T^n),$$

car par sub-additivité $H_\mu\left(\bigvee_{i=0}^{n-1} T^{-i}\alpha\right) \leq \sum_{i=0}^{n-1} H_\mu(T^{-i}\alpha) = nH_\mu(\alpha) < \infty$.

Montrons (2). Il suffit de montrer que $h_\mu(T^{-1}) = h_\mu(T)$. Or,

$$H_\mu\left(\bigvee_{n=0}^{N-1} (T^{-1})^{-n}\alpha\right) = H_\mu\left(T^{-(N-1)}\left(\bigvee_{n=0}^{N-1} T^n\alpha\right)\right) = H_\mu\left(\bigvee_{n=0}^{N-1} T^{-n}\alpha\right),$$

donc $h_\mu(\alpha, T^{-1}) = h_\mu(\alpha, T)$ pour toute partition mesurable et dénombrable α . $\square$

Soient $(X_i, \mathcal{B}_i, \mu_i, T_i)$, $i = 1, 2$ deux s.d.m. séparables. On définit $T_1 \otimes T_2 : X_1 \times X_2 \rightarrow X_1 \times X_2$ par $T_1 \otimes T_2(x_1, x_2) = (T_1x_1, T_2x_2)$. Cette application préserve la mesure $\mu_1 \otimes \mu_2$.

Proposition 102. *Soient $(X_i, \mathcal{B}_i, \mu_i, T_i)$, $i = 1, 2$, deux s.d.m. séparables. Alors,*

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

DÉMONSTRATION. Nous avons

$$\begin{aligned} & H_{\mu_1 \otimes \mu_2}\left(\bigvee_{k=0}^{N-1} (T_1 \otimes T_2)^{-k}(\alpha_n^1 \times \alpha_n^2)\right) \\ &= H_{\mu_1 \otimes \mu_2}\left(\left(\bigvee_{k=0}^{N-1} T_1^{-k}(\alpha_n^1)\right) \times \left(\bigvee_{k=0}^{N-1} T_2^{-k}(\alpha_n^2)\right)\right) \\ &= - \sum_{A \in \alpha_n^1, B \in \alpha_n^2} \mu_1 \otimes \mu_2(A \times B) \log(\mu_1 \otimes \mu_2(A \times B)) \\ &= - \sum_{A \in \alpha_n^1, B \in \alpha_n^2} \mu_1(A)\mu_2(B) \log(\mu_1(A)\mu_2(B)) \\ &= - \sum_{A \in \alpha_n^1, B \in \alpha_n^2} \mu_1(A)\mu_2(B) \log(\mu_1(A)) - \sum_{A \in \alpha_n^1, B \in \alpha_n^2} \mu_1(A)\mu_2(B) \log(\mu_2(B)) \\ &= - \sum_{A \in \alpha_n^1} \mu_1(A) \log(\mu_1(A)) - \sum_{B \in \alpha_n^2} \mu_2(B) \log(\mu_2(B)) \\ &= H_{\mu_1}\left(\bigvee_{k=0}^{N-1} (T_1)^{-k} \alpha_n^1\right) + H_{\mu_2}\left(\bigvee_{k=0}^{N-1} (T_2)^{-k} \alpha_n^2\right). \end{aligned}$$

Ainsi, en divisant par N et en passant à la limite, nous avons

$$h_{\mu_1 \otimes \mu_2}(T_1 \otimes T_2, \alpha_n^1 \times \alpha_n^2) = h_{\mu_1}(T_1, \alpha_n^1) + h_{\mu_2}(T_2, \alpha_n^2).$$

Le résultat s'en déduit. $\square$

Pour la suite, nous aurons besoin de généraliser les définitions d'information et d'entropie par rapport, non plus à une partition, mais à une sous-tribu.

Définition 103. Soient $(X, \mathcal{B}, \mu)$ un espace de probabilité, α une partition dénombrable et mesurable, et $\mathcal{A} \subseteq \mathcal{B}$ une sous- σ -algèbre de $\mathcal{B}$. La fonction d'information de α par rapport à $\mathcal{A}$ se définit par

$$I_{\alpha|\mathcal{A}} = - \sum_{A \in \alpha} 1_A \cdot \log(\mathbb{E}(1_A|\mathcal{A})).$$

L'entropie conditionnelle de α par rapport à $\mathcal{A}$ est

$$H(\alpha | \mathcal{A}) = \int_X I_{\alpha|\mathcal{A}} d\mu.$$

Lorsqu'il sera nécessaire de spécifier la mesure utilisée, nous noterons $I_\mu(\alpha|\mathcal{A})$ et $H_\mu(\alpha|\mathcal{A})$ respectivement. Cette définition généralise celle utilisant les partitions puisque $I_{\alpha|\beta} = I_{\alpha|\sigma(\beta)}$.

Proposition 104. Si $\mathcal{A}_1 \supseteq \mathcal{A}_2$ alors $H(\alpha | \mathcal{A}_1) \leq H(\alpha | \mathcal{A}_2)$.

Cette proposition généralise le (4) de Lemme 91 avec $\mathcal{A}_2$ la tribu grossière. Pour prouver la Proposition 104 nous aurons besoin de l'inégalité de Jensen.

Lemme 105 (Inégalité de Jensen). Soit $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ une fonction concave. Alors, pour tout $f \in L^1(X, \mathcal{B}, \mu)$ et toute sous- σ -algèbre $\mathcal{A} \subseteq \mathcal{B}$, nous avons

$$\varphi(\mathbb{E}(f | \mathcal{A})) \geq \mathbb{E}(\varphi(f) | \mathcal{A}).$$

DÉMONSTRATION DE LA PROPOSITION 104. La fonction φ de $[0, 1]$ dans $\mathbb{R}_+$ définie par $\varphi(t) = -t \cdot \log(t)$ est concave (on pose $\varphi(0) = 0$).

$$\begin{aligned} H(\alpha | \mathcal{A}) &= - \sum_{A \in \alpha} \int_A \log(\mathbb{E}(1_A | \mathcal{A})) d\mu \\ &= - \sum_{A \in \alpha} \int_X \mathbb{E}(1_A \log(\mathbb{E}(1_A | \mathcal{A})) | \mathcal{A}) d\mu \\ &= - \sum_{A \in \alpha} \int_X \log(\mathbb{E}(1_A | \mathcal{A})) \mathbb{E}(1_A | \mathcal{A}) d\mu \\ &= \sum_{A \in \alpha} \int_A \varphi(\mathbb{E}(1_A | \mathcal{A})) d\mu. \end{aligned}$$

Utilisons l'inégalité de Jensen avec $A \in \alpha$, $f = \mathbb{E}(1_A | \mathcal{A}_1)$ et $\mathcal{A} = \mathcal{A}_2$:

$$\varphi(\mathbb{E}(\mathbb{E}(1_A | \mathcal{A}_1) | \mathcal{A}_2)) \geq \mathbb{E}(\varphi(\mathbb{E}(1_A | \mathcal{A}_1)) | \mathcal{A}_2)$$

c'est-à-dire

$$\varphi(\mathbb{E}(1_A | \mathcal{A}_2)) \geq \mathbb{E}(\varphi(\mathbb{E}(1_A | \mathcal{A}_1)) | \mathcal{A}_2).$$

En intégrant nous obtenons

$$\int_X \varphi(\mathbb{E}(1_A | \mathcal{A}_2)) d\mu \geq \int_X \varphi(\mathbb{E}(1_A | \mathcal{A}_1)) d\mu.$$

En sommant sur $A \in \alpha$ et en utilisant les premières égalités obtenues nous en déduisons $H(\alpha | \mathcal{A}_2) \geq H(\alpha | \mathcal{A}_1)$. $\square$

4. Rappels sur les martingales

Nous rappelons ici certains résultats qui seront nécessaires pour prouver les résultats de la section 2.

Le lemme suivant est un lemme technique très utile pour prouver les convergences.

Lemme 106 (Lemme maximal). *Soient $(X, \mathcal{B}, \mu)$ un espace de probabilité et $\mathcal{B}_1 \subseteq \mathcal{B}_2 \subseteq \dots \subseteq \mathcal{B}_N \subseteq \dots \subseteq \mathcal{B}$ une suite croissante de sous- σ -algèbres de $\mathcal{B}$. Alors, pour tout $f \in L^1(X, \mathcal{B}, \mu)$, $\lambda > 0$ et $N \in \mathcal{N}^*$, nous avons :*

$$\text{pour tout } \lambda > 0, \quad \mu(E) \leq \frac{1}{\lambda} \int_X |f| d\mu,$$

où $E = \{x \in X; \max_{1 \leq n \leq N} \mathbb{E}(f | \mathcal{B}_n)(x) > \lambda\}$.

DÉMONSTRATION. Nous pouvons supposer $f \geq 0$. Nous avons $E = \bigcup_{n=1}^N E_n$ où $E_n =$

$$\{x \in X; (\forall k \in \{1, \dots, n-1\}, \mathbb{E}(f | \mathcal{B}_k)(x) \leq \lambda) \text{ et } \mathbb{E}(f | \mathcal{B}_n)(x) > \lambda\}.$$

Or, les E_n sont deux à deux disjoints et $E_n \in \mathcal{B}_n$, donc

$$\begin{aligned} \int_X f d\mu &\geq \int_E f d\mu = \sum_{n=1}^N \int_{E_n} f d\mu = \sum_{n=1}^N \int_{E_n} \mathbb{E}(f | \mathcal{B}_n) d\mu \\ &\geq \sum_{n=1}^N \lambda \cdot \mu(E_n) = \lambda \cdot \mu(E). \end{aligned}$$

$\square$

Corollaire 107. *Soit $(\mathcal{B}_n)_{n \in \mathbb{N}}$ une suite croissante de sous- σ -algèbres de β . Alors, pour tout $f \in L^1(X, \mathcal{B}, \mu)$ et tout $\lambda > 0$,*

$$\mu \left\{ x \in X; \sup_{n \in \mathbb{N}} \mathbb{E}(f | \mathcal{B}_n) > \lambda \right\} \leq \frac{1}{\lambda} \int_X |f| d\mu.$$

DÉMONSTRATION. Exercice. $\square$

Théorème 108 (Théorème de convergence des martingales croissantes).

Prenons $(X, \mathcal{B}, \mu)$ un espace de probabilité, $(\mathcal{B}_n)_n$ une suite croissante de sous- σ -algèbres de $\mathcal{B}$ et

$$\mathcal{B}_\infty = \{B \in \mathcal{B}; \forall \epsilon > 0, \exists n \in \mathbb{N}, \exists B_n \in \mathcal{B}_n \text{ tel que } \mu(B \Delta B_n) < \epsilon\}$$

sa limite. Alors, pour tout $f \in L^1(X, \mathcal{B}, \mu)$ nous avons

$$\lim_{n \rightarrow \infty} \mathbb{E}(f | \mathcal{B}_n) = \mathbb{E}(f | \mathcal{B}_\infty) \text{ dans } L^1(X, \mathcal{B}, \mu) \text{ et } \mu\text{-p.p.}$$

DÉMONSTRATION. Prouvons la convergence dans $L^1(X, \mathcal{B}, \mu)$. Le résultat est vrai pour f dans $\cup_{n \in \mathbb{N}} L^1(X, \mathcal{B}_n, \mu)$. En effet, si f appartient à $L^1(X, \mathcal{B}_n, \mu)$, alors pour tout $n \leq n' \leq \infty$, $\mathbb{E}(f | \mathcal{B}_{n'}) = f$.

Par ailleurs, $\cup_{n \in \mathbb{N}} L^1(X, \mathcal{B}_n, \mu)$ est dense dans $L^1(X, \mathcal{B}_\infty, \mu)$ (exercice).

Prenons $f \in L^1(X, \mathcal{B}_\infty, \mu)$ et $\epsilon > 0$. Il existe $n_0 \in \mathbb{N}$ et $g \in L^1(X, \mathcal{B}_{n_0}, \mu)$ tels que $\|f - g\|_1 < \epsilon$. Pour $n \geq n_0$ nous avons

$$\begin{aligned} \|\mathbb{E}(f | \mathcal{B}_n) - f\|_1 &= \|\mathbb{E}(f | \mathcal{B}_n) - \mathbb{E}(g | \mathcal{B}_n) + \mathbb{E}(g | \mathcal{B}_n) - g + g - f\|_1 \\ &\leq \|\mathbb{E}(f - g | \mathcal{B}_n)\|_1 + \|g - f\|_1 \leq 2\epsilon. \end{aligned}$$

puisque $\mathbb{E}(g | \mathcal{B}_n) = g$ pour tout $n \geq n_0$. Ainsi nous avons le résultat pour $f \in L^1(X, \mathcal{B}, \mu)$ puisque $\mathbb{E}(f | \mathcal{B}_\infty) \in L^1(X, \mathcal{B}_\infty, \mu)$ et $\mathbb{E}(f | \mathcal{B}_n) = \mathbb{E}(\mathbb{E}(f | \mathcal{B}_\infty) | \mathcal{B}_n)$.

Montrons la convergence μ -p.p.. Comme précédemment, il suffit de faire la preuve pour $f \in L^1(X, \mathcal{B}_\infty, \mu)$. Soit $\epsilon > 0$. Les inégalités suivantes permettent de conclure.

$$\begin{aligned} &\mu \left\{ \limsup_{n \rightarrow \infty} |\mathbb{E}(f | \mathcal{B}_n) - f| > \sqrt{\epsilon} \right\} \\ &= \mu \left\{ \limsup_{n \rightarrow \infty} |\mathbb{E}(f | \mathcal{B}_n) - \mathbb{E}(g | \mathcal{B}_n) + \mathbb{E}(g | \mathcal{B}_n) - g + g - f| > \sqrt{\epsilon} \right\} \\ &\leq \mu \left\{ \limsup_{n \rightarrow \infty} |\mathbb{E}(f - g | \mathcal{B}_n)| > \frac{\sqrt{\epsilon}}{2} \right\} + \mu \left\{ |g - f| > \frac{\sqrt{\epsilon}}{2} \right\} \\ &\leq \mu \left\{ \sup_{n \in \mathbb{N}} |\mathbb{E}(f - g | \mathcal{B}_n)| > \frac{1}{2} \sqrt{\epsilon} \right\} + \mu \left\{ |f - g| > \frac{1}{2} \sqrt{\epsilon} \right\} \\ &\leq \frac{2}{\sqrt{\epsilon}} \int_X |f - g| d\mu + \frac{2}{\sqrt{\epsilon}} \int_X |f - g| d\mu \leq 4\sqrt{\epsilon}, \end{aligned}$$

où nous utilisons le corollaire du Lemme Maximal et l'inégalité $\mu\{|f| > \lambda\} \leq \frac{1}{\lambda} \int_X |f| d\mu$. $\square$

Nous voulons appliquer le théorème des martingales croissantes à la fonction d'information. Pour cela nous avons besoin du lemme technique.

Lemme 109 (Lemme de Chung). *Soit $(X, \mathcal{B}, \mu)$ un espace de probabilité. Soit α une partition mesurable et dénombrable d'entropie finie. Soit $(\mathcal{A}_n)_{n \in \mathbb{N}}$ une suite croissante de sous- σ -algèbres de $\mathcal{B}$. Alors,*

$$\int_X \sup_{n \in \mathbb{N}} I_{\alpha | \mathcal{A}_n} d\mu \leq H_\mu(\alpha) + 1.$$

En particulier, $\sup_{n \in \mathbb{N}} I_{\alpha | \mathcal{A}_n} \in L^1(X, \mathcal{B}, \mu)$.

Nous sommes enfin en mesure de prouver le résultat fondamental suivant.

Si $(\mathcal{A}_n)_{n \in \mathbb{N}}$ est une suite de partitions ou de sous- σ -algèbres, notons $\bigvee_n \mathcal{A}_n$ la sous- σ -algèbre engendrée par $\cup_n \mathcal{A}_n$. On peut remarquer que $\mathcal{A}_\infty = \bigvee_n \mathcal{A}_n \pmod{\mu}$.

Corollaire 110. *Soient α une partition mesurable finie, $(\mathcal{A}_n)_{n \in \mathbb{N}}$ une suite croissante de sous- σ -algèbres et $\mathcal{A}_\infty = \bigvee_n \mathcal{A}_n$.*

Alors, $I_{\alpha | \mathcal{A}_n} \rightarrow_{n \rightarrow \infty} I_{\alpha | \mathcal{A}_\infty}$ μ -presque partout et dans $L^1(X, \mu)$. Ainsi

$$H_\mu(\alpha | \mathcal{A}_n) \rightarrow_{n \rightarrow \infty} H_\mu(\alpha | \mathcal{A}_\infty).$$

DÉMONSTRATION. Fixons $A \in \alpha$. Grâce au théorème des martingales croissantes nous avons $\mu(A | \mathcal{A}_n) \rightarrow_{n \rightarrow \infty} \mu(A | \mathcal{A}_\infty)$ μ -presque partout et dans $L^1(X, \mathcal{B}, \mu)$, où $\mu(A | \mathcal{A}_n) = \mathbb{E}(1_A | \mathcal{A}_n)$ et $\mu(A | \mathcal{A}_\infty) = \mathbb{E}(1_A | \mathcal{A}_\infty)$. Ceci implique que $I_{\alpha | \mathcal{A}_n}(x) \rightarrow I_{\alpha | \mathcal{A}_\infty}(x)$ μ presque partout.

Par le Lemme de Chung, nous avons que la suite $(I_{\alpha | \mathcal{A}_n})_n$ est dominée par la fonction intégrable $\sup_n I_{\alpha | \mathcal{A}_n}$. Ainsi par le théorème de convergence dominée de Lebesgue, nous avons $I_{\alpha | \mathcal{A}_n}(x) \rightarrow I_{\alpha | \mathcal{A}_\infty}(x)$ dans L^1 .

En intégrant, nous obtenons le résultat pour l'entropie associée, i.e. $H(\alpha | \mathcal{A}_n)$ converge vers $H(\alpha | \mathcal{A}_\infty)$ quand $n \rightarrow +\infty$. $\square$

5. Preuve du théorème de Shanon-Breiman-Mac Millan

Donnons tout d'abord une définition équivalente de $h_\mu(T, \alpha)$.

Proposition 111. *Soient $(X, \mathcal{B}, \mu, T)$ un s.d.m. et α une partition mesurable dénombrable. Alors*

$$h_\mu(T, \alpha) = \lim_{n \rightarrow +\infty} H(\alpha | \bigvee_{i=0}^{n-1} T^{-i} \alpha) = H(\alpha | \alpha^-),$$

où $\alpha^- = \bigvee_{i \geq 1} T^{-i} \alpha$.

DÉMONSTRATION. En utilisant la Proposition 92 et l'invariance de H par T , nous obtenons

$$\begin{aligned}
H\left(\bigvee_{i=0}^{n-1} T^{-i}\alpha\right) &= H\left(\alpha \mid \bigvee_{i=1}^{n-1} T^{-i}\alpha\right) + H\left(\bigvee_{i=0}^{n-2} T^{-i}\alpha\right) \\
&= H\left(\alpha \mid \bigvee_{i=1}^{n-1} T^{-i}\alpha\right) + H\left(\alpha \mid \bigvee_{i=1}^{n-2} T^{-i}\alpha\right) + H\left(\bigvee_{i=0}^{n-3} T^{-i}\alpha\right) \\
&\dots \\
&= H(\alpha) + \sum_{r=1}^n H\left(\alpha \mid \bigvee_{i=1}^{r-1} T^{-i}\alpha\right).
\end{aligned}$$

D'autre part, nous avons par le corollaire 110 appliqué à $\mathcal{A}_n = \sigma(\bigvee_{i=0}^{n-1} T^{-i}\alpha)$ que la suite $H(\alpha \mid \bigvee_{i=1}^{n-1} T^{-i}\alpha)$ converge, quand n tend vers l'infini, vers $H(\alpha \mid \alpha^-)$.

Puisque la convergence implique la convergence en moyenne de Cesaro, par les égalités précédentes, nous avons

$$\lim_{n \rightarrow +\infty} \frac{1}{n} H\left(\bigvee_{i=0}^{n-1} T^{-i}\alpha\right) = \lim_{n \rightarrow +\infty} H\left(\alpha \mid \bigvee_{i=0}^{n-1} T^{-i}\alpha\right) = H(\alpha \mid \alpha^-).$$

□

Faisons quelques commentaires. Supposons $h_\mu(T) = 0$. Alors, toute partition α dénombrable et mesurable d'entropie finie est telle que $H_\mu(\alpha \mid \alpha^-) = 0$. Ceci implique que α est contenue dans α^- et donc que le s.d.m. est en quelque sorte "déterministe".

De plus $T^{-1}\mathcal{B} = \mathcal{B} \pmod{\mu}$. En effet, supposons qu'il existe $B \in \mathcal{B} \setminus T^{-1}\mathcal{B}$ de mesure non-nulle. Considérons $\alpha = \{B, B^c\}$. D'après ce qui précède nous devons avoir $\alpha \subseteq \alpha^-$. Or, $\alpha^- \subseteq T^{-1}\mathcal{B}$, donc $\alpha \subseteq T^{-1}\mathcal{B}$. D'où $B \in T^{-1}\mathcal{B}$, contradiction.

Nous sommes alors en mesure de démontrer le théorème principal de cette section :

Théorème 112 (Shannon-Mac Millan-Breiman).

Soient $(X, \mathcal{B}, \mu, T)$ un s.d.m. et α une partition mesurable et dénombrable telle que $h_\mu(T, \alpha) < \infty$. Alors,

$$\lim_{N \rightarrow \infty} \frac{1}{N} I_{\bigvee_{n=0}^{N-1} T^{-n}\alpha} = \mathbb{E}(f \mid \mathcal{I}) \quad \mu - p.p. \text{ et dans } L^1(X, \mathcal{B}, \mu),$$

où $f = I_{\alpha \mid \alpha^-}$ et $\mathcal{I}$ désigne la tribu engendrée par les ensembles T invariants ($T^{-1}B = B$).

Utilisons le, à présent, pour prouver le Corollaire 98.

DÉMONSTRATION DU COROLLAIRE 98. En appliquant le théorème de Shannon-Mc Millan-Breiman, l'ergodicité de la mesure implique que $\mathcal{I}$ est la tribu grossière et donc que $\mathbb{E}(f \mid \mathcal{I})$ est constante et vaut

$$\int_X \mathbb{E}(f \mid \mathcal{I}) d\mu = \int_X f d\mu = H(\alpha \mid \alpha^-) = h_\mu(T, \alpha).$$

□

Le théorème de Shanon-McMillan-Breiman découle en fait, du théorème ergodique.

DÉMONSTRATION DU THÉORÈME 112. Posons $f_0 = I_\alpha$, $f_k = I_{\alpha|\bigvee_{n=1}^k T^{-n}\alpha}$ et $f = I_{\alpha|\alpha^-}$. Nous avons

$$\begin{aligned} I_{\bigvee_{n=0}^{N-1} T^{-n}\alpha} &= I_\alpha \bigvee_{n=1}^{N-1} T^{-n}\alpha = I_{\alpha|\bigvee_{n=1}^{N-1} T^{-n}\alpha} + I_{\bigvee_{n=1}^{N-1} T^{-n}\alpha} \\ &= f_{N-1} + I_{T^{-1}(\bigvee_{n=0}^{N-2} T^{-n}\alpha)} = f_{N-1} + I_{\bigvee_{n=0}^{N-2} T^{-n}\alpha} \circ T \\ &= f_{N-1} + f_{N-2} \circ T + \cdots + f_0 \circ T^{N-1}. \end{aligned}$$

Ainsi

$$\begin{aligned} &\left| \frac{1}{N} I_{\bigvee_{n=0}^{N-1} T^{-n}\alpha} - \mathbb{E}(f | \mathcal{I}) \right| = \left| \frac{1}{N} \sum_{k=0}^{N-1} f_{N-1-k} \circ T^k - \mathbb{E}(f | \mathcal{I}) \right| \\ &\leq \left| \frac{1}{N} \sum_{k=0}^{N-1} (f_{N-1-k} \circ T^k - f \circ T^k) \right| + \left| \frac{1}{N} \sum_{k=0}^{N-1} f \circ T^k - \mathbb{E}(f | \mathcal{I}) \right| \end{aligned}$$

Par hypothèse $H_\mu(\alpha) < \infty$. Donc, d'après le Lemme de Chung et le corollaire 110, f appartient à $L^1(X, \mathcal{B}, \mu)$. Le Théorème de Birkhoff implique que

$$\left| \frac{1}{N} \sum_{k=0}^{N-1} f \circ T^k - \mathbb{E}(f | \mathcal{I}) \right|$$

converge vers 0 μ -presque partout et dans $L^1(X, \mathcal{B}, \mu)$.

Étudions l'autre terme. Posons $g_k = |f_k - f|$. Nous avons

$$(1) \quad \left| \frac{1}{N} \sum_{k=0}^{N-1} (f_{N-1-k} - f) \circ T^k \right| \leq \frac{1}{N} \sum_{k=0}^{N-1} g_{N-1-k} \circ T^k.$$

Montrons que le second terme converge vers 0 μ -presque partout et dans $L^1(X, \mathcal{B}, \mu)$. Grâce au corollaire 110, nous observons que $f_k = I_{\alpha|\bigvee_{n=1}^k T^{-n}\alpha}$ converge vers $f = I_{\alpha|\alpha^-}$ μ -p.p. et dans L^1 . Ainsi, $\lim_{k \rightarrow \infty} g_k = 0$ μ -presque partout et dans $L^1(X, \mathcal{B}, \mu)$. Prouvons la convergence dans $L^1(X, \mathcal{B}, \mu)$ du second terme de (1). Il suffit de montrer que son intégrale tend vers 0. Nous avons

$$\begin{aligned} \frac{1}{N} \int_X \sum_{k=0}^{N-1} g_{N-1-k} \circ T^k d\mu &= \frac{1}{N} \sum_{k=0}^{N-1} \int_X g_{N-1-k} d\mu \\ &= \frac{1}{N} \sum_{k=0}^{N-1} \int_X g_k d\mu \xrightarrow{N \rightarrow \infty} 0 \end{aligned}$$

Prouvons la convergence μ -presque sûre du second terme de (1). Nous avons

$$G_N = \sup_{k \geq N} g_k \leq \left(\sup_{k \geq 1} f_k \right) + f \in L^1(X, \mathcal{B}, \mu),$$

par le Lemme de Chung. Fixons M . Alors,

$$\begin{aligned} \frac{1}{N} \sum_{k=0}^{N-1} g_{N-1-k} \circ T^k &= \frac{1}{N} \sum_{k=0}^{N-M-1} g_{N-1-k} \circ T^k + \frac{1}{N} \sum_{k=N-M}^{N-1} g_{N-1-k} \circ T^k \\ &\leq \frac{1}{N} \sum_{k=0}^{N-M-1} G_M \circ T^k + \frac{1}{N} \sum_{k=N-M}^{N-1} g_{N-1-k} \circ T^k. \end{aligned}$$

D'après le Lemme de Chung, $\sup_{k \geq 1} f_k$ appartient à $L^1(X, \mathcal{B}, \mu)$ et est donc fini μ presque partout. Comme $\frac{1}{N} \sum_{k=N-M}^{N-1} g_{N-1-k} \circ T^k$ est une moyenne d'une somme de M termes chacun majoré par $\sup_{k \geq 1} f_k$, nous avons

$$\lim_{N \rightarrow +\infty} \frac{1}{N} \sum_{k=N-M}^{N-1} g_{N-1-k} \circ T^k = 0 \text{ } \mu\text{-presque partout.}$$

Le Théorème de Birkhoff implique que

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=0}^{N-M-1} G_M \circ T^k = \int_X G_M d\mu.$$

Remarquons que G_M tend vers 0 μ -presque partout. Le théorème de convergence dominée de Lebesgue entraîne que $\int_X G_M d\mu$ tend vers 0. Ceci achève la preuve. $\square$

Corollaire 113. *Soit $(X, \mathcal{B}, \mu, T)$ un s.d.m. ergodique et α une partition dénombrable et mesurable telle que $H_\mu(\alpha) < \infty$. Pour tout $\epsilon > 0$ il existe $N(\epsilon)$ tel que pour tout $N \geq N(\epsilon)$, la famille d'atomes*

$$\Gamma_N(\alpha) = \left\{ A \in \bigvee_{n=0}^{N-1} T^{-n}\alpha : e^{-N(h_\mu(\alpha, T) + \epsilon)} \leq \mu(A) \leq e^{-N(h_\mu(\alpha, T) - \epsilon)} \right\}$$

vérifie $\mu\left(\bigcup_{A \in \Gamma_N(\alpha)} A\right) > 1 - \epsilon$.

DÉMONSTRATION. D'après le Théorème de Shannon-Mac Millan-Breiman, $\frac{1}{N} I_{\bigvee_{n=0}^{N-1} T^{-n}\alpha} \xrightarrow{N \rightarrow \infty} h_\mu(\alpha, T)$ μ -presque partout. Soit $\epsilon > 0$, il existe $N(\epsilon)$ tel que pour tout $N \geq N(\epsilon)$

$$\mu \left\{ x \in X; \left| \frac{1}{N} I_{\bigvee_{n=0}^{N-1} T^{-n}\alpha}(x) - h_\mu(\alpha, T) \right| > \epsilon \right\} < \epsilon.$$

Remarquons que $I_{\bigvee_{n=0}^{N-1} T^{-n}\alpha}$ est constante sur les atomes de $\bigvee_{n=0}^{N-1} T^{-n}\alpha$: elle vaut $-\log(A)$ sur $A \in \bigvee_{n=0}^{N-1} T^{-n}\alpha$. Par conséquent il suffit de poser

$$\Gamma_N(\alpha) = \left\{ A \in \bigvee_{n=0}^{N-1} T^{-n}\alpha; \left| -\frac{1}{N} \log \mu(A) - h_\mu(\alpha, T) \right| \leq \epsilon \right\}.$$

□

6. Preuves des théorèmes d'Abramov et de Sinaiï

Lemme 114. *Soit $(X, \mathcal{B}, \mu, T)$ un s.d.m.. Si α et β sont deux partitions mesurables et dénombrables d'entropies finies alors*

$$h_\mu(\alpha, T) \leq h_\mu(\beta, T) + H(\alpha | \beta) \text{ et} \\ |h_\mu(\alpha, T) - h_\mu(\beta, T)| \leq H_\mu(\alpha | \beta) + H(\beta | \alpha).$$

DÉMONSTRATION. Nous avons

$$H \left(\bigvee_{i=0}^{n-1} T^{-i}\alpha \right) \leq H \left(\bigvee_{i=0}^{n-1} T^{-i}\alpha \vee \bigvee_{i=0}^{n-1} T^{-i}\beta \right) \\ = H \left(\bigvee_{i=0}^{n-1} T^{-i}\beta \right) + H \left(\bigvee_{i=0}^{n-1} T^{-i}\alpha \mid \bigvee_{i=0}^{n-1} T^{-i}\beta \right),$$

par la proposition 92. Nous estimons ensuite

$$H \left(\bigvee_{i=0}^{n-1} T^{-i}\alpha \mid \bigvee_{i=0}^{n-1} T^{-i}\beta \right) \\ = H \left(\alpha \mid \bigvee_{i=0}^{n-1} T^{n-1}\beta \right) + H \left(\bigvee_{i=1}^{n-1} T^{-i}\alpha \mid \alpha \vee \bigvee_{i=0}^{n-1} T^{-i}\beta \right) \\ \leq H(\alpha | \beta) + H \left(\bigvee_{i=1}^{n-1} T^{-i}\alpha \mid \bigvee_{i=1}^{n-1} T^{-i}\beta \right) \\ = H(\alpha | \beta) + H \left(\bigvee_{i=0}^{n-2} T^{-i}\alpha \mid \bigvee_{i=0}^{n-2} T^{-i}\beta \right).$$

En procédant de façon inductive sur n , nous obtenons

$$H \left(\bigvee_{i=0}^{n-1} T^{-i} \alpha \mid \bigvee_{i=0}^{n-1} T^{-i} \beta \right) \leq nH(\alpha \mid \beta)$$

Finalement avec l'inégalité du début, nous voyons que

$$H \left(\bigvee_{i=0}^{n-1} T^{-i} \alpha \right) \leq H \left(\bigvee_{i=0}^{n-1} T^{-i} \beta \right) + nH(\alpha \mid \beta).$$

En divisant par n , puis en faisant tendre n vers $+\infty$, on obtient le résultat voulu. L'autre inégalité en découle. $\square$

Théorème 115 (Abramov). *Soient $(X, \mathcal{B}, \mu, T)$ un s.d.m. et $(\alpha_n; n \in \mathbb{N})$ une suite croissante de partition d'entropie finie convergeant vers $\mathcal{B} \pmod{\mu}$. Alors,*

$$h_\mu(T) = \lim_{n \rightarrow \infty} h_\mu(T, \alpha_n).$$

DÉMONSTRATION. Soit δ une partition d'entropie finie. En raison du lemme précédent nous avons

$$h_\mu(T, \delta) \leq h_\mu(T, \alpha_n) + H(\delta \mid \alpha_n).$$

D'après le corollaire au Lemme de Chung, $(H(\delta \mid \alpha_n))_n$ tend vers $H(\delta \mid \mathcal{B})$, quand $n \rightarrow +\infty$. Ainsi, $h_\mu(T) \leq \lim_{n \rightarrow \infty} h_\mu(T, \alpha_n) \leq h_\mu(T)$. $\square$

Rappelons la définition de générateur.

Définition 116. *Soit $(X, \mathcal{B}, \mu, T)$ un s.d.m.. Une partition α dénombrable et mesurable d'entropie finie est un générateur fort si $\alpha^- \vee \sigma(\alpha) = \mathcal{B} \pmod{\mu}$.*

Si T est inversible, α est un générateur si $\alpha_T = \mathcal{B} \pmod{\mu}$, où $\alpha_T = \bigvee_{i \in \mathbb{Z}} T^i \alpha$.

Lemme 117. *Soit $(X, \mathcal{B}, \mu, T)$ un s.d.m.. Si α et γ sont des partitions d'entropie finie, alors*

- (1) *Si $\sigma(\alpha) \vee \alpha^- \supseteq \sigma(\gamma)$ alors $h_\mu(T, \gamma) \leq h_\mu(T, \alpha)$ et*
- (2) *pour T inversible, si $T\alpha_T \supseteq \sigma(\gamma)$ alors $h_\mu(T, \gamma) \leq h_\mu(T, \alpha)$.*

DÉMONSTRATION. Supposons $\gamma \preceq \alpha^- \vee \alpha$. D'après le lemme 114, nous avons

$$h_\mu(T, \gamma) \leq h_\mu \left(T, \bigvee_{k=0}^{n-1} T^{-k} \alpha \right) + H \left(\gamma \mid \bigvee_{k=0}^{n-1} T^{-k} \alpha \right)$$

Soit $\epsilon > 0$. Puisque $\bigvee_{k=0}^{n-1} T^{-k} \alpha$ tend vers $\alpha \vee \alpha^-$ quand n tend vers $+\infty$ et que γ est $\alpha \vee \alpha^-$ -mesurable, par le corollaire 110, il existe $n(\epsilon) \in \mathbb{N}$ tel que pour tout $n \geq n(\epsilon)$, $H_\mu \left(\gamma \mid \bigvee_{k=0}^{n-1} T^{-k} \alpha \right) < \epsilon$. Soit $n \geq n(\epsilon)$. On remarque que

$$\begin{aligned}
h_\mu \left(T, \bigvee_{k=0}^{n-1} T^{-k} \alpha \right) &= \lim_{N \rightarrow \infty} \frac{1}{N} H_\mu \left(\bigvee_{l=0}^{N-1} T^{-l} \left(\bigvee_{k=0}^{n-1} T^{-k} \alpha \right) \right) \\
&= \lim_{N \rightarrow \infty} \frac{1}{N} H_\mu \left(\bigvee_{n=0}^{(N+n-2)} T^{-n} \alpha \right) = h_\mu(T, \alpha).
\end{aligned}$$

d'où l'on conclut que $h_\mu(T, \gamma) \leq h_\mu(T, \alpha) + \epsilon$ pour tout $\epsilon > 0$. Ce qui permet de conclure. L'autre partie du lemme se prouve de façon analogue. $\square$

On en déduit directement le théorème 97 de Sinaiï.

Relation entre l'entropie topologique et l'entropie mesurée

Dans ce chapitre, X est un espace métrique compact et $T : X \rightarrow X$ une application continue. La tribu des boréliens de X sera notée $\mathcal{B}(X)$.

Dans cette partie, nous prouvons la relation simple qui existe entre l'entropie topologique et l'entropie mesurée : si T est une fonction continue d'un espace métrique compact, alors

$$h_\mu(T) = \sup\{h_\mu(T) \mid \mu \in M(X, T)\}.$$

L'inégalité $\sup\{h_\mu(T) \mid \mu \in M(X, T)\} \leq h(T)$ a été prouvé en 1968 par L.W. Goodwyn. En 1970, T.N.T. Goodman a prouvé l'égalité dans le cas général. Dans le lemme suivant, ∂B désigne l'ensemble $\partial B = \bar{B} \setminus \overset{\circ}{B}$.

Lemme 118. *Soient X un espace métrique compact et $\mu \in M(X, T)$.*

- (1) *Si $x \in X$ et $\delta > 0$, alors il existe $\delta' < \delta$ tel que $\mu(\partial B(x, \delta')) = 0$.*
- (2) *Si $\delta > 0$, alors il existe une partition finie $\xi = \{A_1, \dots, A_k\}$ de $(X, \mathcal{B}(X))$ telle que $\text{diam}(A_j) < \delta$ et $\mu(\partial A_j) = 0$ pour chaque j .*

DÉMONSTRATION. Le (1) est laissé en exercice. Montrons (2). D'après (1), il existe un recouvrement ouvert fini $\beta = \{B_1, \dots, B_r\}$ de X par les boules de rayon inférieur à $\delta/2$ avec $\mu(\partial B_j) = 0$ pour tout j . Posons $A_1 = \bar{B}_1$ et pour tout $n > 1$, $A_n = \bar{B}_n \setminus (\bar{B}_1 \cup \bar{B}_2 \cup \dots \cup \bar{B}_{n-1})$. Alors $\xi = \{A_1, \dots, A_k\}$ est une partition de $(X, \mathcal{B}(X))$ avec $\text{diam}(A_n) < \delta$, et puisque $\partial A_n \subset \bigcup_{i=1}^n \partial B_i$ nous avons $\mu(\partial A_n) = 0$ pour tout n . $\square$

Avant de montrer le principe variationnel, nous allons rassembler un certain nombre de remarques que nous utiliserons pour le montrer. Dans ce qui suit, X est un espace métrique compact et $\mathcal{B}(X)$ la tribu des boréliens.

Remarques.

- (1) Si $\mu_i \in M(X)$, $1 \leq i \leq n$, et $p_i \geq 0$, $\sum_{i=1}^n p_i = 1$, alors

$$H_{\sum_{i=1}^n p_i \mu_i}(\xi) \geq \sum_{i=1}^n p_i H_{\mu_i}(\xi)$$

pour une toute partition finie ξ de $(X, \mathcal{B}(X))$. En fait, il suffit de le montrer pour deux mesures (**Exercice**). Prenons deux mesures $\mu, m \in M(X, T)$ et $p \in [0, 1]$. Puisque $\phi(x) = x \log(x)$ est convexe, nous obtenons que si $B \in \mathcal{B}(X)$ alors

$$\phi(p\mu(B) + (1-p)m(B)) \leq p\phi(\mu(B)) + (1-p)\phi(m(B)).$$

Donc, si ξ est une partition finie quelconque de $(X, \mathcal{B}(X))$, alors

$$H_{p\mu+(1-p)m}(\xi) \geq pH_\mu(\xi) + (1-p)H_m(\xi).$$

(2) Soient q et n des entiers positifs avec $1 \leq q \leq n$. Pour $0 \leq j \leq q-1$ posons $a(j) = [(n-j)/q]$, où $[b]$ désigne la partie intégrale de $b > 0$. Nous avons :

(a) $a(0) \geq a(1) \geq \dots \geq a(q-1)$.

(b) Fixons $0 \leq j \leq q-1$. Alors

$$\{0, 1, \dots, n-1\} = \{j + rq + i; 0 \leq r \leq a(j) - 1, 0 \leq i \leq q-1\} \cup S$$

où $S = \{0, 1, \dots, j-1, j + a(j)q, j + a(j)q + 1, \dots, n-1\}$. De plus, $j + a(j)q \geq j + [(n-j)/q - 1]q = n - q$, donc $\#S \leq 2q$.

(c) Pour chaque $0 \leq j \leq q-1$, $(a(j) - 1)q + j \leq [(n-j)/q - 1]q + j = n - q$.

Les nombres $\{j + rq; 0 \leq j \leq q-1, 0 \leq r \leq a(j) - 1\}$ sont tous distincts et ne sont pas plus grands que $n - q$.

(3) Si $\mu \in M(X, T)$ et si $\mu(\partial A_i) = 0$, $0 \leq i \leq n-1$, alors

$$\mu \left(\partial \left(\bigcap_{i=0}^{n-1} T^{-i} A_i \right) \right) = 0$$

car $\partial \left(\bigcap_{i=0}^{n-1} T^{-i} A_i \right) \subset \bigcup_{i=0}^{n-1} T^{-i} \partial A_i$.

Théorème 119. Soit $T : X \rightarrow X$ une fonction continue d'un espace métrique compact X . Alors $h(T) = \sup\{h_\mu(T) | \mu \in M(X, T)\}$.

DÉMONSTRATION. Soit $\mu \in M(X, T)$. Montrons $h_\mu(T) \leq h(T)$. Considérons $\xi = \{A_1, \dots, A_k\}$ une partition finie de $(X, \mathcal{B}(X))$. Choisissons $\varepsilon > 0$ tel que $\varepsilon < 1/(k \log k)$. La mesure μ étant régulière, il existe des ensembles compacts $B_j \subset A_j$, $1 \leq j \leq k$, avec $\mu(A_j \setminus B_j) < \varepsilon$. Soit η la partition $\{B_0, \dots, B_k\}$, où $B_0 = X \setminus \bigcup_{j=1}^k B_j$. Nous avons $\mu(B_0) < k\varepsilon$ et, en utilisant les propriétés de ϕ ,

$$\begin{aligned}
H_\mu(\xi|\eta) &= - \sum_{i=0}^k \sum_{j=1}^k \mu(B_i) \phi \left(\frac{\mu(B_i \cap A_j)}{\mu(B_i)} \right) \\
&= -\mu(B_0) \sum_{j=1}^k \phi \left(\frac{\mu(B_0 \cap A_j)}{\mu(B_0)} \right) \text{ car si } i \neq 0, \frac{\mu(B_i \cap A_j)}{\mu(B_i)} = 0 \text{ ou } 1 \\
&= -k\mu(B_0) \sum_{j=1}^k \frac{1}{k} \phi \left(\frac{\mu(B_0 \cap A_j)}{\mu(B_0)} \right) \leq -k\mu(B_0) \phi \left(\sum_{j=1}^k \frac{\mu(B_0 \cap A_j)}{k\mu(B_0)} \right) \\
&= -k\mu(B_0) \phi \left(\frac{1}{k} \right) = \mu(B_0) \log(k) < k\varepsilon \log(k) < 1.
\end{aligned}$$

Remarquons que pour chaque $i \neq 0$, $B_0 \cup B_i = X \setminus \bigcup_{j \neq i} B_j$ est ouvert, et donc $\beta = \{B_0 \cup B_1, \dots, B_0 \cup B_k\}$ est un recouvrement ouvert de X . Posons $N = N \left(\bigvee_{i=0}^{n-1} T^{-i}\eta \right)$, le nombre d'ensembles non-vides dans la partition $\bigvee_{i=0}^{n-1} T^{-i}\eta$. Nous avons, pour $n \geq 1$,

$$\begin{aligned}
H_\mu \left(\bigvee_{i=0}^{n-1} T^{-i}\eta \right) &= \sum_{A \in \bigvee_{i=0}^{n-1} T^{-i}\eta} -\phi(\mu(A)) \\
&= N \sum_{A \in \bigvee_{i=0}^{n-1} T^{-i}\eta} \frac{1}{N} (-\phi)(\mu(A)) \\
&\leq N(-\phi) \left(\sum_{A \in \bigvee_{i=0}^{n-1} T^{-i}\eta} \frac{1}{N} \mu(A) \right) \\
&= N(-\phi) \left(\frac{1}{N} \right) = \log N,
\end{aligned}$$

Montrons que $N \left(\bigvee_{i=0}^{n-1} T^{-i}\eta \right) \leq 2^n N \left(\bigvee_{i=0}^{n-1} T^{-i}\beta \right)$. Soit α un sous-recouvrement de $\bigvee_{i=0}^{n-1} T^{-i}\beta$ de cardinalité $N(\alpha) = N \left(\bigvee_{i=0}^{n-1} T^{-i}\beta \right)$. On peut décrire α de la façon suivante :

$$\alpha = \left\{ \bigcap_{i=0}^{n-1} T^{-i} (B_0 \cup B_{s(i,j)}) ; 1 \leq j \leq N(\alpha) \right\},$$

où $\{s(i,j); 0 \leq i \leq n-1, 1 \leq j \leq N(\alpha)\} \subset \{1, \dots, k-1\}$. Or pour chaque j on peut écrire l'élément de α correspondant de la façon suivante :

$$\bigcup_{l=1}^{2^n} \bigcap_{i=0}^{n-1} T^{-i} B_{t(i,j,l)},$$

où $t(i, j, l) \in \{0, s(i, j)\}$. Par conséquent

$$\alpha' = \left\{ \bigcap_{i=0}^{n-1} T^{-i} (B_{t(i,j,l)}) ; 1 \leq l \leq 2^n, 1 \leq j \leq N(\alpha) \right\}$$

est un recouvrement de X tel que

$$N(\alpha') \leq 2^n N(\alpha).$$

De plus, $\bigvee_{i=0}^{n-1} T^{-i} \eta \preceq \alpha'$, par conséquent,

$$H_\mu \left(\bigvee_{i=0}^{n-1} T^{-i} \eta \right) \leq \log N \left(\bigvee_{i=0}^{n-1} T^{-i} \beta \right) + n \log 2$$

puis

$$h_\mu(T, \eta) \leq h(T, \beta) + \log 2 \leq h(T) + \log 2.$$

Ainsi, d'après le Lemme 114 nous avons

$$h_\mu(T, \xi) \leq h_\mu(T, \eta) + H_\mu(\xi|\eta) \leq h(T) + \log 2 + 1,$$

puis $h_\mu(T) \leq h(T) + \log 2 + 1$. Ceci étant vrai pour toute autre application continue laissant μ invariante, c'est également vrai pour les applications T^n . Ceci donne, grâce aux théorèmes 101 et 78, $nh_\mu(T) \leq nh(T) + \log 2 + 1$. D'où $h_\mu(T) \leq h(T)$.

Soit $\varepsilon > 0$. Montrons que nous pouvons trouver $\mu \in M(X, T)$ telle que $h_\mu(T) \geq s(\varepsilon, X, T)$. Ce qui impliquera que $\sup\{h_\mu(T) | \mu \in M(X, T)\} \geq h(T)$.

Soient E_n un ensemble (n, ε) -séparé pour X de cardinalité $s_n(\varepsilon, X)$ et $\sigma_n \in M(X)$ la mesure atomique concentrée de façon uniforme sur les points de E_n : $\sigma_n = (1/s_n(\varepsilon, X)) \sum_{x \in E_n} \sigma_x$.

Considérons $\mu_n \in M(X)$ définie par $\mu_n = (1/n) \sum_{i=0}^{n-1} \sigma_n \circ T^{-i}$. Puisque $M(X)$ est compact, nous pouvons choisir une suite strictement croissante (n_j) d'entiers telle que $\lim_{n \rightarrow \infty} \frac{1}{n_j} \log s_{n_j}(\varepsilon, X) = s(\varepsilon, X, T)$ et $(\mu_{n_j})_j$ converge dans $M(X)$ vers $\mu \in M(X)$.

D'après la Proposition 28, nous savons que $\mu \in M(X, T)$. Nous allons montrer $h_\mu(T) \geq s(\varepsilon, X, T)$. Le Lemme 118 nous assure que nous pouvons choisir une partition $\xi = \{A_1, \dots, A_k\}$ de $(X, \mathcal{B}(X))$ telle que $\text{diam}(A_i) < \varepsilon$ et $\mu(\partial A_i) = 0$ pour $1 \leq i \leq k$. Les éléments de $\bigvee_{i=0}^{n-1} T^{-i} \xi$ contiennent au plus un élément de E_n , donc il y a exactement $s_n(\varepsilon, X)$ éléments de $\bigvee_{i=0}^{n-1} T^{-i} \xi$ de σ_n -mesure $1/s_n(\varepsilon, X)$, les autres étant de σ_n -mesure nulle. Par conséquent $H_{\sigma_n} \left(\bigvee_{i=0}^{n-1} T^{-i} \xi \right) = \log s_n(\varepsilon, X)$.

Fixons des entiers positifs q, n avec $1 < q < n$ et, comme dans la remarque (2), définissons $a(j)$, pour $0 \leq j \leq q-1$, par $a(j) = \lfloor (n-j)/q \rfloor$.

Soit $0 \leq j \leq q-1$. D'après la Remarque (2) (b) nous avons :

$$\bigvee_{i=0}^{n-1} T^{-i}\xi = \left(\bigvee_{r=0}^{a(j)-1} T^{-(rq+j)} \left(\bigvee_{i=0}^{q-1} T^{-i}\xi \right) \right) \vee \bigvee_{l \in S} T^{-l}\xi$$

et S à une cardinalité au plus égale à $2q$. Par conséquent,

$$\begin{aligned} \log s_n(\varepsilon, X) &= H_{\sigma_n} \left(\bigvee_{i=0}^{n-1} T^{-i}\xi \right) \\ &\leq \sum_{r=0}^{a(j)-1} H_{\sigma_n} \left(T^{-(rq+j)} \bigvee_{i=0}^{q-1} T^{-i}\xi \right) + \sum_{l \in S} H_{\sigma_n} (T^{-l}\xi) \\ &\leq \sum_{r=0}^{a(j)-1} H_{\sigma_n \circ T^{-(rq+j)}} \left(\bigvee_{i=0}^{q-1} T^{-i}\xi \right) + 2q \log(k). \end{aligned}$$

En sommant sur j de 0 à $q-1$ et en utilisant la Remarque (2) (c) nous obtenons

$$q \log s_n(\varepsilon, X) \leq \sum_{p=0}^{n-1} H_{\sigma_n \circ T^{-p}} \left(\bigvee_{i=0}^{q-1} T^{-i}\xi \right) + 2q^2 \log(k).$$

En divisant par n et en utilisant la Remarque (1) nous obtenons

$$(2) \quad \frac{q}{n} \log s_n(\varepsilon, X) \leq H_{\mu_n} \left(\bigvee_{i=0}^{q-1} T^{-i}\xi \right) + \frac{2q^2}{n} \log(k). (*)$$

Grâce à la Remarque (3) nous savons que les éléments de $\bigvee_{i=0}^{q-1} T^{-i}\xi$ ont des frontières de mesure nulle. Ainsi $\lim_{j \rightarrow \infty} \mu_{n_j}(B) = \mu(B)$ pour chaque élément B de $\bigvee_{i=0}^{q-1} T^{-i}\xi$ (Remarque (4)) et par conséquent

$$\lim_{j \rightarrow \infty} H_{\mu_{n_j}} \left(\bigvee_{i=0}^{q-1} T^{-i}\xi \right) = H_{\mu} \left(\bigvee_{i=0}^{q-1} T^{-i}\xi \right).$$

En remplaçant n par n_j dans l'inégalité (2) et en faisant tendre j vers $+\infty$ nous en déduisons

$$qs(\varepsilon, X, T) \leq H_{\mu} \left(\bigvee_{i=0}^{q-1} T^{-i}\xi \right).$$

En divisant par q et en faisant tendre q vers $+\infty$ nous concluons $s(\varepsilon, X, T) \leq h_{\mu}(T, \xi) \leq h_{\mu}(T)$. $\square$

Bibliographie

- [D] Devaney, *An introduction to chaotic dynamical systems*. The Benjamin/Cummings Publishing Co., Inc., Menlo Park, CA, 1986. 320 pp.
- [KH] A. Katok, B. Hasselblatt, *Introduction to the modern theory of dynamical systems*, Cambridge University Press, Cambridge, 1995.
- [Pa] W. Parry, *Topics in Ergodic Theory*, Cambridge University Press 1981.
- [Pe] K. Petersen, *Ergodic Theory*, Cambridge Studies in Advanced Mathematics 2, 1983.
- [Q] Queffélec, M. *Substitution dynamical systems – spectral analysis*. Lecture Notes in Mathematics, 1294. Springer-Verlag, Berlin, 1987.
- [Wa] P. Walters, *An Introduction to Ergodic Theory*, Graduate Texts in Mathematics 79, Springer-Verlag 1982.
- [Wi] D. Williams, *Probability with martingales*, Cambridge Mathematical Textbooks, Cambridge University Press, Cambridge, 1991.
- [Z] Zinsmeister, M. *Formalisme thermodynamique et systèmes dynamiques holomorphes*. Panoramas et Synthèses , 4. Société Mathématique de France, Paris, 1996.