

Systemes dynamiques élémentaires

Yves Benoist

Frédéric Paulin

Cours aux carrés

1 Introduction

Ces notes correspondent à un cours à l'Ecole Normale Supérieure du premier auteur les années 2000-2001 et 2001-2002 et du second auteur en 2002-2003.

Nous renvoyons par exemple à l'article [Yoc] pour donner une petite idée de tout ce dont nous ne traiterons pas dans le domaine des systèmes dynamiques. Nous renvoyons à [Sin2, Lect. 1] pour les grands types de problèmes qui se posent dans la théorie des systèmes dynamiques.

En mécanique classique, on étudie l'évolution au cours du temps de certains systèmes physiques comme une toupie, un gaz, une étoile et ses planètes ... Si, au temps initial $t = 0$, le système est représenté par un point x de l'espace des phases X , alors au temps t , ce système est représenté par un point $f^t(x)$.

Lorsque les équations différentielles qui régissent ce mouvement sont indépendantes du temps, on a l'égalité $f^{t+t'} = f^t \circ f^{t'}$ pour tous t, t' . On dit que f est un groupe à un paramètre de transformations de X .

Lorsque l'on ne dispose pas de formule explicite pour f^t , on cherche à comprendre le comportement de $f^t(x)$ pour t grand. Une telle étude qualitative a été initiée par Poincaré.

Dans les exemples issus de la mécanique céleste, il arrive souvent, à cause de la conservation de l'énergie, que l'espace des phases soit compact et qu'il existe sur X une mesure finie invariante par f^t .

Pour clarifier les phénomènes, certains mathématiciens sont sortis du cadre des équations différentielles, en ne gardant que l'espace X et le groupe à un paramètre de transformations $(f^t)_{t \in \mathbb{R}}$ [système dynamique continu] ou encore, que l'espace X et une transformation f (par toujours inversible) de l'espace X [système dynamique discret]. Dans ce dernier cas, on s'intéresse au comportement asymptotique des itérations f^n de f . Pour mener l'étude qualitative du comportement de $f^t(x)$ pour t grand, deux cadres s'avèrent particulièrement bien adaptés : celui de la topologie [système dynamique topologique] et celui de la théorie de la mesure [théorie ergodique].

Nous étudierons ces deux points de vue et leurs interactions. Bien plus que de clarifier les idées de ce sujet, ces cadres plus larges et plus naturels ont permis leur application à d'autres domaines des mathématiques (théorie des nombres, théorie des groupes). Nous en verrons quelques-unes.

Remerciements : Nous remercions C. Wormser pour ses nombreuses corrections sur une première version de ce texte.

Table des matières

1	Introduction	2
2	Exemples fondamentaux	6
2.1	Systèmes dynamiques topologiques et mesurables	6
2.2	Systèmes de Kronecker	8
2.3	Systèmes symboliques ou de Bernoulli	8
2.4	Mesures de Liouville et systèmes hamiltoniens	10
2.5	Billards	12
2.6	Exercices	14
2.7	Indications pour la résolution des exercices	16
3	Récurrence	18
3.1	Récurrence et minimalité	18
3.2	Récurrence multiple	19
3.3	Le théorème de récurrence de Poincaré	22
3.4	Exercices	23
3.5	Indications pour la résolution des exercices	25
4	Ergodicité	28
4.1	Transformations ergodiques	28
4.2	Le théorème ergodique	30
4.3	Mesures invariantes et mesures ergodiques	33
4.4	Exercices	35
4.5	Indications pour la résolution des exercices	37
5	Unique ergodicité	39
5.1	Unique ergodicité	39
5.2	Unique ergodicité des translations sur le tore	40
5.3	Equirépartition modulo 1 de la suite $P(n)$	41
5.4	Appendice: critère de Weyl et lemme de van der Corput	42
5.5	Exercices	45
5.6	Indications pour la résolution des exercices	45
6	Mélange	46
6.1	Transformations mélangeantes	46
6.2	Transformations linéaires du tore	47
6.3	Fractions continues	49
6.3.1	Préliminaires	49
6.3.2	Ergodicité de la transformation de Gauss	51
6.3.3	Taux de croissance du développement en fractions continues	52
6.3.4	Mélange de la transformation de Gauss	53
6.4	Mélange des systèmes symboliques	54
6.5	Exercices	55
6.6	Indications pour la résolution des exercices	57

7	Stabilité structurelle	62
7.1	Le théorème du point fixe	62
7.2	Automorphismes linéaires hyperboliques de $\mathbb{R}^N$	63
7.3	Le théorème de Grobman-Hartman	66
7.4	Quelques lemmes de relèvement	67
7.5	Stabilité structurelle des automorphismes hyperboliques du tore.	69
7.6	Exercices	70
7.7	Indications pour la résolution des exercices	71
8	Représentations unitaires	72
8.1	Sur $SL_N(\mathbb{R})$	72
8.2	Représentations unitaires	74
8.3	Ergodicité des quotients de $SL_2(\mathbb{R})$ par un réseau	75
8.4	Décroissance des coefficients et mélange	76
8.5	Une construction de réseaux uniformes de $SL_2(\mathbb{R})$	79
8.6	Interprétation géométrique	81
9	Entropie métrique	85
9.1	Information et entropie d'une partition : définitions	85
9.2	Information et entropie d'une partition : propriétés	87
9.3	Entropie d'un système dynamique mesurable : définitions	90
9.4	Entropie d'un système dynamique mesuré : propriétés	91
9.5	L'entropie comme fonction de la mesure	93
9.6	Exercices	94
9.7	Indications pour la résolution des exercices	95
10	Entropie topologique	96
10.1	Recouvrements ouverts	96
10.2	Entropie topologique : définition	97
10.3	Entropie topologique : propriétés	99
10.4	Le principe variationnel : première inégalité	100
10.5	Le principe variationnel : seconde inégalité	101
10.6	Exercices	103
10.7	Indications pour la résolution des exercices	104
11	Sous-décalages de type fini	107
11.1	Systèmes de Markov	107
11.2	Chaînes de Markov	110
11.3	Mesure de Parry	112
11.4	Exercices	115
11.5	Indications pour la résolution des exercices	121
12	Codage	131
12.1	Variétés stables, lemme de pistage et lemme de fermeture	131
12.2	Partition de Markov	133
12.3	Un codage sur le tore $\mathbb{T}^2$	137
12.4	Constructions de partitions de Markov	140
12.5	Entropie des automorphismes hyperboliques du tore	142

12.6 Exercices	144
12.7 Indications pour la résolution des exercices	144
Index	145
Bibliographie	149

2 Exemples fondamentaux

Nous donnons quelques exemples fondamentaux de systèmes dynamiques (voir aussi [KH, CFS, PY]). Ils nous serviront de motivation et d'illustration tout au long de ce cours. D'autres exemples seront introduits et développés dans des chapitres ultérieurs. Certains sont les archétypes de comportement de systèmes dynamiques, dont l'étude générale est renvoyée aux références.

2.1 Systèmes dynamiques topologiques et mesurables

Donnons dans ce paragraphe un petit peu de vocabulaire.

Si $f : X \rightarrow X$ est une application et x un point de X , on appelle *orbite (positive)* de x l'ensemble $\{f^n(x) / n \geq 0\}$. Si f est bijective, l'*orbite* de x est l'ensemble $\{f^n(x) / n \in \mathbb{Z}\}$. On prendra bien garde de ne pas confondre, lorsque X est un groupe multiplicatif, les notations ambiguës $f^n(x) = f \circ \dots \circ f(x)$ et $f(x)^n = f(x) \times \dots \times f(x)$.

Une partie A de X est dite *invariante* par f (ou f -invariante) si $f(A) \subset A$. Dans certains ouvrages, cette expression peut aussi signifier par exemple que $f^{-1}(A) = A$. Le lecteur est invité à bien faire attention au contexte.

Un *système dynamique topologique* (à temps discrets) est un couple (X, f) avec X un espace topologique et $f : X \rightarrow X$ une application continue. Il est dit *inversible* si f est un homéomorphisme. On s'intéresse en particulier au comportement topologique des orbites.

Deux systèmes dynamiques topologiques (X, f) et (Y, g) sont (*topologiquement*) *conjugués* s'il existe un homéomorphisme $\psi : X \rightarrow Y$ tel que

$$\psi \circ f = g \circ \psi .$$

Deux transformations continues (topologiquement) conjuguées ont “même dynamique” (topologique), au sens suivant : un tel homéomorphisme ψ envoie une orbite de f dans X sur une orbite de g dans Y , une orbite périodique de f dans X sur une orbite périodique de même période de g dans Y , une orbite dense de f dans X sur une orbite dense de g dans Y , une orbite récurrente de f dans X sur une orbite récurrente de g dans Y , ... (voir plus loin les définitions d'une orbite périodique ou récurrente). Deux systèmes dynamiques topologiques (X, f) et (Y, g) sont (*topologiquement*) *semi-conjugués* s'il existe une application continue $\psi : X \rightarrow Y$ telle que

$$\psi \circ f = g \circ \psi .$$

(on dit aussi que (Y, g) est un *quotient* de (X, f)).

Soient $(X, \mathcal{A}, \mu)$ et $(Y, \mathcal{B}, \nu)$ deux espaces mesurés. Une application mesurable $f : X \rightarrow Y$ *préserve la mesure* si

$$\forall B \in \mathcal{B}, \quad \mu(f^{-1}(B)) = \nu(B) .$$

Lorsque $(X, \mathcal{A}, \mu) = (Y, \mathcal{B}, \nu)$, on dit aussi que la mesure μ est *invariante* par f ou f -invariante.

Un *système dynamique mesurable* (respectivement *mesuré*, *probabilisé*) est un triplet $(X, \mathcal{B}, f)$ (respectivement un quadruplet $(X, \mathcal{B}, \mu, f)$) avec $(X, \mathcal{B})$ (respectivement $(X, \mathcal{B}, \mu)$) un espace mesurable (respectivement mesuré, de probabilité) et $f : X \rightarrow X$ une application mesurable (respectivement mesurable et préservant la mesure μ). Il est dit *inversible* si f

est bijective d'inverse mesurable (respectivement mesurable et préservant la mesure μ). La *théorie ergodique* est l'étude des systèmes dynamiques mesurables, en s'intéressant surtout au comportement des (ou de presque toutes les) orbites.

Deux systèmes dynamiques mesurables $(X, \mathcal{A}, \mu, f)$ et $(Y, \mathcal{B}, \nu, g)$ sont (*mesurablement*) *conjugués* (on dit aussi *conjugués au sens de la mesure*) s'il existe une partie X' (respectivement Y') dans X (respectivement Y) de mesure totale et f -invariante (respectivement g -invariante), et une bijection $\psi : X' \rightarrow Y'$ mesurable et préservant la mesure, ainsi que son inverse, telle que

$$\forall x \in X', \quad \psi \circ f(x) = g \circ \psi(x) .$$

Deux transformations préservant la mesure, qui sont (mesurablement) conjugues ont "même dynamique" (mesurable), au sens suivant : une telle application ψ envoie presque toute orbite de f dans X sur une orbite de g dans Y , une partie mesurable de X dans laquelle revient presque tout orbite de f sur une partie mesurable de Y dans laquelle revient presque tout orbite de g ,

Les notions que nous allons définir dans ce cours (ergodicité au chapitre 4, mélange au chapitre 6) sont invariantes par conjugaison (i.e. si $(X, \mathcal{A}, \mu, f)$ et $(Y, \mathcal{B}, \nu, g)$ sont deux systèmes dynamiques mesurables conjugus, alors l'un est ergodique (respectivement mélangeant) si et seulement si l'autre l'est).

Donnons un critère pratique pour montrer qu'une application préserve la mesure. Rappelons d'abord le résultat suivant, que nous admettrons (voir par exemple [CohD],[Neu, page 23]).

Soit X un ensemble. Une *algèbre de Boole* dans X est une partie $\mathcal{E}$ de $\mathcal{P}(X)$ contenant $\emptyset$, telle que l'intersection de deux éléments de $\mathcal{E}$, et le complémentaire d'un élément de $\mathcal{E}$, soient des éléments de $\mathcal{E}$.

Théorème 2.1 (Théorème de Carathéodory) *Soient $\mathcal{E}$ une algèbre de Boole de parties d'un ensemble X , $\mathcal{B}$ la σ -algèbre engendrée par $\mathcal{E}$ et $\mu : \mathcal{E} \rightarrow [0, +\infty]$ une application vérifiant*

- i) μ est σ -finie sur $\mathcal{E}$ (i.e. X est une union dénombrable d'éléments de $\mathcal{E}$ de mesure finie pour μ).*
- ii) pour tous les éléments C_1, C_2 de $\mathcal{E}$ tels que $C_1 \cap C_2 = \emptyset$, on a $\mu(C_1 \cup C_2) = \mu(C_1) + \mu(C_2)$,*
- iii) pour toute suite décroissante C_n d'éléments de $\mathcal{E}$ telle que $\mu(C_0) < \infty$ et $\bigcap_{n \in \mathbb{N}} C_n = \emptyset$, on a $\lim_{n \rightarrow \infty} \mu(C_n) = 0$.*

Alors μ se prolonge de manière unique en une mesure σ -finie (encore notée μ) sur $\mathcal{B}$. $\square$

En particulier, si l'application $\mu : \mathcal{E} \rightarrow [0, +\infty]$ vérifie $\mu(X) = 1$, alors son extension μ est une mesure de probabilité.

Corollaire 2.2 *Soit $(X, \mathcal{B})$ un espace mesurable. Alors deux mesures sur $(X, \mathcal{B})$, dont au moins une est σ -finie, et qui coïncident sur une sous-algèbre de Boole de $\mathcal{B}$, sont égales. $\square$*

Rappelons que la *mesure image* $f_*\mu$ d'une mesure μ par une application mesurable $f : (X, \mathcal{A}) \rightarrow (Y, \mathcal{B})$ est définie par $f_*\mu(B) = \mu(f^{-1}(B))$ pour tout B dans $\mathcal{B}$. Remarquons que si $(X, \mathcal{A}, \mu)$ et $(Y, \mathcal{B}, \nu)$ sont deux espaces mesurés, alors dire qu'une application mesurable $f : X \rightarrow Y$ préserve la mesure équivaut à dire que

$$f_*\mu = \nu.$$

Notons que si $f : (X, \mathcal{A}) \rightarrow (Y, \mathcal{B})$ et $g : (Y, \mathcal{B}) \rightarrow (Z, \mathcal{C})$ sont deux applications mesurables, alors

$$(g \circ f)_*\mu = g_*(f_*\mu)$$

pour toute mesure μ sur $(X, \mathcal{A})$.

Corollaire 2.3 *Soient $(X, \mathcal{A}, \mu)$ et $(Y, \mathcal{B}, \nu)$ deux espaces mesurés et $f : X \rightarrow Y$ une application mesurable. Soit $\mathcal{C}$ une semi-algèbre dans $\mathcal{B}$ (i.e. une partie de $\mathcal{B}$, contenant $\emptyset$, telle que l'intersection de deux éléments de $\mathcal{C}$, et le complémentaire d'un élément de $\mathcal{C}$, soient une union finie disjointe d'éléments de $\mathcal{C}$), qui engendre la σ -algèbre $\mathcal{B}$. Si ν est σ -finie sur $\mathcal{C}$ et si f préserve la mesure sur $\mathcal{C}$, i.e. si*

$$\forall B \in \mathcal{C}, \quad f^{-1}(B) \in \mathcal{A} \quad \text{et} \quad \mu(f^{-1}(B)) = \nu(B),$$

alors f préserve la mesure.

Preuve. Soit $\mathcal{E}$ l'ensemble des unions finies disjointes d'éléments de $\mathcal{C}$. Alors $\mathcal{E}$ est une algèbre de Boole, et les mesures $f_*\mu$ et ν coïncident sur $\mathcal{E}$. Par le corollaire précédent, ces mesures sont égales, donc f préserve la mesure. $\square$

2.2 Systèmes de Kronecker

Soit K un groupe topologique compact, par exemple le cercle $\mathbb{S}^1 = \{z \in \mathbb{C} / |z| = 1\}$. Soit k un point de K et $\tau_k : x \mapsto kx$ la translation (à gauche) par k . Soit μ_K la mesure de Haar (à gauche) de K , normalisée par $\mu(K) = 1$.

Rappelons (voir par exemple [Wei, CohD], ainsi que le chapitre 8.1 pour $SL_N\mathbb{R}$) que tout groupe topologique localement compact admet une mesure de Radon positive, invariante par translation à gauche, unique à scalaire multiplicatif près, appelée *mesure de Haar*. Par exemple, si $K = \mathbb{S}^1$, alors μ_K est la mesure, notée $d\theta$, définie par $\int_K f d\mu_K = \int_0^1 f(e^{2i\pi t}) dt$.

Alors τ_k est continue et préserve μ_K par définition. On appelle *système de Kronecker* tout système dynamique topologique (K, τ_k) ou tout système dynamique mesurable (K, μ_K, τ_k) .

2.3 Systèmes symboliques ou de Bernoulli

Soit $\Lambda = \{1, \dots, q\}$ un *alphabet* (i.e. un ensemble) fini que l'on munit de la topologie discrète, par exemple induite par la distance discrète

$$d(a, b) = \begin{cases} 1 & \text{si } a \neq b \\ 0 & \text{si } a = b \end{cases}.$$

On pose $\mathbb{E}$ l'ensemble $\mathbb{N}$ ou $\mathbb{Z}$. Soit $X = \Lambda^{\mathbb{E}}$, muni de la topologie produit. Un élément ω de X est un *mot* i.e. une suite $(\omega_i)_{i \in \mathbb{E}}$, et on appelle ω_i la i -ème lettre. L'espace X est compact, et métrisable par exemple par la distance, encore notée d , suivante :

$$d(\omega, \omega') = \sup_{i \in \mathbb{E}} \frac{1}{2^{|i|}} d(\omega_i, \omega'_i) .$$

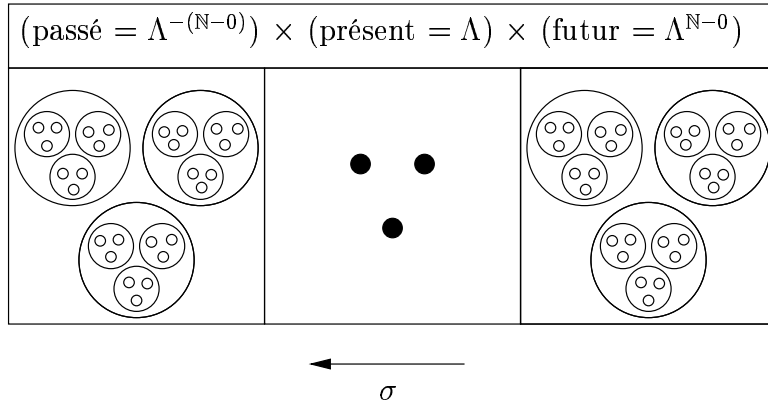
L'espace X est homéomorphe à l'espace de Cantor.

On note $\sigma : X \rightarrow X$, et on appelle *décalage* (à gauche) (et “shift” en anglais) l'application définie par

$$\forall i \in \mathbb{E}, \quad \sigma(\omega)_i = \omega_{i+1} .$$

Elle est 2-Lipschitzienne (donc continue) pour la distance d :

$$\forall \omega, \omega' \in X, \quad d(\sigma(\omega), \sigma(\omega')) \leq 2d(\omega, \omega') .$$



Plus généralement, si $(Y, \mathcal{A}, \nu)$ est un espace de probabilité, (par exemple $Y = \Lambda$, $\mathcal{A} = \mathcal{P}(\Lambda)$ et ν l'équiprobabilité $\nu(\{y\}) = \frac{1}{q}$), notons encore X l'ensemble produit $Y^{\mathbb{E}}$, $\mathcal{B}$ la σ -algèbre produit $\mathcal{A}^{\mathbb{E}}$, et μ la *mesure de probabilité produit* $\nu^{\mathbb{E}}$.

Rappelons (voir par exemple [CohD, Dud]) que $\mathcal{B}$ est la σ -algèbre engendrée par l'algèbre de Boole des unions finies disjointes de cylindres de X . Par définition, un *cylindre* est une partie de X de la forme

$$C_{m, A_0, A_1, \dots, A_n} = \{(\omega_i)_i \in X / \forall i = 0, \dots, n, \omega_{m+i} \in A_i\},$$

avec m dans $\mathbb{E}$, n dans $\mathbb{N}$ et A_i dans $\mathcal{A}$. Notons que la préimage par le décalage σ d'un cylindre est un cylindre :

$$\sigma^{-1}(C_{m, A_0, \dots, A_n}) = C_{m+1, A_0, \dots, A_n} .$$

Lorsque Y est un alphabet fini, nous appellerons plus précisément cylindre tout tel ensemble avec les A_i des singletons (voir paragraphe 11.2), ce qui suffit pour définir la mesure produit.

Pour définir la mesure produit, on utilise alors le théorème de Carathéodory rappelé au paragraphe 2.1. On définit une application μ sur l'algèbre de Boole des unions finies disjointes de cylindres, en posant

$$\mu(C_{m, A_0, \dots, A_n}) = \prod_{i=0}^n \nu(A_i) .$$

Cette application vérifie les conditions du théorème de Carathéodory 2.1. Lorsque $Y = \Lambda$ est un alphabet fini, la condition iii) du théorème de Carathéodory est automatiquement satisfaite, car les cylindres $C_{m, A_0, \dots, A_n}$ sont des compacts de X , donc la condition $\bigcap_{n \in \mathbb{N}} C_n = \emptyset$ assure qu'il existe n_0 tel que, pour tout $n \geq n_0$, on a $C_n = \emptyset$. D'où une mesure μ sur X .

Proposition 2.4 *La mesure μ sur X est invariante par le décalage σ .*

Preuve. Les mesures μ et $\sigma_*\mu$ coïncident sur les cylindres, donc sur les unions finies disjointes de cylindres. Le résultat découle alors de l'unicité dans le théorème de Carathéodory (voir paragraphe 2.1). $\square$

On appelle *système de Bernoulli* (ou *système symbolique*) tout système dynamique topologique (X, σ) , ainsi que tout système dynamique mesurable $(X, \mathcal{B}, \mu, \sigma)$, construit à partir d'un alphabet fini Λ .

2.4 Mesures de Liouville et systèmes hamiltoniens

Soit $F : \Omega \rightarrow \mathbb{R}^n$ un champ de vecteurs de classe C^∞ sur un ouvert Ω de $\mathbb{R}^n$. On considère l'équation différentielle ordinaire

$$\frac{dx}{dt} = F(x)$$

sur Ω . On note $f^t(x)$ la valeur en l'instant t de l'unique solution valant x à l'instant $t = 0$. On suppose pour simplifier que $f^t(x)$ est défini pour tout t dans $\mathbb{R}$ (sinon, ce qui suit reste vrai localement). Alors $(f^t)_{t \in \mathbb{R}}$ est un groupe à un paramètre de classe C^∞ de difféomorphismes de Ω (i.e. $(t, x) \mapsto f^t(x)$ est de classe C^∞ , f^t est un difféomorphisme de Ω pour tout t , $f^0 = id$ et $f^{t+s} = f^t \circ f^s$ pour tous s, t).

Le résultat suivant donne un critère d'existence d'une mesure f^t -invariante absolument continue par rapport à la mesure de Lebesgue sur Ω . Une telle mesure est appelée une *mesure de Liouville*.

Proposition 2.5 (Théorème de Liouville) *Soit λ la mesure de Lebesgue sur Ω , et $\rho : \Omega \rightarrow [0, +\infty[$ une application de classe C^∞ . Le flot (f^t) préserve la mesure $\rho d\lambda$ si et seulement si (en notant F_k la k -ème coordonnée de F)*

$$\sum_{k=1}^n \frac{\partial}{\partial x^k} (\rho F_k) = 0 .$$

Preuve. On note $\mathcal{C}_c^\infty(\Omega, \mathbb{R})$ l'ensemble des applications de classe C^∞ à support compact sur Ω . Par densité de $\mathcal{C}_c^\infty(\Omega, \mathbb{R})$ dans $L^1(\Omega, \lambda)$, il suffit de montrer que, pour tout t dans $\mathbb{R}$ et tout φ dans $\mathcal{C}_c^\infty(\Omega, \mathbb{R})$, on a

$$\int_{\Omega} \varphi \rho d\lambda = \int_{\Omega} \varphi \circ f^t \rho d\lambda .$$

Comme $f^0 = id$, il suffit de montrer que, pour tout φ dans $\mathcal{C}_c^\infty(\Omega, \mathbb{R})$, on a

$$\frac{d}{dt} \left(\int_{\Omega} \varphi \circ f^t \rho d\lambda \right) = 0 .$$

Comme f^t est un groupe à un paramètre de difféomorphismes, il suffit de montrer que, pour tout φ dans $\mathcal{C}_c^\infty(\Omega, \mathbb{R})$, on a

$$\int_{\Omega} \frac{d}{dt}(\varphi \circ f^t) \big|_{t=0} \rho \, d\lambda = 0 .$$

Or, en notant I l'intégrale ci-dessus,

$$I = \int_{\Omega} d\varphi(x) \cdot F(x) \rho(x) \, d\lambda(x) = \int_{\Omega} \sum_{k=1}^n \rho F_k \frac{\partial \varphi}{\partial x^k} \, d\lambda = \int_{\Omega} \left(\sum_{k=1}^n \frac{\partial}{\partial x^k} (\rho F_k) \right) \varphi \, d\lambda .$$

Le résultat en découle. $\square$

Corollaire 2.6 *Le flot local d'un champ de vecteurs sur un ouvert de $\mathbb{R}^n$ préserve le volume si et seulement si sa divergence est nulle.* $\square$

En mécanique classique, de nombreux systèmes sont régis par un *système (d'équations) hamiltonien*. Soit Ω un ouvert de $\mathbb{R}^{2n}$, de coordonnées $(q_1, \dots, q_n, p_1, \dots, p_n)$, et $H : \Omega \rightarrow \mathbb{R}$ une application de classe C^∞ , appelé un *hamiltonien*. Par exemple, dans le cas de l'attraction universelle, H est l'énergie totale, somme des énergies potentielle et cinétique, $q = (q_1, \dots, q_n)$ la position et $p = (p_1, \dots, p_n)$ la quantité de mouvement. Le système d'équations d'hamiltonien H est alors

$$\begin{cases} \frac{dq_i}{dt} = \frac{\partial H}{\partial p_i} \\ \frac{dp_i}{dt} = -\frac{\partial H}{\partial q_i} \end{cases} .$$

Le flot local d'un système hamiltonien est de divergence nulle, car

$$\frac{\partial}{\partial q_i} \left(\frac{\partial H}{\partial p_i} \right) + \frac{\partial}{\partial p_i} \left(-\frac{\partial H}{\partial q_i} \right) = 0 .$$

Donc un flot local d'un système hamiltonien préserve la mesure de Lebesgue sur $\mathbb{R}^{2n}$.

L'hamiltonien est une *intégrale première* pour son flot local, i.e. les surfaces de niveau Σ_c définies par $H = c$, pour c une constante, sont invariantes par f^t . Lorsque Σ_c est une sous-variété compacte de Ω , on montre que le flot local sur Σ_c est défini pour tout temps, et qu'il existe une mesure de probabilité sur Σ_c , qui est invariante par le flot hamiltonien (voir un cours de géométrie différentielle, par exemple [Spi]; la mesure est la mesure riemannienne pour la métrique riemannienne induite sur Σ_c par la métrique euclidienne de $\mathbb{R}^n$, ou, autrement dit, si $d\lambda$ est la forme volume euclidienne sur $\mathbb{R}^n$, et X le champ de vecteurs unitaire (pour la norme euclidienne) orthogonal à Σ_c , alors la forme volume de Σ_c est $i_X(d\lambda)$; l'invariance de cette mesure sur Σ_c vient de l'invariance de la mesure de Lebesgue par le flot hamiltonien).

Un gaz idéal contenu dans une boîte parallélépipédique V est un exemple de tel système hamiltonien. On suppose que ce gaz suit les lois de la mécanique classique, deux molécules n'étant soumises qu'à une interaction répulsive $U(r)$ fonction de leur distance mutuelle r . On suppose que les chocs sur les parois de la boîte sont élastiques. Si N est le nombre

de molécules, l'espace des phases est une partie de $\mathbb{R}^{6N}$ de coordonnées (p, q) avec p dans $(\mathbb{R}^3)^N$ la quantité de mouvement et q dans $(\mathbb{R}^3)^N$ la position des N particules. Ce gaz est donc un système hamiltonien pour l'hamiltonien $H = \sum_{i=1}^N \frac{p_i^2}{2m_i} + \sum_{i \neq j} U(\|q_i - q_j\|)$. Si H_0 est l'énergie initiale du système, l'évolution du gaz se fait sur l'hypersurface $H = H_0$, compacte car V l'est, et préserve donc une mesure de probabilité.

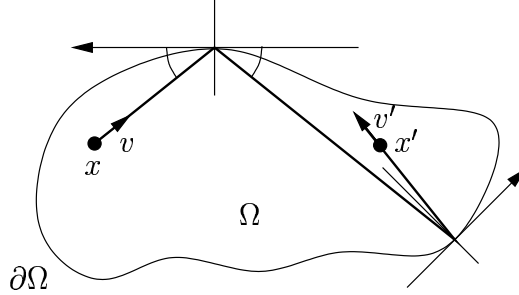
C'est une des motivations de la théorie ergodique d'arriver à comprendre le comportement statistique des orbites d'un système dynamique de la mécanique classique, précisément en fonction du fait qu'il laisse invariant une mesure.

2.5 Billards

Soit Ω un ouvert borné de $\mathbb{R}^2$, à bord $\partial\Omega$ de classe C^∞ par morceaux, orienté par la normale extérieure. On fixe un paramétrage (préservant l'orientation) de $\partial\Omega$ proportionnellement à la longueur d'arc, de sorte que la longueur totale de $\partial\Omega$ soit 2π . On pose $\overline{\Omega} = \Omega \cup \partial\Omega$.

Soit X l'ensemble des couples (x, v) avec x dans $\overline{\Omega}$ et v un vecteur unitaire basé en x dans $\mathbb{R}^2$, de sorte que si x est dans $\partial\Omega$, alors v pointe vers l'intérieur. On munit X (contenu dans $\overline{\Omega} \times \mathbb{S}^1$) de la mesure induite par la mesure produit $d\lambda d\theta$, avec $d\lambda$ la mesure de Lebesgue sur $\mathbb{R}^2$ et $d\theta$ la mesure de Haar sur $\mathbb{S}^1$.

Pour tout (x, v) dans X , on définit la *trajectoire de billard* issue de (x, v) comme la courbe affine par morceaux partant de x en suivant la droite dirigée par v , puis en se réfléchissant avec chocs élastiques sur le bord de Ω .



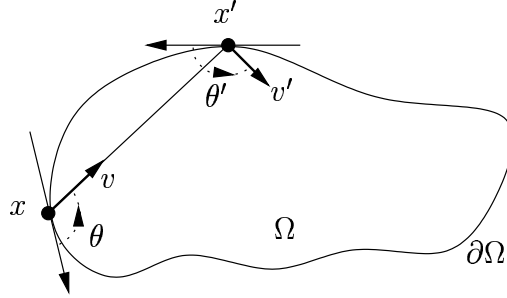
Soit $(f^t)_{t \in \mathbb{R}}$ le groupe à un paramètre d'applications mesurables de X dans X , définies (sauf sur l'ensemble de mesure nulle des points dont la trajectoire passe par un point anguleux de $\partial\Omega$) par $(x, v) \mapsto f^t(x, v) = (x', v')$ avec x' le point de la trajectoire de (x, v) à distance (algébrique) t de x le long de cette trajectoire, et v' le vecteur tangent à la trajectoire (orientée) en x' (et si x' est un point de $\partial\Omega$, alors v' est le vecteur suivant immédiatement la réflexion). On appelle $(f^t)_{t \in \mathbb{R}}$ le *flot du billard* sur Ω .

Proposition 2.7 *Le flot du billard préserve la mesure $d\lambda d\theta$.*

Preuve. Soit t dans $\mathbb{R}$ (que l'on peut supposer positif, le raisonnement étant analogue sinon) et (x_0, v_0) un point de X . Quitte à enlever un ensemble de mesure nulle, on suppose que $f^t(x_0, v_0)$ est défini et n'appartient pas à $\partial\Omega \times \mathbb{S}^1$. Si la trajectoire de (x_0, v_0) entre les instants 0 et t ne rencontre pas $\partial\Omega$, alors au voisinage de (x_0, v_0) , l'application f^t est une translation $(x, v) \mapsto (x + tv, v)$, qui préserve bien la mesure $d\lambda d\theta$.

Si la trajectoire de (x_0, v_0) entre les instants 0 et t rencontre $\partial\Omega$ une et une seule fois, alors au voisinage de (x_0, v_0) , alors le point $f^t(x, v)$ s'obtient en composant la translation $(x, v) \mapsto (x + tv, v)$ et la symétrie par rapport à la droite orthogonale à la droite $\Delta_{x,v}$ tangente à $\partial\Omega$ au premier point $y_{x,v}$ d'intersection en partant de x de la droite $x + \mathbb{R}v$. Donc f_t est de classe C^1 au voisinage de (x_0, v_0) , et pour montrer qu'elle préserve la mesure $d\lambda d\theta$, il suffit de montrer que son jacobien est 1. En raisonnant au premier ordre, nous pouvons supposer que la droite $\Delta_{x,v}$ est constante, et, après changement de coordonnées, horizontale. Alors la symétrie ci-dessus transforme $x = (x_1, x_2)$ en $x = (x_1, -x_2)$ et l'angle θ de v par rapport à l'horizontale en $\pi - \theta$. Donc son jacobien est 1, ce qui montre le résultat. $\square$

Soit Σ le sous-espace (mesurable) de X formé des points (x, v) avec $x \in \partial\Omega$. Notons $T : \Sigma \rightarrow \Sigma$ l'application de premier retour du flot du billard f^t sur Σ , qui est appelée *la transformation du billard*. Rappelons que $T(x, v) = f^t(x, v) = (x', v')$ si t est la distance entre x et le premier point x' d'intersection avec $\partial\Omega$ avec la trajectoire du billard issue de (x, v) (voir figure ci-dessous).



Si Ω est convexe, on identifie $(x, v) \in \Sigma$ avec $(s, \theta) \in \mathbb{S}^1 \times]0, \pi[$, avec s le paramétrage de x sur $\partial\Omega$, et θ l'angle de v en x par rapport à la tangente (orientée) en x à $\partial\Omega$. On note alors μ la mesure suivante sur Σ :

$$d\mu = \sin \theta \, ds \, d\theta .$$

Notons que l'application T est définie presque partout pour cette mesure (en tout point (x, v) tel que x et x' sont des points réguliers de $\partial\Omega$).

Proposition 2.8 *Si Ω est convexe, la transformation du billard T préserve la mesure μ .*

Preuve. Aux points où elle est définie, l'application T est différentiable. Calculons son jacobien. Au second ordre près, nous pouvons supposer que le bord $\partial\Omega$ est affine par morceaux. Notons $T(s, \theta) = (s', \theta')$. On a $\frac{\partial \theta'}{\partial s} = 0$, il suffit donc de calculer $\frac{\partial s'}{\partial s}$ et $\frac{\partial \theta'}{\partial \theta}$.

Pour calculer $\frac{\partial s'}{\partial s}$, on travaille à θ constant (voir figure ci-dessous). Rappelons que dans un triangle euclidien de côtés de longueur a, b, c et d'angles opposés α, β, γ , on a

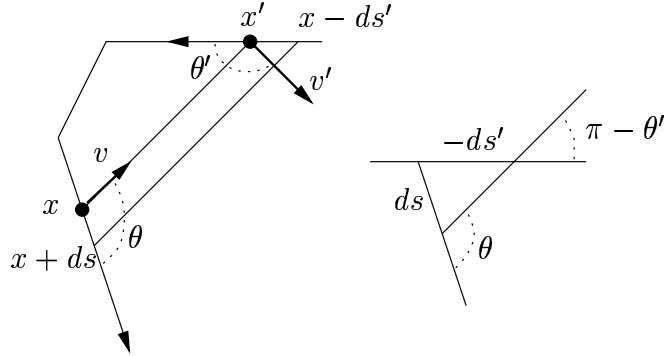
$$\frac{a}{\sin \alpha} = \frac{b}{\sin \beta} = \frac{c}{\sin \gamma} .$$

En tenant compte des orientations, on a

$$\frac{ds}{\sin(\pi - \theta')} = \frac{-ds'}{\sin(\pi - \theta)},$$

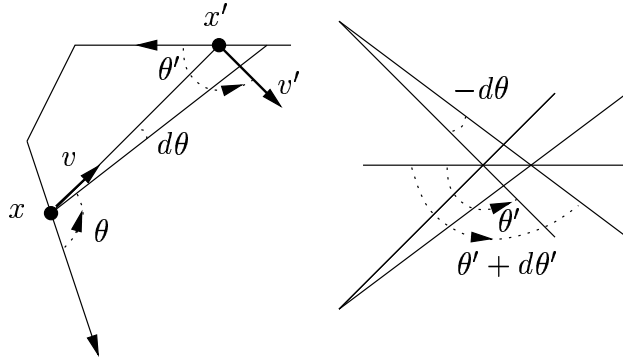
donc

$$\frac{\partial s'}{\partial s} = -\frac{\sin \theta}{\sin \theta'}.$$



Pour calculer $\frac{\partial \theta'}{\partial \theta}$, on travaille à s constant (voir figure ci-dessous). Une symétrie préservant les angles et renversant l'orientation, on a $d\theta' = -d\theta$, et donc

$$\frac{\partial \theta'}{\partial \theta} = -1.$$



Par le théorème de changement de variable, le résultat en découle. $\square$

2.6 Exercices

Exercice E.2.1 Soit $f : [0, 1] \rightarrow [0, 1]$ l'application donnée par $f(x) = 2\sqrt{x(1-x)}$. Montrer que f préserve la mesure de probabilité $\mu = \frac{dx}{2\sqrt{1-x}}$.

Exercice E.2.2 Soient α dans $\mathbb{R}$, $a_1, \dots, a_n$ dans $\mathbb{C}$ avec $|a_i| < 1$, et $f : \mathbb{C} \rightarrow \mathbb{C}$ l'application définie par

$$f(z) = e^{2i\pi\alpha} z \frac{z - a_1}{1 - \overline{a_1}z} \dots \frac{z - a_n}{1 - \overline{a_n}z}.$$

1. Montrer que si $|z| = 1$, alors $|f(z)| = 1$.
2. Est-ce que l'application $z \mapsto \frac{z - a_1}{1 - \overline{a_1}z}$ préserve la mesure de Haar sur $\mathbb{S}^1$?

3. Montrer que $f : \mathbb{S}^1 \rightarrow \mathbb{S}^1$ préserve la mesure de Haar.

Exercice E.2.3 Soit $X = [0, 1[$. On rappelle que tout élément x de X admet un unique développement diadique $x = \sum_{n=1}^{\infty} \frac{x_n}{2^n}$ où x_n vaut 0 ou 1 et n'est pas constant égal à 1 à partir d'un certain rang. Soit $f : X \rightarrow X$ l'application définie par $f(x) = y$ où $y_n = x_{n+2}$ si n est impair, $y_2 = x_1$, et $y_n = x_{n-2}$ si n est pair au moins égal à 4.

1. Montrer que f préserve la mesure de Lebesgue.
2. Montrer qu'il existe un point x dans X dont l'orbite est dense dans X .
3. Montrer que f est conjuguée à un décalage de Bernoulli.

Exercice E.2.4 Soit $X =]0, +\infty[$ et $f : X \rightarrow X$ l'application définie par $f(x) = 2x$.

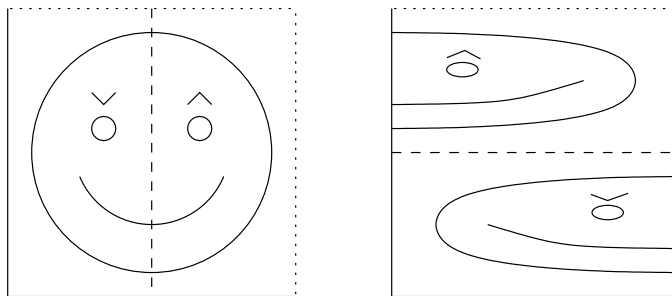
1. Montrer qu'il n'existe pas de mesure borélienne finie invariante par f .
2. Construire une mesure borélienne μ invariante par f , sans atome, i.e.

$$\forall x \in X, \quad \mu(\{x\}) = 0$$

et σ -finie (i.e. telle que X est union dénombrable de parties mesurables de mesure finie.) Existe-t-il une telle mesure μ qui soit équivalente à la mesure de Lebesgue λ (i.e. telle que pour tout borélien B , on a $\mu(B) = 0$ si et seulement si $\lambda(B) = 0$) ?

Exercice E.2.5 Soit $Y = \{0, 1\}$ et ν l'équiprobabilité sur Y .

1. Montrer que le décalage de Bernoulli sur $(Y^{\mathbb{N}}, \nu^{\mathbb{N}})$ est conjugué (au sens de la mesure) à l'application de *doublement de l'angle* de $\mathbb{R}/\mathbb{Z}$ dans $\mathbb{R}/\mathbb{Z}$ définie par $x \mapsto 2x$.
2. Montrer que le décalage de Bernoulli sur $(Y^{\mathbb{Z}}, \nu^{\mathbb{Z}})$ est conjugué (au sens de la mesure) à l'application du carré $[0, 1[\times [0, 1[$ dans lui-même définie par le dessin ci-dessous.



Exercice E.2.6 Soit λ la mesure de Haar sur $\mathbb{S}^1 = \mathbb{R}/\mathbb{Z}$ (normalisée pour être de probabilité). Si A, B sont des parties mesurables de $\mathbb{S}^1$, on note $\varphi_{A,B}$ l'application de $\mathbb{S}^1$ dans $\mathbb{R}$ définie par

$$t \mapsto \lambda(A \cap (B + t)) .$$

1. Montrer que $\varphi_{A,B}$ est continue.
2. Montrer que $\int_{\mathbb{S}^1} \varphi_{A,B} d\lambda = \lambda(A) \times \lambda(B)$.

3. En déduire que si α est irrationnel, alors la rotation R_α d'angle α sur le cercle est ergodique (voir la définition au paragraphe 4.1).
4. Montrer que si α est irrationnel, alors les rotations R_α et $R_{2\alpha}$ ne sont pas conjuguées au sens de la mesure.
5. Montrer que les deux rotations $R_{\frac{1}{5}}$ et $R_{\frac{2}{5}}$ sont conjuguées (au sens de la mesure) mais ne sont pas topologiquement conjuguées.
6. Donner une condition nécessaire et suffisante pour que deux rotations du cercle soient conjuguées au sens de la mesure.

2.7 Indications pour la résolution des exercices

Exercice E.2.2 (1) Calcul.

(2) Oui si $a_1 = 0$, non sinon. Si $a_1 = \rho e^{i\theta}$ avec $0 < \rho < 1$, alors cette application g admet $e^{i\theta}$ pour point fixe, qui est répulsif car $g'(e^{i\theta}) = \frac{1+\rho}{1-\rho} > 1$. Il existe donc un voisinage ouvert U de $e^{i\theta}$ dans $\mathbb{S}^1$ tel que $\overline{g^{-1}(U)} \subset U$. Donc g ne préserve pas la mesure de Haar sur le cercle.

(3) Il s'agit de montrer que, pour toute fonction continue $\varphi : \mathbb{S}^1 \rightarrow \mathbb{S}^1$, on a

$$\int_{\mathbb{S}^1} \varphi \, d\theta = \int_{\mathbb{S}^1} \varphi \circ f \, d\theta .$$

Si $\overline{\varphi}$ est l'extension harmonique de φ à l'intérieur du disque unité, alors, comme f envoie le disque unité dans lui même (par connexité) et est holomorphe, l'application $\varphi \circ f$ (qui est harmonique, par la caractérisation des fonctions harmoniques comme les parties réelles des fonctions holomorphes) est (par unicité) le prolongement harmonique de $\varphi \circ f$. Donc

$$\int_{\mathbb{S}^1} \varphi \, d\theta = \overline{\varphi}(0) = \overline{\varphi} \circ f(0) = \int_{\mathbb{S}^1} \varphi \circ f \, d\theta .$$

Exercice E.2.3 (3) Considérer $\Psi : X \rightarrow \{0, 1\}^{\mathbb{Z}}$ définie par

$$\Psi \left(\sum_{n=1}^{\infty} \frac{x_n}{2^n} \right) = (\dots, x_6, x_4, x_2, x_1, x_3, x_5, x_7, \dots)$$

(avec x_1 en position 0), et montrer que Ψ envoie la mesure de Lebesgue sur la mesure produit $\nu^{\mathbb{Z}}$ avec ν l'équiprobabilité sur $\{0, 1\}$.

(2) découle alors de (3), en utilisant la preuve de la proposition 4.2 et l'ergodicité du décalage de Bernoulli (voir l'exercice E.4.3 et le paragraphe 6.4).

(1) découle de (3).

Exercice E.2.5 La transformation du dessin est définie par $(x, y) \mapsto (x', y')$ avec

$$x' = \begin{cases} 2x & \text{si } x \in [0, \frac{1}{2}[\\ 2(x - \frac{1}{2}) & \text{si } x \in [\frac{1}{2}, 1[\end{cases} \quad \text{et} \quad y' = \begin{cases} \frac{1}{2}y & \text{si } x \in [0, \frac{1}{2}[\\ \frac{1}{2}(y + 1) & \text{si } x \in [\frac{1}{2}, 1[\end{cases}$$

(1) On pourra considérer l'application $\psi : \{0, 1\}^{\mathbb{N}} \rightarrow \mathbb{R}/\mathbb{Z}$ définie par

$$\psi((x_n)_{n \in \mathbb{N}}) = \sum_{n=0}^{\infty} \frac{x_n}{2^{n+1}} \pmod{1}.$$

(2) On pourra considérer l'application $\psi : \{0, 1\}^{\mathbb{Z}} \rightarrow [0, 1[\times [0, 1[$ définie par

$$\psi((x_n)_{n \in \mathbb{Z}}) = \left(\sum_{n=0}^{+\infty} \frac{x_n}{2^{n+1}}, \sum_{n=-1}^{-\infty} \frac{x_n}{2^n} \right).$$

Exercice E.2.6 (1) Utiliser le fait que pour tout borélien A , et tout $\epsilon > 0$, il existe deux unions finies d'intervalles U, V telles que $U \subset A \subset V$ et $\lambda(U - V) \leq \epsilon$.

(2) En notant $\mathbb{I}_E$ la *fonction caractéristique* d'une partie E , et comme $\mathbb{I}_{E \cap E'} = \mathbb{I}_E \mathbb{I}_{E'}$, on a

$$\begin{aligned} \int_{\mathbb{S}^1} \varphi_{A,B} d\lambda &= \int_{\mathbb{S}^1} \int_{\mathbb{S}^1} \mathbb{I}_{A \cap (B+t)}(x) dx dt = \int_{\mathbb{S}^1} \mathbb{I}_A(x) \left(\int_{\mathbb{S}^1} \mathbb{I}_{B+t}(x) dt \right) dx = \\ &= \int_{\mathbb{S}^1} \mathbb{I}_A(x) \left(\int_{\mathbb{S}^1} \mathbb{I}_{-B+x}(t) dt \right) dx = \int_{\mathbb{S}^1} \mathbb{I}_A(x) \left(\int_{\mathbb{S}^1} \mathbb{I}_{-B}(t) dt \right) dx = \lambda(A) \times \lambda(B). \end{aligned}$$

(3) Si A est un borélien tel que $R_\alpha^{-1}(A) = A$, alors $\varphi_{A,A}$ vaut $\lambda(A \cap A) = \lambda(A)$ sur tous les points de $\mathbb{R}/\mathbb{Z}$ images de $\mathbb{Z} + \alpha\mathbb{Z}$. Si α est irrationnel, ce sous-groupe de $\mathbb{R}$ est dense, et par continuité, l'application $\varphi_{A,A}$ est constante de valeur $\lambda(A)$. Par (2), on a donc $\lambda(A)$ vaut 0 ou 1.

(4) Soit $A = [0, \frac{1}{4}[$ et n_i une suite d'entiers tels que $n_i \alpha$ converge vers $\frac{1}{2}$. Alors pour i assez grand, $R_\alpha^{n_i}(A)$ et A sont disjoints, donc $\lambda(R_\alpha^{n_i}(A) \cap A) = 0$. Si R_α et $R_{2\alpha}$ étaient mesurablement conjugués, il existerait un borélien A' de même mesure que A tel que

$$\lambda(R_\alpha^{n_i}(A) \cap A) = \lambda(R_{2\alpha}^{n_i}(A') \cap A').$$

Or ce dernier terme converge vers $\lambda(A')$ par continuité de $\varphi_{A',A'}$, ce qui contredit le fait que $\lambda(A') > 0$.

(5) Pour montrer que $R_{\frac{1}{5}}$ et $R_{\frac{2}{5}}$ sont conjuguées au sens de la mesure, il suffit de découper le cercle en cinq intervalles de même longueur.

Un homéomorphisme du cercle préserve l'ordre cyclique, et l'ordre cyclique sur une orbite de $R_{\frac{1}{5}}$ et une orbite de $R_{\frac{2}{5}}$ ne coïncident pas.

(6) Utiliser le développement en série de Fourier.

3 Récurrence

La récurrence est une propriété de l'orbite d'un point, qui consiste à revenir une infinité de fois vers son point de départ, un peu comme une comète qui revient assez régulièrement près de la Terre.

Nous montrons dans ce premier chapitre plusieurs théorèmes d'existence de points récurrents (voir aussi [KH, Part 1, 3.3 et 4.1][PY, Chap. 1]). Comme application (voir aussi [Fur][PY, Chap. 2]), nous montrerons que pour toute partition finie de $\mathbb{N}$, il existe une partie qui contient des suites arithmétiques de longueur aussi grande que l'on veut. Voir [Fur][PY, Chap. 16] pour une généralisation de ce résultat, le théorème de Szemerédi.

3.1 Récurrence et minimalité

Soit f une application continue d'un espace topologique X dans lui-même. Un point x de X est dit *récurrent* (pour f) s'il existe une suite strictement croissante d'entiers n_k telle que

$$\lim_{k \rightarrow +\infty} f^{n_k}(x) = x.$$

Un point x est dit *périodique* pour f s'il existe $n \geq 1$ tel que $f^n(x) = x$.

Théorème 3.1 (Théorème de récurrence de Birkhoff) *Soit X un espace métrisable compact. Toute application continue $f : X \rightarrow X$ admet un point récurrent.*

Remarques. (1) L'hypothèse X métrisable n'est pas indispensable. Elle est vérifiée dans tous les exemples qui nous intéresseront.

(2) Ce résultat est faux si X n'est pas compact. Par exemple, prendre $X = \mathbb{R}$ et $f(x) = x + 1$.

(3) Il peut n'y avoir qu'un seul point récurrent. Par exemple, prendre $X = \mathbb{R} \cup \{\infty\} \simeq \mathbb{S}^1$ et $f(x) = x + 1$, $f(\infty) = \infty$.

(4) Il peut n'y avoir aucun point périodique. Par exemple, prendre une rotation irrationnelle du cercle, et même $X = \mathbb{S}^1 = \{z \in \mathbb{C} / |z| = 1\}$ et $f(x) = \lambda x$ pour λ dans $\mathbb{S}^1$ non racine de l'unité.

(5) Nous donnons une première preuve de ce résultat, qui repose sur l'axiome du choix. Nous verrons dans le chapitre suivant une preuve qui ne l'utilise pas.

Une partie fermée Y de X est dite *minimale* (pour f) si elle est non vide, *invariante* par f (i.e. $f(Y) \subset Y$) et si Y ne contient pas de fermé non vide invariant par f autre que Y .

Par exemple, un point fixe ou une orbite périodique est minimale. Dans l'exemple d'une rotation irrationnelle du cercle, le cercle est minimal.

Remarque. Comme X est métrisable, une partie fermée non vide Y de X est minimale si et seulement si, pour tout y dans Y , l'orbite positive $\{f^n(y) / n \geq 0\}$ est dense dans Y . En particulier, tout point d'une partie minimale est récurrent.

Proposition 3.2 *Soit X un espace métrisable compact non vide et $f : X \rightarrow X$ une application continue. Alors X contient un fermé minimal.*

Preuve. Soit $\mathcal{F}$ l'ensemble des fermés non vides de X invariants par f , partiellement ordonné par l'inclusion. On veut montrer que $\mathcal{F}$ contient un élément minimal. Ceci résulte du théorème de Zorn [Kri], car toute partie $\mathcal{F}'$ de $\mathcal{F}$ totalement ordonnée admet un minorant : l'intersection des éléments de $\mathcal{F}'$, qui est non vide par compacité. $\square$

Preuve du théorème de récurrence de Birkhoff: Il découle de la remarque et de la proposition précédente. $\square$

Exemple 1 : Soit (K, τ_k) un système de Kronecker (voir 2.2).

Lemme 3.3 *Tout point x de K est récurrent pour τ_k .*

Preuve. Soit x_0 un point de K récurrent pour τ_k . Soient $k' = x_0^{-1}x$ et $\rho_{k'} : y \mapsto yk'$, de sorte que $\rho_{k'}(x_0) = x$. Comme $\rho_{k'}$ et τ_k commutent et $\rho_{k'}$ est continue, le point x est aussi récurrent pour τ_k . $\square$

Exemple 2 : Soit $(\Lambda^{\mathbb{N}}, \sigma)$ ou $(\Lambda^{\mathbb{Z}}, \sigma)$ un système de Bernoulli (voir 2.3) sur un alphabet fini Λ .

Les points récurrents sont les mots qui chaque fois qu'ils contiennent un mot fini, contiennent dans la suite une deuxième copie de celui-ci.

3.2 Récurrence multiple

Voici une généralisation du théorème de récurrence de Birkhoff.

Théorème 3.4 (Furstenberg-Weiss) *Soit X un espace métrisable compact et $f : X \rightarrow X$ une application continue. Alors, pour tout ℓ dans $\mathbb{N} - \{0\}$, il existe un point x de X et une suite strictement croissante d'entiers n_k tels que*

$$\forall i \in \{1, \dots, \ell\}, \quad \lim_{k \rightarrow \infty} f^{i n_k}(x) = x.$$

Autrement dit, il existe un point x récurrent pour $f, f^2, \dots, f^\ell$, avec les mêmes temps de retour. Nous utiliserons le lemme indépendant suivant.

Lemme 3.5 *Soit Y un espace métrisable complet et φ une fonction semi-continue supérieurement (i.e. pour tout y_0 dans Y , on a $\lim_{y \rightarrow y_0} \varphi(y) \leq \varphi(y_0)$, voir par exemple [Bre, page 8]). Alors l'ensemble des points de continuité de φ est dense dans Y .*

Preuve. Quitte à remplacer φ par e^φ , on peut supposer que φ est minorée. Notons, pour $\epsilon > 0$,

$$Y_\epsilon = \{y \in Y \mid \exists (y_k) \in Y^{\mathbb{N}}, \lim_{k \rightarrow \infty} y_k = y \text{ et } \lim_{k \rightarrow \infty} \varphi(y_k) \leq \varphi(y) - \epsilon\}.$$

Comme φ est semicontinue supérieurement, l'ensemble Y_ϵ est fermé et l'ensemble des points de continuité de φ est l'intersection des ouverts ${}^c Y_\epsilon$. Par le théorème de Baire, il suffit de voir que chacun de ces ouverts est dense, c'est-à-dire que chacun des fermés Y_ϵ est d'intérieur vide.

Si, par l'absurde, l'intérieur de Y_ϵ contenait un point y_0 , il contiendrait aussi un point y_1 tel que $\varphi(y_1) \leq \varphi(y_0) - \frac{\epsilon}{2}$. On construirait ainsi une suite de points y_k dans l'intérieur de Y_ϵ telle que $\varphi(y_k) \leq \varphi(y_{k-1}) - \frac{\epsilon}{2}$. Ceci contredirait le fait que φ est minorée. $\square$

Preuve du théorème de Furstenberg-Weiss. On peut bien sûr supposer que X est minimal. On procède par récurrence sur ℓ . Pour $\ell = 1$, c'est le théorème de récurrence de Birkhoff 3.1. On suppose le résultat vrai pour $\ell - 1$. On note x un point de X et $n_k \rightarrow \infty$ une suite telle que

$$\forall i \in \{1, \dots, \ell - 1\}, \quad \lim_{k \rightarrow \infty} f^{i n_k}(x) = x .$$

On introduit l'espace produit $Y = X^\ell$, la diagonale $\Delta = \{(x, \dots, x) / x \in X\}$ de Y et les deux transformations continues g et h de Y données par

$$g(x_1, \dots, x_\ell) = (f(x_1), f^2(x_2), \dots, f^\ell(x_\ell))$$

$$h(x_1, \dots, x_\ell) = (f(x_1), f(x_2), \dots, f(x_\ell)) .$$

On veut trouver un point récurrent de g qui soit sur Δ . Pour cela, on utilisera les trois propriétés suivantes de g et h .

$$\text{Étape 1 : } \left\{ \begin{array}{l} \text{(i) } g \text{ et } h \text{ commutent (i.e. } g \circ h = h \circ g). \\ \text{(ii) } h \text{ préserve } \Delta \text{ et } \Delta \text{ est minimal pour } h. \\ \text{(iii) Il existe } y_0 \text{ dans } \Delta \text{ et des points } z_{0,k} \text{ dans } \Delta \text{ tels que} \\ \lim_{k \rightarrow \infty} g^{n_k}(z_{0,k}) = y_0 . \end{array} \right.$$

Preuve. Les points (i) et (ii) sont clairs. Pour vérifier le point (iii), on prend $y_0 = (x, \dots, x)$ et $z_{0,k} = (x_k, \dots, x_k)$ où $x_k \in X$ vérifie $f^{n_k}(x_k) = x$. C'est possible, car par minimalité, f est surjective. On a alors

$$g^{n_k}(z_{0,k}) = (x, f^{n_k}(x), \dots, f^{(\ell-1)n_k}(x)) \xrightarrow{k \rightarrow \infty} (x, x, \dots, x) = y_0 . \quad \square$$

Étape 2 : $\forall y \in \Delta, \forall \epsilon > 0, \exists z \in \Delta, \exists n \geq 1, \quad d(g^n z, y) < \epsilon$.

Preuve. Par minimalité de Δ pour h , il existe $m \geq 0$ tel que $d(h^m y_0, y) < \epsilon$. On prend alors $z = h^m(z_{0,k})$ pour k suffisamment grand, car

$$g^{n_k} z = h^m g^{n_k}(z_{0,k}) \xrightarrow{k \rightarrow \infty} h^m y_0 . \quad \square$$

Étape 3 : $\forall \epsilon > 0, \exists y \in \Delta, \exists n \geq 1, \quad d(g^n y, y) < \epsilon$.

Preuve. Soit $\epsilon_0 = \frac{\epsilon}{2}$ et y_0 un point de Δ . L'étape 2 permet de trouver un point y_1 dans Δ et $n_1 \geq 1$ tel que $d(g^{n_1} y_1, y_0) < \epsilon_0$. La continuité de g^{n_1} permet de trouver $\epsilon_1 < \epsilon_0$ tel que

$$\forall y \in \Delta, \quad d(y, y_1) < \epsilon_1 \Rightarrow d(g^{n_1} y, y_0) < \epsilon_0 .$$

On recommence avec ϵ_1 et y_1 . On choisit ainsi, de proche en proche, pour tout $k \geq 1$ des entiers $n_k \geq 1$, des points y_k dans Δ et des réels $\epsilon_k < \epsilon_{k-1}$ tels que

$$\forall y \in \Delta, \quad d(y, y_k) < \epsilon_k \Rightarrow d(g^{n_k} y, y_{k-1}) < \epsilon_{k-1} .$$

On a alors

$$\forall i < j, \quad d(g^{n_j + n_{j-1} + \dots + n_{i+1}} y_j, y_i) < \epsilon_i < \frac{\epsilon}{2} .$$

Par compacité de Δ , on peut trouver $i < j$ tel que $d(y_i, y_j) < \frac{\epsilon}{2}$. On prend alors $y = y_j$ et $n = n_j + n_{j-1} + \dots + n_{i+1}$. On a $d(g^n y, y) < \epsilon$. $\square$

Pour conclure, introduisons la fonction $\varphi : \Delta \rightarrow [0, +\infty[$ définie par

$$\varphi(y) = \inf_{n \geq 1} d(y, g^n y) .$$

Cette fonction φ est semi-continue supérieurement. D'après le lemme 3.5 ci-dessus, il existe un point de continuité a de φ . Il nous suffit de montrer que $\varphi(a) = 0$.

Supposons, par l'absurde, que $\varphi(a) > 0$. Il existe alors un réel $\delta > 0$ et un ouvert non vide V de Δ tel que, pour tout y dans V , on a $\varphi(y) > \delta$. On a l'égalité

$$\bigcup_{m \geq 1} h^{-m}(V) = \Delta$$

car le complémentaire de cette union est un fermé h -invariant, donc vide par minimalité de Δ . Par compacité de Δ , on peut trouver une partie finie F de $\mathbb{N} - \{0\}$ telle que on a

$$\bigcup_{m \in F} h^{-m}(V) = \Delta .$$

Par uniforme continuité des h^m sur Δ , il existe $\epsilon > 0$ tel que

$$\forall y', y'' \in \Delta, \quad d(y', y'') < \epsilon \Rightarrow \forall m \in F, \quad d(h^m y', h^m y'') < \delta .$$

On considère alors le point y de Δ et l'entier $n \geq 1$ donnés par l'étape 3. On a donc $d(g^n y, y) < \epsilon$. Ce point y est dans un des ouverts $h^{-m}(V)$ pour un m dans F . On a donc $d(g^n h^m y, h^m y) < \delta$. En particulier, le point $h^m y$ de V vérifie $\varphi(h^m y) < \delta$. Contradiction. $\square$

Voici une application du théorème de Furstenberg-Weiss.

Théorème 3.6 (van der Waerden) *Soit $\mathbb{N} = B_1 \sqcup \dots \sqcup B_q$ une partition finie de $\mathbb{N}$, alors il existe une partie B_j qui contient une suite arithmétique $\{a, a+n, a+2n, \dots, a+\ell n\}$ avec $n \geq 1$, de longueur ℓ aussi grande que l'on veut.*

Preuve. Il suffit de voir que, pour ℓ fixé, une des parties B_j contient une suite arithmétique de longueur ℓ . La partition définit un mot ω du système de Bernoulli $\{1, \dots, q\}^{\mathbb{N}}$ par

$$\forall i \in \mathbb{N}, \quad \omega_i = j \Leftrightarrow i \in B_j .$$

L'espace $\{1, \dots, q\}^{\mathbb{N}}$ est muni de la distance d définie dans la section 2.3. Soit $X = \{\sigma^n \omega, n \geq 0\}$ l'adhérence de l'orbite de ω par le décalage σ . D'après le théorème de Furstenberg-Weiss, il existe un point ω' de X et un entier $n \geq 1$ tel que $d(\omega', \sigma^j \omega') < 1$ pour tout $j = 1, \dots, \ell$, c'est-à-dire, tel que $\omega'_0 = \omega'_n = \omega'_{2n} = \dots = \omega'_{\ell n}$. Comme ω' est dans l'adhérence de l'orbite de ω , il existe un entier $a \geq 0$ tel que $d(\omega', \sigma^a \omega) < 2^{-\ell n}$, c'est-à-dire, tel que $\omega_{a+i} = \omega'_i$ pour tout $i = 1, \dots, \ell n$. En particulier, on a $\omega_a = \omega_{a+n} = \dots = \omega_{a+\ell n}$. Autrement dit, en notant j cet entier, la suite arithmétique $\{a, a+n, a+2n, \dots, a+\ell n\}$ est dans B_j . $\square$

3.3 Le théorème de récurrence de Poincaré

En présence d'une mesure invariante, nous nous intéressons ci-dessous à la propriété de récurrence presque sûre des orbites.

Théorème 3.7 (Théorème de récurrence de Poincaré) *Soit $(X, \mathcal{B}, \mu)$ un espace de probabilité, $f : X \rightarrow X$ une application mesurable préservant μ , et A un élément de $\mathcal{B}$. Alors, pour μ -presque tout point x de A , l'orbite $(f^n(x))_{n \in \mathbb{N}}$ repasse une infinité de fois dans A .*

Preuve. Sinon, il existe $n_0 \geq 1$ tel que l'ensemble

$$B = \{x \in A \mid \forall n \geq n_0, f^n(x) \notin A\}$$

vérifie $\mu(B) > 0$. Quitte à remplacer f par f^{n_0} , on peut supposer $n_0 = 1$. Les ensembles $(f^{-n}(B))_{n \in \mathbb{N}}$ ne sont pas disjoints, car $\sum_{n \in \mathbb{N}} \mu(f^{-n}(B)) = \infty$. Donc il existe $m < n$ tel que $f^{-n}(B) \cap f^{-m}(B) \neq \emptyset$. D'où on en déduit que $f^{n-m}(B) \cap B \neq \emptyset$. Contradiction. $\square$

Autre preuve : Notons $A_n = \bigcup_{k \geq n} f^{-k}(A)$ l'ensemble des points de X qui passent dans A lorsque l'on itère f au moins n fois. Notons $A_\infty = \bigcap_{n \in \mathbb{N}} \bigcup_{k \geq n} f^{-k}(A)$ l'ensemble des points qui passent une infinité de fois dans A . Alors $\mu(A_{n+1}) = \mu(f^{-1}(A_n)) = \mu(A_n)$, et $A \subset A_0$. Puisque la suite (A_n) est décroissante, on a $\mu(A_n \cap A) = \mu(A_0 \cap A) = \mu(A)$. Donc $\mu(A_\infty \cap A) = \mu(A)$ par les propriétés des mesures finies. Ceci démontre le résultat. $\square$

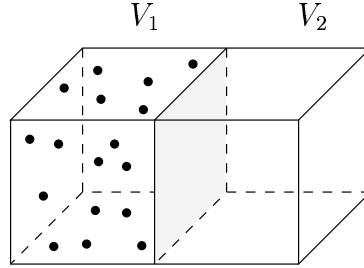
Corollaire 3.8 *Soit X un espace métrisable séparable, μ une mesure de probabilité borélienne sur X et $f : X \rightarrow X$ une application continue préservant μ . Alors μ -presque tout point de X est récurrent pour f .*

Preuve. Soit $\mathcal{C}$ une base dénombrable d'ouverts dans X (par exemple l'ensemble des boules centrées en un point d'une partie dénombrable dense de X , et de rayon rationnel strictement positif). Pour tout élément B de $\mathcal{C}$, le théorème de récurrence de Poincaré fournit un ensemble μ -négligeable N_B tel que tout point de $B - N_B$ repasse une infinité de fois dans B . La réunion $N = \bigcup_{B \in \mathcal{C}} N_B$ est encore μ -négligeable. Or, par construction, tout point du complémentaire de N est récurrent. $\square$

Exemple 1 : Considérons un flot $(f^t)_{t \in \mathbb{R}}$ préservant un ouvert Ω de $\mathbb{R}^n$, et défini par une équation différentielle $\frac{dx}{dt} = F(x)$ sur Ω . Si la mesure de Lebesgue de Ω est finie, et si F est à divergence nulle, alors la trajectoire de presque tout point de Ω revient arbitrairement près de x (i.e. $\liminf_{t \rightarrow +\infty} \|f^t(x) - x\| = 0$).

Exemple 2 (Paradoxe de Zermelo) : Faisons l'expérience de mécanique statistique suivante. On considère un gaz idéal contenu dans une boîte parallélépipédique V subdivisée en deux parties cubiques V_1, V_2 de longueur de côté 10 cm. On suppose qu'à l'instant $t = 0$, toutes les molécules sont dans V_1 (par exemple parce que l'on a fait le vide dans V_2), et on

enlève la cloison entre V_1 et V_2 .



La théorie prédit que pour presque toutes ces positions initiales d'énergie donnée, le gaz doit revenir dans la partie V_1 . Cela semble paradoxal, car ce n'est jamais observé. En fait le premier temps de retour moyen est fini, mais très très très grand, de l'ordre de $2^{10^{23}}$ (ce que l'on suppose être plus long que la durée de vie de l'univers!).

3.4 Exercices

Exercice E.3.1 Soit X un espace métrisable compact, $f : X \rightarrow X$ une application continue et Y un fermé minimal de X . Montrer que $f(Y) = Y$. A-t-on $f^{-1}(Y) = Y$?

Exercice E.3.2 (Birkhoff) Soit X un espace métrisable compact et $f : X \rightarrow X$ une application continue. Un point x de X est dit *quasipériodique* si, pour tout voisinage V de x , il existe une suite d'entiers n_k qui tend vers l'infini, telle que $f^{n_k}(x) \in V$, et telle que la suite $n_{k+1} - n_k$ des "temps de retour dans V " est bornée.

1. Montrer que tout point d'un fermé minimal Y de X est quasipériodique.
2. En déduire que X contient des points quasipériodiques.
3. Montrer que réciproquement, si x est quasipériodique, alors le fermé $\overline{\{f^n(x), n \in \mathbb{N}\}}$ est minimal.

Exercice E.3.3 On considère le système symbolique $\Lambda^{\mathbb{Z}}$ ou $\Lambda^{\mathbb{N}}$ avec $\Lambda = \{0, 1\}$.

1. Construire un point ω de $\Lambda^{\mathbb{Z}}$ qui soit récurrent pour σ mais pas pour σ^{-1} .
2. Construire un point ω de $\Lambda^{\mathbb{N}}$ qui soit quasipériodique mais pas périodique.
3. Construire un point ω de $\Lambda^{\mathbb{N}}$ qui soit récurrent mais pas quasipériodique.

Exercice E.3.4 Construire une partie A de $\mathbb{N}$ telle que ni A ni son complémentaire ne contienne de suite arithmétique infinie.

Exercice E.3.5 (Théorème de Furstenberg-Weiss II) Soit X un espace métrisable compact et $f : X \rightarrow X$ une application continue. On veut montrer que, pour tout point x de X , il existe un point quasipériodique y de X et une suite $n_k \rightarrow \infty$ telle que

$$\lim_{k \rightarrow \infty} d(f^{n_k}(x), f^{n_k}(y)) = 0.$$

Pour cela, on introduit l'espace topologique produit X^X muni de la topologie produit. C'est un espace compact par le théorème de Tychonoff. On note E l'adhérence de l'ensemble $\{f, f^2, f^3, \dots\}$ dans X^X . Par définition, on a

$$E = \{u \in X^X / \forall x_1, \dots, x_p \in X, \exists (n_k) \in \mathbb{N}^{\mathbb{N}}, \forall i = 1, \dots, p, \lim_{k \rightarrow \infty} f^{n_k}(x_i) = u(x_i)\}.$$

1. Montrer que pour tout v dans X^X , l'application $u \mapsto u \circ v$ de X^X dans lui-même est continue. En déduire de $E^2 \subset E$ (i.e. $\forall u, v \in E, u \circ v \in E$).
2. Montrer qu'il existe un fermé minimal Y dans X et un élément u de E tel que $u(x) \in Y$.
3. Soit $\mathcal{F}$ l'ensemble des parties fermées F de E non vides, telles que $F^2 \subset F$ et telles que, pour tout u dans F , on a $u(x) \in Y$. Montrer que $\mathcal{F}$ est non vide. Montrer, à l'aide du lemme de Zorn, que $\mathcal{F}$ contient un élément minimal F_0 pour l'inclusion.
4. Montrer que pour tout u dans F_0 , la partie $F_0 \circ u$ est encore dans $\mathcal{F}$. En déduire que $F_0 = \{u\}$.
5. En déduire qu'il existe $u \in E$ tel que $u(x) \in Y$ et $u^2 = u$.
6. Montrer que le point $y = u(x)$ convient.

Exercice E.3.6 (Théorème de Hindman) Soit $\mathbb{N} = B_1 \sqcup \dots \sqcup B_q$ une partition finie de $\mathbb{N}$. Montrer qu'il existe une partie B_j qui contient un parallélogramme infini, i.e. une partie P de $\mathbb{N}$ de la forme

$$P = \{p_{i_1} + \dots + p_{i_\ell} / i_1 < \dots < i_\ell\}$$

où $p_{i_1} \leq \dots \leq p_{i_\ell} \leq \dots$ est une suite infinie d'entiers.

Exercice E.3.7 (Théorème de Furstenberg) On considère le cercle $\mathbb{S}^1 = \mathbb{R}/\mathbb{Z}$, muni des structures usuelles de groupe abélien et d'espace métrique, et, pour tout n dans $\mathbb{N}$, l'application "de multiplication par n de l'angle", définie par $x \mapsto nx$.

Le but de ce problème est de démontrer un théorème de Furstenberg, disant que toute partie du cercle, fermée, non vide, invariante par doublement et triplement de l'angle, et minimale pour ces propriétés, est de cardinal fini.

On note $\Sigma = \{2^r 3^s / r, s \in \mathbb{N}\}$.

- (1) Construire une partie fermée infinie et propre du cercle, invariante par triplement de l'angle.
- (2) Soit (X, d) un espace métrique et $\epsilon > 0$, une partie Y de X est dite ϵ -dense si

$$\forall x \in X, \exists y \in Y, d(x, y) \leq \epsilon.$$

a) Si $S = \{m \log 2 + p \log 3 / m, p \in \mathbb{N}\}$ et $S_n = \{x - n \log 6 / x \in S\}$, montrer que, pour tout $\epsilon > 0$, pour tout n suffisamment grand, la partie $S_n \cap [0, +\infty[$ est ϵ -dense dans $[0, +\infty[$.

b) En déduire qu'il existe une suite strictement croissante $(s_n)_{n \in \mathbb{N}}$ dans Σ , telle que $\frac{s_{n+1}}{s_n}$ converge vers 1.

(3) Si A, B sont des parties de $\mathbb{S}^1$, et n un élément de $\mathbb{N}$, on note

$$nA = \{na \mid a \in A\} \quad \text{et} \quad A - B = \{a - b \mid a \in A, b \in B\}.$$

Si Ω est une partie de $\mathbb{N}$, une partie A de $\mathbb{S}^1$ est dite Ω -invariante si $nA \subset A$ pour tout n dans Ω . Elle est dite Ω -minimale si elle est fermée, non vide, Ω -invariante et ne contient pas de partie fermée, non vide, Ω -invariante, propre.

a) Montrer qu'avec les notations de la question (2), pour tout $\epsilon > 0$, si x est un point du cercle $\mathbb{S}^1$ différent de 0, suffisamment proche de 0 et à droite de 0, alors la partie $\{s_k x \mid k \in \mathbb{N}\}$ est ϵ -dense dans $\mathbb{S}^1$.

b) Montrer que toute partie Σ -invariante du cercle, contenant le point 0 comme point non isolé, est dense dans le cercle.

c) Montrer que si M est une partie fermée Σ -invariante du cercle, alors ou bien M est finie (et formée de points rationnels), ou bien $M - M = \mathbb{S}^1$.

(4) Le but de cette dernière question est de montrer que si M est une partie Σ -minimale du cercle telle que $M - M = \mathbb{S}^1$, alors $M = \mathbb{S}^1$.

On pose $\Sigma_n = \{x \in \Sigma \mid x \equiv 1 \pmod{5^n}\}$. On fixe M_n une partie Σ_n -minimale du cercle, contenue dans M , avec $M_{n+1} \subset M_n$.

a) Montrer que pour tout n dans $\mathbb{N}$ et x dans Σ , il existe ℓ dans $\mathbb{N} - \{0\}$ tel que x^ℓ appartienne à Σ_n .

b) On note $x_1, \dots, x_{k_n}$ des représentants de Σ modulo 5^n . Montrer que $\bigcup_{i=1}^{k_n} x_i M_n = M$.

c) Montrer que $M_n - M = \mathbb{S}^1$.

d) Montrer que, pour m, n dans $\mathbb{N}$ et ξ dans M_n , si $\theta_{m,n} = m/5^n \pmod{1}$, alors $\theta_{m,n} \in \{\xi\} - M$.

e) En déduire le résultat.

(5) Conclure.

3.5 Indications pour la résolution des exercices

Exercice E.3.2 1) On montrera que, pour tout voisinage ouvert V de x , on a $Y \subset \bigcup_{n \in \mathbb{N}} f^{-n}(V)$ et on extraira un recouvrement fini.

3) On montrera que, pour tout y dans Y et tout voisinage V de x , il existe $n \geq 0$ tel que $f^n(y) \in V$, et donc que $x \in \{f^n(x), n \in \mathbb{N}\}$.

Exercice E.3.6 Comme pour le théorème de van der Waerden, on introduira le mot ω du système symbolique $\{1, \dots, q\}^{\mathbb{N}}$ donné par $\omega_i = j$ si et seulement si $i \in B_j$. On appliquera l'exercice E.3.5 avec $X = \{\sigma^n \omega \mid n \in \mathbb{N}\}$ et $x = \omega$.

Exercice E.3.7 (1) L'image dans le cercle $\mathbb{R}/\mathbb{Z}$ du Cantor triadique $\{\sum_{i=1}^{+\infty} \frac{a_i}{3^i} \mid a_i = 0, 2\}$ dans $\mathbb{R}$ est invariant par triplement de l'angle.

(2) a) Il est bien connu que le sous-groupe $G = \mathbb{Z} \log 2 + \mathbb{Z} \log 3$ de $\mathbb{R}$ est dense, car $\log 2$ et $\log 3$ sont rationnellement indépendants. Les S_n forment une suite croissante de parties de G , de réunion égale à G , donc dense. Supposons par l'absurde qu'il existe une sous-suite S_{n_k} qui ne rencontre pas un intervalle I contenu dans $[0, +\infty[$, de longueur $\epsilon > 0$. Soit q_k dans $\mathbb{N}$ tel que $I - q_k \log 2 \subset [-\log 2, 0]$. Comme $S_{n_k} \subset S_{n_k} - q_k \log 2$, pour tout k dans $\mathbb{N}$, il existe un intervalle de longueur ϵ contenu dans l'intervalle $[-\log 2, 0]$ que la partie S_{n_k} ne rencontre pas. Ceci contredit le fait que $\bigcup_{k \in \mathbb{N}} S_{n_k}$ est dense dans $\mathbb{R}$.

b) Donc S est ϵ -dense dans $[n \log 6, +\infty[$ pour n assez grand. Si $(u_n)_{n \in \mathbb{N}}$ est l'énumération des éléments de la partie discrete S de $\mathbb{R}$, de sorte que $u_n < u_{n+1}$, alors u_n tend vers $+\infty$ et $u_{n+1} - u_n$ tend vers 0. En posant $s_n = e^{u_n}$, le résultat en découle.

(3) On identifie $\mathbb{S}^1 = \mathbb{R}/\mathbb{Z}$ et $[0, 1[$.

a) Soit $n_1 \in \mathbb{N}$ tel que $s_{k+1}/s_k \leq 1 + \epsilon$ pour $k \geq n_1$. Soit x dans $\mathbb{R}$ tel que $s_{n_1}x < \epsilon$. Alors pour tout $k \geq n_1$, on a

$$(s_{k+1} - s_k)x = (s_{k+1}/s_k - 1)s_kx \leq \epsilon s_kx.$$

On considère la suite croissante $(s_kx)_{k \geq n_1}$, qui sur l'axe réel part à gauche de ϵ et augmente au plus de ϵ à chaque pas tant qu'elle reste à gauche de 1, car elle vérifie $s_{k+1}x - s_kx \leq \epsilon$ si $s_kx \leq 1$. Donc $\{s_kx / k \in \mathbb{N}\}$ est ϵ -dense dans $[0, 1[$.

b) Soit Y une telle partie. Pour tout $\epsilon > 0$, soit y dans Y proche, à droite, et différent de 0, tel que $\{s_ky / k \in \mathbb{N}\}$ soit ϵ -dense. Comme Y est Σ -invariante, on en déduit que Y est ϵ -dense dans $\mathbb{S}^1$, et ceci pour tout ϵ , ce qui montre le résultat.

c) Soit M une partie fermée Σ -invariante infinie. Alors par compacité du cercle, $M - M$ est une partie Σ -invariante, fermée, et contenant 0 comme point non isolé à droite. Par b), elle est égale au cercle.

4) a) Comme tout élément x de Σ est premier avec 5, il est inversible dans l'anneau $\mathbb{Z}/5^n\mathbb{Z}$, qui est fini. Comme il existe des entiers $p > q$ tels que $x^p = x^q \pmod{5^n}$, on a donc $x^{p-q} = 1 \pmod{5^n}$.

b) Soit $F = \bigcup_{i=1}^n M_n$. Alors F est un fermé non vide, contenu dans M et Σ_n -invariant. Montrons qu'il est Σ -invariant, ce qui par minimalité montrera que $F = M$. Soit x dans Σ . Pour tout i , soit j tel que $xx_i \equiv x_j \pmod{5^n}$. Par 4a), soit ℓ dans $\mathbb{N} - \{0\}$ tel que $x_j^\ell \in \Sigma_n$, et en particulier, $x_j^\ell M_n \subset M_n$.

Alors $xx_i x_j^{\ell-1} \equiv x_j^\ell \equiv 1 \pmod{5^n}$, donc $xx_i x_j^{\ell-1}$ appartient à Σ_n . D'où $xx_i x_j^{\ell-1} M_n \subset M_n$, ce qui entraîne que $xx_i x_j^\ell M_n \subset x_j M_n$. En particulier $xx_i M_n \cap x_j M_n$ est non vide, donc les deux ensembles Σ_n -minimaux $xx_i M_n$ et $x_j M_n$ coïncident. Par conséquent $x F \subset F$, ce qui montre le résultat.

c) Soit $F = M_n - M$. D'après 3c) et 4b), on a $\bigcup_{i=1}^n x_i F = \mathbb{S}^1$. Donc le fermé F est d'intérieur non vide, et Σ_n -invariant. Soit x dans $\Sigma_n - \{1\}$. Par ergodicité de l'application de multiplication par x sur le cercle (voir chapitre suivant, mais la minimalité, et même l'existence d'une orbite dense suffit), on en déduit que $F = \mathbb{S}^1$.

d) Écrivons $\theta_{m,n} = \xi_0 - m$ avec ξ_0 dans M_n et m dans M . Par minimalité, il existe une suite x_i dans Σ_n telle que $x_i \xi_0$ converge vers ξ . Quitte à extraire, on peut supposer que $x_i m$ converge vers un point m' de M . Comme $x_i \equiv 1 \pmod{5^n}$, on a $x_i \frac{m}{5^n} \equiv \frac{m}{5^n} \pmod{1}$. Donc $\theta = \xi - m'$, ce qui montre le résultat.

e) Soit ξ un point de $\bigcap_{n \in \mathbb{N}} M_n$ (qui est non vide par le théorème des compacts emboîtés). Alors pour tous les entiers m, n , le point $\xi - \theta_{m,n}$ appartient à M . Par densité des $\theta_{m,n}$ dans $\mathbb{S}^1$, et comme M est fermé, on en déduit que $M = \mathbb{S}^1$.

(5) Comme $\{0\}$ est une partie fermée Σ -invariante non vide de $\mathbb{S}^1$, on obtient une contradiction à partir de la question (4), et donc seule la première possibilité dans l'alternative de la question (3) c) est possible. Ceci démontre le théorème de Furstenberg.

4 Ergodicité

Historiquement, le mot ergodique ($\epsilon\rho\gamma\omicron\nu$ = énergie, $\omicron\delta\omicron\varsigma$ = chemin) reflète l'idée, due à Boltzmann, que, le plus souvent, un système physique décrit au cours du temps un chemin sur l'hypersurface à énergie constante, chemin qui est dense sur cette hypersurface. De nos jours, le concept mathématique d'ergodicité ne fait plus référence à la physique ou à la topologie. Voir le paragraphe 2.4 pour une explication de l'évolution des idées.

4.1 Transformations ergodiques

Soit $(X, \mathcal{B}, \mu)$ un espace de probabilité, et $f : X \rightarrow X$ une application mesurable qui préserve μ .

On dit que f est *ergodique* si

$$\forall A \in \mathcal{B}, \quad f^{-1}(A) = A \Rightarrow \mu(A) = 0 \text{ ou } 1.$$

Lorsque la transformation f est sous-entendue c'est la mesure μ que l'on qualifie d'*ergodique*. Voir [Wal, KH] pour l'extension au cas où la mesure μ n'est pas une probabilité, et au cas où f ne préserve pas μ , mais préserve les ensembles de mesure nulle.

La proposition suivante donne une interprétation $\mathbb{L}^2$ de l'ergodicité d'une transformation f . Cette interprétation relie les propriétés dynamiques de f aux propriétés spectrales de l'opérateur unitaire de $\mathbb{L}^2(X, \mu)$

$$U_f : \varphi \mapsto \varphi \circ f$$

(qui est bien unitaire car f préserve μ .)

Nous dirons qu'une application $\varphi : X \rightarrow \mathbb{C}$ est *f-invariante* si $\varphi \circ f = \varphi$ et *presque partout f-invariante* si $\varphi \circ f$ coïncide presque partout avec φ .

Proposition 4.1 *Avec ces notations, les assertions suivantes sont équivalentes :*

1. *f est ergodique,*
2. *toute application $\varphi : X \rightarrow \mathbb{C}$ mesurable, telle que $\varphi \circ f = \varphi$ presque partout, est presque partout constante.*
3. *toute application φ dans $\mathbb{L}^1(X, \mu)$, telle que $\varphi \circ f = \varphi$ presque partout, est presque partout constante.*
4. *toute application φ dans $\mathbb{L}^2(X, \mu)$, telle que $\varphi \circ f = \varphi$ presque partout, est presque partout constante.*

Preuve. Il est immédiat, car $\mathbb{L}^2(X, \mu) \subset \mathbb{L}^1(X, \mu)$ et en prenant pour φ la *fonction caractéristique* $\mathbb{I}_A$ de A , que (2) implique (3) implique (4) implique (1). (Rappelons que $\mathbb{I}_A(x)$ vaut 1 si x appartient à A et 0 sinon.)

Pour montrer que (1) implique (2), on peut supposer φ à valeurs réelles. Si φ n'est pas constante presque partout, il existe un réel x tel que, en notant $A' = \varphi^{-1}([x, +\infty[)$, on a $0 < \mu(A') < 1$. L'égalité $\varphi \circ f = \varphi$ presque partout assure que les ensembles A' et $f^{-1}(A')$ coïncident en dehors d'une partie négligeable. L'ensemble $A = \bigcap_{p \in \mathbb{N}} \bigcup_{n \geq p} f^{-n}(A')$

coïncide alors avec A' en dehors d'une partie négligeable. En particulier, on a $0 < \mu(A) < 1$. Comme $f^{-1}(A) = A$, ceci contredit l'ergodicité de μ . $\square$

L'exercice E.4.2 décrit une propriété de densité, au sens de la mesure, des orbites. La proposition suivante renforce encore l'intuition "je passe partout" sous-jacente au mot ergodique.

Proposition 4.2 *Soit X un espace métrisable séparable, μ une mesure de probabilité borélienne sur X de support X , et $f : X \rightarrow X$ une application continue qui préserve μ . Si f est ergodique, alors μ -presque toute orbite est dense dans X (i.e. pour μ -presque tout x dans X , on a $\overline{\{f^n(x), n \in \mathbb{N}\}} = X$).*

Preuve. Soit $\mathcal{C}$ une base dénombrable d'ouverts non vides dans X . Pour tout élément B de $\mathcal{C}$, la partie $A_B = \bigcap_{p \in \mathbb{N}} \bigcup_{n \geq p} f^{-n}(B)$ vérifie $f^{-1}(A_B) = A_B$. Cette partie est une intersection décroissante de parties de mesure supérieure ou égale à $\mu(B)$. Donc $\mu(A_B) \geq \mu(B) > 0$. Par ergodicité de f , on a donc $\mu(A_B) = 1$. L'intersection $A = \bigcap_{B \in \mathcal{C}} A_B$ est aussi de mesure totale, et les orbites des points de A sont denses. $\square$

Exemples : Soit $N \geq 1$ un entier. Considérons le tore $\mathbb{T}^N = (\mathbb{S}_1)^N = (\mathbb{R}/\mathbb{Z})^N \simeq \{\theta = (\theta_1, \dots, \theta_N) \in [0, 1]^N\}$, muni de la mesure de Lebesgue $d\theta = d\theta_1 \dots d\theta_N$.

1. Soit $\alpha = (\alpha_1, \dots, \alpha_N)$ un élément de $\mathbb{R}^N$. La translation du tore $f : \theta \mapsto \theta + \alpha \bmod \mathbb{Z}^N$ est ergodique si et seulement si $1, \alpha_1, \dots, \alpha_N$ sont rationnellement indépendants.
2. Pour tout M dans $\mathbb{N} - \{0, 1\}$, l'application $g : \theta \mapsto M\theta \bmod \mathbb{Z}^N$ est ergodique.

En particulier, une *rotation* $f : \theta \mapsto \theta + \alpha \bmod 1$ sur le cercle, d'angle α irrationnel, ainsi que l'application de *doublement de l'angle* $f : \theta \mapsto 2\theta \bmod 1$ sur le cercle, sont ergodiques (pour la mesure de Lebesgue sur le cercle).

Preuve. Il est immédiat que ces transformations préservent la mesure (voir aussi le paragraphe 2.2 pour le premier cas).

Pour φ dans $\mathbb{L}^2(\mathbb{T}^N)$ et k dans $\mathbb{Z}^N$, en notant $\langle \cdot | \cdot \rangle$ le produit scalaire usuel sur $\mathbb{R}^N$, on pose

$$c_k(\varphi) = \int_{[0,1]^N} e^{-2i\pi \langle k | \theta \rangle} \varphi(\theta) d\theta.$$

La théorie de Fourier assure que la transformée de Fourier $\varphi \mapsto (c_k(\varphi))_{k \in \mathbb{Z}^N}$ est une isométrie de $\mathbb{L}^2(\mathbb{T}^N)$ sur $\ell^2(\mathbb{Z}^N)$.

Soit $\alpha = (\alpha_1, \dots, \alpha_N)$ un élément de $\mathbb{R}^N$. Si $1, \alpha_1, \dots, \alpha_N$ ne sont pas rationnellement indépendants, alors soit k dans $\mathbb{Z}^N - \{0\}$ tel que $\langle k | \alpha \rangle \in \mathbb{Z}$. L'application $\varphi : \theta \mapsto e^{-2i\pi \langle k | \theta \rangle} \varphi(\theta)$ est $\mathbb{L}^2$, invariante par f , non constante presque partout. Donc f n'est pas ergodique.

Supposons maintenant $1, \alpha_1, \dots, \alpha_N$ rationnellement indépendants. Soit φ dans $\mathbb{L}^2(\mathbb{T}^N)$ telle que $\varphi \circ f = \varphi$ ou $\varphi \circ g = \varphi$ presque partout. On veut montrer que φ est constante presque partout. On calcule aisément, pour tout k dans $\mathbb{Z}^N$,

1. $c_k(\varphi) = c_k(\varphi \circ f) = e^{2i\pi \langle k | \alpha \rangle} c_k(\varphi)$. Donc pour tout k dans $\mathbb{Z}^N - \{0\}$, on a $c_k(\varphi) = 0$.
2. $c_{Mk}(\varphi) = c_{Mk}(\varphi \circ g) = c_k(\varphi)$. Donc $c_k(\varphi) = c_{M^m k}(\varphi)$ pour tout entier $m \geq 0$. Comme $\lim_{\|k\| \rightarrow \infty} c_k(\varphi) = 0$, pour tout k dans $\mathbb{Z}^N - \{0\}$, on a $c_k(\varphi) = 0$.

Dans les deux cas, l'injectivité de la transformation de Fourier montre que $\varphi = c_0(\varphi)$ presque partout. $\square$

Nous verrons plus tard que les décalages de Bernoulli sont ergodiques, car mélangeants (voir l'exercice E.4.3 et le paragraphe 6.4).

4.2 Le théorème ergodique

Le théorème suivant est une version “quantitative” de la “densité” des orbites. Pour une transformation ergodique, il exprime que, pour toute partie mesurable A , la proportion de temps passé dans A par presque toutes les orbites est égale à $\mu(A)$. De façon plus précise :

Théorème 4.3 (Théorème ergodique de Birkhoff) *Soit $(X, \mathcal{B}, \mu)$ un espace mesuré, et $f : X \rightarrow X$ une application mesurable qui préserve μ . Pour tout φ dans $\mathbb{L}^1(X, \mu)$, on note*

$$S_n \varphi(x) = \frac{1}{n} \sum_{k=0}^{n-1} \varphi(f^k(x)) .$$

(i) *La limite $\tilde{\varphi}(x) = \lim_{n \rightarrow \infty} S_n \varphi(x)$ existe pour μ -presque tout x .*

(ii) *$\tilde{\varphi} \circ f = \tilde{\varphi}$ presque partout.*

(iii) *$\|\tilde{\varphi}\|_{\mathbb{L}^1} \leq \|\varphi\|_{\mathbb{L}^1}$.*

(iv) *Si la mesure μ est finie, alors la convergence a lieu dans $\mathbb{L}^1$, i.e.*

$$\lim_{n \rightarrow \infty} \|S_n \varphi - \tilde{\varphi}\|_{\mathbb{L}^1} = 0 .$$

(v) *Pour toute partie A dans $\mathcal{B}$, de mesure finie, telle que $f^{-1}(A) = A$, on a*

$$\int_A \varphi d\mu = \int_A \tilde{\varphi} d\mu .$$

(vi) *En particulier, si μ est une mesure de probabilité ergodique, alors*

$$\tilde{\varphi}(x) = \int_X \varphi d\mu$$

pour μ -presque tout x .

La somme $S_n \varphi(x)$ s'appelle une *moyenne de Birkhoff* de φ . La limite (lorsqu'elle existe) $\lim_{n \rightarrow \infty} S_n \varphi(x)$ s'appelle *moyenne orbitale* (ou *temporelle*). L'intégrale $\int_X \varphi d\mu$ s'appelle la *moyenne spatiale* de φ . Le théorème ergodique dit donc, en paraphrasant, que si μ est une probabilité ergodique, alors presque toutes les moyennes temporelles d'une fonction intégrable coïncident avec sa moyenne spatiale.

Ceci traduit une propriété “d'équirépartition” de (presque toute) orbite (voir aussi le chapitre 5 suivant), au sens que l'on peut calculer les intégrales des fonctions en prenant des moyennes des valeurs de cette fonction sur les orbites.

Si μ est une mesure de probabilité ergodique, en prenant pour φ la fonction caractéristique $\mathbb{I}_A$ d'un élément A dans $\mathcal{B}$, le résultat précédent dit que la proportion de temps

$$\lim_{n \rightarrow \infty} \frac{1}{n} \text{Card} \{k \leq n / f^n(x) \in A\}$$

que l'orbite de x passe dans A est égale à $\mu(A)$, pour μ -presque tout point x de X .

Voici un exemple d'application. Pour tout réel x dans $[0, 1[$, on note $0, a_1 \dots a_n \dots$ son écriture décimale (si on a le choix, on prend l'écriture avec a_n nul pour n assez grand).

Corollaire 4.4 *Pour presque tout x dans $[0, 1[$ (pour la mesure de Lebesgue), pour tout chiffre j dans $\{0, \dots, 9\}$, "la proportion de j dans l'écriture décimale de x est égale à $\frac{1}{10}$ " i.e.*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \text{Card} \{k \leq n, a_k = j\} = \frac{1}{10} .$$

Preuve. On prend $X = [0, 1[$, $\mu = dx$, $f(x) = 10x \bmod 1$, $\varphi = \mathbb{I}_{[\frac{j}{10}, \frac{j+1}{10}]}$. On a alors

$$S_n \varphi(x) = \frac{1}{n} \text{Card} \{k \leq n, a_k = j\} \xrightarrow{n \rightarrow \infty} \int_X \varphi d\mu = \frac{1}{10},$$

car f est ergodique. □

Pour φ dans $\mathbb{L}^1(X, \mu; \mathbb{R})$, on note

$$\varphi^*(x) = \sup_{n \geq 1} \frac{1}{n} \sum_{k=0}^{n-1} \varphi(f^k(x)) .$$

Lemme 4.5 (Lemme ergodique maximal) *Pour tout φ dans $\mathbb{L}^1(X, \mu; \mathbb{R})$, on a*

$$\int_{\{\varphi^*(x) > 0\}} \varphi d\mu \geq 0 .$$

Preuve. On pose $\psi_n = \sup \{0, \varphi, \varphi + \varphi \circ f, \dots, \varphi + \varphi \circ f + \dots + \varphi \circ f^{n-1}\}$ et $E_n = \{x \in X / \psi_n(x) > 0\}$. Sur E_n , on a $\psi_n = \varphi + \psi_{n-1} \circ f$. Sur ${}^c E_n$, on a $\psi_n = 0$. Sur X , on a $\psi_{n-1} \circ f \geq 0$. Donc

$$\int_{E_n} \varphi = \int_{E_n} \psi_n - \int_{E_n} \psi_{n-1} \circ f \geq \int_X \psi_n - \int_X \psi_{n-1} \circ f = \int_X (\psi_n - \psi_{n-1}) \geq 0 .$$

Or $\{\varphi^*(x) > 0\} = \bigcup_{n \geq 1} E_n$. Donc $\int_{\{\varphi^*(x) > 0\}} \varphi d\mu \geq 0$. □

Preuve du théorème de Birkhoff 4.3. (i) Pour montrer l'existence de la limite, il suffit de voir que pour tout $\alpha < \beta$, l'ensemble

$$E_{\alpha, \beta} = \{x \in X / \liminf_{n \rightarrow \infty} S_n \varphi(x) < \alpha < \beta < \limsup_{n \rightarrow \infty} S_n \varphi(x)\}$$

est de mesure nulle. Remarquons que $f^{-1}(E_{\alpha, \beta}) = E_{\alpha, \beta}$ (voir (ii) ci-dessous). Comme $\limsup_{n \rightarrow \infty} S_n \varphi \leq \varphi^*$, on a $E_{\alpha, \beta} \cap \{(\varphi - \beta)^* > 0\} = E_{\alpha, \beta}$. De même, comme $\liminf_{n \rightarrow \infty} S_n \varphi \geq -(-\varphi)^*$, on a $E_{\alpha, \beta} \cap \{(\alpha - \varphi)^* > 0\} = E_{\alpha, \beta}$.

Le lemme ergodique maximal appliqué sur la partie $E_{\alpha,\beta}$ aux fonctions $\varphi - \beta$ et $\alpha - \varphi$ donne

$$\int_{E_{\alpha,\beta}} (\varphi - \beta) d\mu \geq 0 \quad \text{et} \quad \int_{E_{\alpha,\beta}} (\alpha - \varphi) d\mu \geq 0.$$

D'où $\int_{E_{\alpha,\beta}} (\alpha - \beta) d\mu \geq 0$, et donc $\mu(E_{\alpha,\beta}) = 0$.

(ii) C'est clair, car la suite $\frac{n+1}{n} S_{n+1}\varphi(x) - S_n\varphi(f(x)) = \frac{1}{n}\varphi(x)$ converge vers 0.

(iii) On peut supposer que φ est positive. Le résultat découle alors du lemme de Fatou (et du fait que f préserve μ) :

$$\int_X \tilde{\varphi} = \int_X \limsup_{n \rightarrow \infty} S_n \varphi \leq \limsup_{n \rightarrow \infty} \int_X S_n \varphi = \int_X \varphi.$$

(iv) Lorsque φ est bornée, cela découle du théorème de convergence dominée de Lebesgue. Le cas général s'en déduit par densité. En effet, pour tout $\epsilon > 0$, on peut trouver ψ bornée telle que $\|\varphi - \psi\|_{\mathbb{L}^1} \leq \frac{\epsilon}{3}$. On a donc, pour tout $n \geq 1$, $\|S_n \varphi - S_n \psi\|_{\mathbb{L}^1} \leq \frac{\epsilon}{3}$ et par (iii), $\|\tilde{\varphi} - \tilde{\psi}\|_{\mathbb{L}^1} \leq \frac{\epsilon}{3}$. Pour n grand, on a alors $\|S_n \psi - \tilde{\psi}\|_{\mathbb{L}^1} \leq \frac{\epsilon}{3}$ et donc $\|S_n \varphi - \tilde{\varphi}\|_{\mathbb{L}^1} \leq \epsilon$. C'est ce que l'on voulait.

(v) Cela résulte de (iv) appliqué à $\varphi|_A$, car $\int_A \varphi = \int_A S_n \varphi$.

(vi) Par (ii), l'application $\tilde{\varphi}$ est constante presque partout, et μ est une probabilité. $\square$

Voici une autre application du théorème ergodique de Birkhoff.

Soit $(\Omega, \mathcal{A}, P)$ un espace de probabilité et $(Y, \mathcal{B})$ un espace mesurable. Si v est une *variable aléatoire* à valeurs dans Y (i.e. une application mesurable de Ω dans Y), on appelle *loi* (ou *distribution*) de φ la mesure image $\nu = v_* P$. Son *espérance* est $E[v] = \int_{\Omega} v dP$. Soit $(v_i)_{i \in \mathbb{N}}$ une suite de variables aléatoires à valeurs dans Y . On dit que les v_i sont *uniformément distribuées* si leurs lois coïncident, et *indépendantes* si pour toute partie finie F de $\mathbb{N}$, pour toute famille finie $(A_i)_{i \in F}$ dans $\mathcal{B}$, on a

$$P\left(\bigcap_{i \in F} v_i^{-1}(A_i)\right) = \prod_{i \in F} P(v_i^{-1}(A_i)).$$

Théorème 4.6 (Loi forte des grands nombres de Kolmogorov) *Soit $(v_i)_{i \in \mathbb{N}}$ une suite de variables aléatoires réelles indépendantes, uniformément distribuées, et intégrables. Alors $\frac{1}{n} \sum_{k=0}^{n-1} v_i$ converge presque sûrement vers $E[v_0]$.*

Preuve. Soit σ le décalage de Bernoulli $\mathbb{R}^{\mathbb{N}} \rightarrow \mathbb{R}^{\mathbb{N}}$, et $v : \Omega \rightarrow \mathbb{R}^{\mathbb{N}}$ l'application définie par $v(\omega) = (v_i(\omega))_{i \in \mathbb{N}}$. Notons $X_k : \mathbb{R}^{\mathbb{N}} \rightarrow \mathbb{R}$ la k -ème projection. Alors $v_k = X_k \circ v$ et $X_k = X_0 \circ \sigma^k$. Soit ν_0 la loi des v_i , et μ la mesure produit $\nu_0^{\mathbb{N}}$. Notons

$$K = \{\omega \in \Omega / \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} v_i(\omega) = E[v_0]\}$$

et

$$K' = \{x \in \mathbb{R}^{\mathbb{N}} / \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} X_0 \circ \sigma^k(x) = E[v_0]\}.$$

Alors $K = v^{-1}(K')$. Puisque le décalage σ est ergodique pour μ (voir l'exercice E.4.3 et le paragraphe 6.4), et puisque $\int_{\mathbb{R}^{\mathbb{N}}} X_0 d\mu = E[v_0]$, on a $\mu(K') = 1$. Puisque les v_i sont indépendantes, on a $v_* P = \prod_{i \in \mathbb{N}} (v_i)_* P = \mu$. Donc $P(K) = v_* P(K') = \mu(K') = 1$. Ceci montre le résultat. $\square$

4.3 Mesures invariantes et mesures ergodiques

Rappelons quelques théorèmes d'intégration et d'analyse fonctionnelle (voir [Rud, CohD]). Ils nous seront utiles pour comprendre l'importance des mesures invariantes pour les systèmes dynamiques topologiques.

Soit X un espace métrisable compact, $\mathcal{B}$ la σ -algèbre des boréliens de X , et $C(X) = C^0(X, \mathbb{C})$ l'algèbre de Banach des applications continues de X dans $\mathbb{C}$, muni de la norme

$$\|\varphi\|_\infty = \sup_{x \in X} |\varphi(x)|.$$

Théorème 4.7 (Théorème de Stone-Weierstrass) *Soit A une sous-algèbre unitaire de $C(X)$ qui sépare les points i.e.*

$$\forall x, y \in X, \quad x \neq y \Rightarrow (\exists \varphi \in A, \quad \varphi(x) \neq \varphi(y)),$$

et qui est stable par conjugaison complexe $\varphi \mapsto \bar{\varphi}$. Alors A est dense dans $C(X)$. □

Corollaire 4.8 *L'espace de Banach $C(X)$ est séparable.*

Preuve. On note D une partie dénombrable dense de X et on prend pour A la sous-algèbre de $C(X)$ engendrée par les applications $y \mapsto d(x, y)$ pour x dans D , qui contient comme partie dense la $\mathbb{Q}[i]$ -sous-algèbre de $C(X)$ engendrée par ces applications. □

Soit $C(X)'$ le dual topologique de $C(X)$ et $\mathcal{M}(X)$ l'ensemble des mesures de probabilités sur $(X, \mathcal{B})$.

Théorème 4.9 (Théorème de Riesz) *Soit $C(X)'_+$ l'ensemble des formes linéaires continues ϕ sur $C(X)$, positives (i.e. $\forall \varphi \in C(X), \quad \varphi \geq 0 \Rightarrow \phi(\varphi) \geq 0$), et telles que $\phi(1) = 1$. Alors l'application*

$$\mu \mapsto (\varphi \mapsto \int_X \varphi d\mu)$$

définit une bijection de $\mathcal{M}(X)$ sur $C(X)'_+$. □

Remarque. C'est pour cela que l'on note parfois $\mu(\varphi)$ pour $\int_X \varphi d\mu$.

On munit $\mathcal{M}(X)$ de la topologie faible. Celle-ci est métrisable, car on peut prendre pour distance

$$d(\mu, \nu) = \sup_{k \in \mathbb{N}} |\mu(\varphi_k) - \nu(\varphi_k)|,$$

où $(\varphi_k)_{k \in \mathbb{N}}$ est une suite dénombrable dense dans $C(X)$ (cela résulte du fait qu'un élément de $\mathcal{M}(X)$, vu comme application de $C(X)$ dans $\mathbb{C}$, est équicontinu (en fait 1-lipschitzien, i.e. $|\mu(f) - \mu(g)| \leq \|f - g\|_\infty$ pour tous f, g dans $C(X)$)).

Proposition 4.10 (Théorème de Banach-Alaoglu) *L'espace $\mathcal{M}(X)$ est compact (pour la topologie faible).*

Preuve. Soit $(\mu_n)_{n \in \mathbb{N}}$ une suite dans $\mathcal{M}(X)$. Reprenons $(\varphi_k)_{k \in \mathbb{N}}$ une suite dénombrable dense dans $C(X)$. Pour k fixé, la suite $(\mu_n(\varphi_k))_{n \in \mathbb{N}}$ est bornée par $\|\varphi_k\|_\infty$. Un procédé d'extraction diagonal permet de construire une sous-suite $\tau(n)$ telle que, pour tout k , la suite $(\mu_{\tau(n)}(\varphi_k))_{n \in \mathbb{N}}$ converge.

On en déduit, par équicontinuité des μ_n , que, pour tout φ dans $\mathcal{C}(X)$, la suite réelle $(\mu_{\tau(n)}(\varphi))_{k \in \mathbb{N}}$ est de Cauchy, donc converge vers un nombre $\phi(\varphi)$.

Par passage à la limite, ϕ est une forme linéaire dans $C(X)'_+$. Le théorème de Riesz assure que ϕ est donnée par une mesure de probabilité μ . La suite $\mu_{\tau(n)}$ converge donc faiblement vers μ . $\square$

Proposition 4.11 *Soit X un espace métrisable compact non vide et $f : X \rightarrow X$ une application continue. Alors il existe au moins une mesure de probabilité borélienne invariante par f sur X .*

Remarques (1) Cette mesure de probabilité invariante est rarement unique (considérer par exemple f valant l'identité). Nous construirons dans les chapitres 11 et 12 de nombreuses mesures invariantes ergodiques pour les systèmes dynamiques symboliques et les automorphismes linéaires du tore.

(2) Cette proposition permet de voir le théorème de récurrence de Birkhoff 3.1 comme une conséquence du théorème de récurrence de Poincaré 3.8. Ce qui donne une démonstration du théorème de récurrence de Birkhoff qui n'utilise pas l'axiome du choix.

Preuve. Partons d'une probabilité ν sur X (par exemple la masse de Dirac δ_x en un point x). Posons

$$\nu_n = \frac{1}{n} \sum_{k=0}^{n-1} (f^k)_* \nu,$$

qui est une probabilité. La compacité de $\mathcal{M}(X)$ assure qu'il existe une sous-suite ν_{k_n} qui converge faiblement vers une probabilité μ . Or la différence $\nu_n - f_*(\nu_n) = \frac{1}{n}(\nu - (f^n)_* \nu)$ converge faiblement vers 0. Donc $f_* \mu = \mu$. $\square$

Notons $\mathcal{M}(X)^f$ le sous-espace de $\mathcal{M}(X)$ des probabilités f -invariantes sur X . C'est un convexe compact de l'espace vectoriel $C(X)'$ des mesures finies sur X , muni de la topologie faible. Le théorème de Krein-Milman [Bre] assure que $\mathcal{M}(X)^f$ est égal à l'adhérence de l'enveloppe convexe de ses points extrémaux. Par le résultat suivant, sous les hypothèses de la proposition 4.11, il existe donc au moins une mesure de probabilité borélienne f -invariante et ergodique sur X .

Proposition 4.12 *Soit X un espace métrisable compact et $f : X \rightarrow X$ une application continue. Les mesures de probabilités boréliennes f -invariantes ergodiques sur X sont exactement les points extrémaux de $\mathcal{M}(X)^f$.*

Preuve. Il s'agit de montrer le résultat suivant. Soient $(X, \mathcal{B}, \mu)$ un espace de probabilité et $f : X \rightarrow X$ une application mesurable préservant μ . Alors μ est ergodique si et seulement si, pour toutes mesures de probabilités f -invariantes μ_1, μ_2 sur X et tout t dans $]0, 1[$ tel que $\mu = t\mu_1 + (1 - t)\mu_2$, on a $\mu = \mu_1 = \mu_2$.

Supposons d'abord μ ergodique. Les ensembles de mesure nulle pour μ sont aussi de mesure nulle pour μ_1 . Le théorème de Radon-Nikodym [Rud] assure qu'il existe une fonction ϕ_1 dans $\mathbb{L}^1(X, \mu)$ telle que $\mu_1 = \phi_1 \mu$. Comme μ et μ_1 sont f -invariantes, on a $\phi_1 = \phi_1 \circ f$ presque partout pour μ . Comme μ est ergodique, l'application ϕ_1 est constante presque partout et les mesures de probabilité μ et μ_1 sont égales.

Réciproquement, soit A_1 une partie mesurable de X telle que $f^{-1}(A_1) = A_1$ et $0 < \mu(A_1) < 1$, et soit $A_2 = {}^c A_1$. On pose $\mu_i = \frac{1}{\mu(A_i)} \mu|_{A_i}$. Les mesures de probabilité μ_i sont f -invariantes, différentes de μ et $\mu = \mu(A_1)\mu_1 + (1 - \mu(A_1))\mu_2$. $\square$

4.4 Exercices

Exercice E.4.1 Soit $(X, \mathcal{B}, \mu)$ un espace mesuré, et $f : X \rightarrow X$ une application mesurable, préservant les ensembles de mesure nulle (i.e. pour tout A dans $\mathcal{B}$, si $\mu(A) = 0$, alors $\mu(f^{-1}(A)) = 0$).

Montrer que si une application $\varphi : X \rightarrow \mathbb{C}$ est presque partout invariante par f (i.e. $\varphi \circ f = \varphi$ μ -presque partout), alors il existe une application $\varphi' : X \rightarrow \mathbb{C}$ invariante par f , et presque partout égale à φ .

Exercice E.4.2 Soit $(X, \mathcal{B}, \mu)$ un espace de probabilité, et $f : X \rightarrow X$ une application mesurable qui préserve μ . Montrer que f est ergodique, si et seulement si pour tous A, B dans $\mathcal{B}$ tels que $\mu(A) > 0$ et $\mu(B) > 0$, il existe n dans $\mathbb{N} - \{0\}$ tel que $\mu(f^{-n}(A) \cap B) > 0$.

Exercice E.4.3 Soit $(X, \mathcal{B}, \mu)$ un espace de probabilité. Montrer que le décalage de Bernoulli σ sur $X^{\mathbb{Z}}$ et sur $X^{\mathbb{N}}$ est ergodique pour la mesure $\mu^{\mathbb{Z}}$ et $\mu^{\mathbb{N}}$ respectivement.

Exercice E.4.4 1. Montrer que, dans le théorème ergodique de Birkhoff, si φ est dans $\mathbb{L}^2(X, \mu)$, alors

- $\tilde{\varphi} \in \mathbb{L}^2(X, \mu)$ et $\|\tilde{\varphi}\|_{\mathbb{L}^2} \leq \|\varphi\|_{\mathbb{L}^2}$,
- si la mesure μ est finie, alors la convergence a lieu dans $\mathbb{L}^2$, i.e.

$$\lim_{n \rightarrow \infty} \|S_n \varphi - \tilde{\varphi}\|_{\mathbb{L}^2} = 0 .$$

2. Soit $(X, \mathcal{B}, \mu)$ un espace mesuré. Un *semi-flot mesurable* est une famille $(f_t)_{t \geq 0}$ telle que $f_0 = id$, $f_{t+s} = f_t \circ f_s$, et $\phi : X \times [0, +\infty[\rightarrow X$ avec $\phi(x, t) = f_t(x)$ est mesurable.

Montrer une version continue du théorème ergodique de Birkhoff: si μ est une mesure de probabilité, et les f_t préservent la mesure, pour tout φ dans $\mathbb{L}^1(X, \mu)$, alors

$$\frac{1}{T} \int_0^T \varphi \circ f_t(x) dt$$

converge, pour μ -presque tout x (ainsi que dans $\mathbb{L}^1$), vers $\tilde{\varphi}(x)$ avec $\tilde{\varphi}$ dans $\mathbb{L}^1(X, \mu)$, f_t -invariante pour tout t , d'intégrale égale à celle de φ .

3. Dans le théorème ergodique de Birkhoff, en supposant la mesure finie, si f est inversible d'inverse mesurable, montrer que $\frac{1}{n} \sum_{k=0}^{n-1} \varphi \circ f^k(x)$ ainsi que $\frac{1}{n} \sum_{k=0}^{n-1} \varphi \circ f^{-k}(x)$ convergent, pour μ -presque tout x , vers $\tilde{\varphi}(x)$.
4. Soit $(X, \mathcal{B}, \mu)$ un espace de probabilité et $f : X \rightarrow X$ une application mesurable préservant μ . Montrer que f est ergodique si et seulement si, pour tout φ dans $\mathbb{L}^1(X, \mu)$, on a $\tilde{\varphi}(x) = \int_X \varphi d\mu$ pour μ -presque tout x .

Exercice E.4.5 Soit $\mathcal{H}$ un espace de Hilbert et $U : \mathcal{H} \rightarrow \mathcal{H}$ une transformation unitaire. On note $\mathcal{H}^U$ le sous-espace fermé $\{h \in \mathcal{H} \mid U(h) = h\}$ et p^U la projection orthogonale sur $\mathcal{H}^U$. On veut montrer que

$$\forall h \in \mathcal{H}, \quad \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} U^k(h) = p^U(h) . \quad (*)$$

- Montrer $(*)$ pour h dans $\mathcal{H}^U$.

- Montrer (*) pour $h = U(h_0) - h_0$ avec h_0 dans $\mathcal{H}$.
- Montrer que l'espace vectoriel $(U - Id)(\mathcal{H})$ est dense dans l'orthogonal de $\mathcal{H}^U$.
- Conclure.

Exercice E.4.6 Soit $(X, \mathcal{B}, \mu)$ un espace de probabilité et $f : X \rightarrow X$ une application mesurable préservant μ . Soit $\varphi : X \rightarrow]0, +\infty[$ une application mesurable, telle que l'application $\log g$ est intégrable. Montrer que

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log (\varphi \circ f^n(x))$$

existe et vaut 0 pour μ -presque tout x dans X .

Exercice E.4.7 Soit $(X, \mathcal{B}, \mu)$ un espace de probabilité et $f : X \rightarrow X$ une application mesurable préservant μ . On suppose f ergodique. Montrer que, pour toutes parties mesurables A, B de X , on a

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \mu(f^{-k}(A) \cap B) = \mu(A)\mu(B) .$$

Exercice E.4.8 Soit $(X, \mathcal{B})$ un espace mesurable et $f : X \rightarrow X$ une application mesurable. Soient μ_1, μ_2 deux mesures de probabilité f -invariantes et ergodiques sur X telles que $\mu_1 \neq \mu_2$. Montrer que μ_1 et μ_2 sont *étrangères* (i.e. il existe une partie mesurable A de X telle que $\mu_1(A) = 1$ et $\mu_2(A) = 0$).

Exercice E.4.9 (Théorème de récurrence de Kac) Soient $(X, \mathcal{B}, \mu)$ un espace de probabilité, $f : X \rightarrow X$ une application mesurable préservant μ et ergodique, et A une partie mesurable de X telle que $\mu(A) > 0$. Pour tout x dans X , on note

$$n_A(x) = \inf\{n \geq 1 \mid f^n(x) \in A\}$$

le *temps de premier retour dans A*. Montrer que

$$\int_A n_A d\mu = 1$$

(autrement dit, la valeur moyenne du temps de premier retour dans A est égale à $\frac{1}{\mu(A)}$).

Exercice E.4.10 Soient X un espace métrique localement compact séparable et $f : X \rightarrow X$ une application continue. On suppose que, pour tous les ouverts non vides U, V de X , il existe n dans $\mathbb{N}$ tel que $f^n(U) \cap V \neq \emptyset$. Montrer qu'il existe une orbite $\{f^n(x), n \in \mathbb{N}\}$ dense dans X .

Exercice E.4.11 Est-ce que la transformation du billard d'une ellipse est ergodique?

Exercice E.4.12 Soit X un espace métrique compact, et $f : X \rightarrow X$ un homéomorphisme. Soit $\mathcal{M} = \mathcal{M}(X)^f$ l'espace des mesures de probabilité sur X , invariantes par f . Soit $C(X) = C^0(X, \mathbb{R})$ l'espace des fonctions continues de X dans $\mathbb{R}$. Pour u, v dans $C(X)$, on note $u \sim v$ s'il existe une application φ dans $C(X)$ telle que

$$u = v + \varphi \circ f - \varphi .$$

1. Montrer que $\sim$ est une relation d'équivalence et que si $u \sim v$, alors $\int u d\mu = \int v d\mu$ pour tout μ dans $\mathcal{M}$.

- 37

Exercice E.4.12 (2) On pourra considérer l'application f_n définie par

$$v_n(x) = \sum_{k=1}^{n-1} \left(1 - \frac{k}{n}\right) u(f^{k-1}(x))$$

et calculer $v_n \circ f - v_n$.

(3) Si $\int u \, d\mu < 0$ pour tout μ dans $\mathcal{M}$, on montrera par l'absurde que u_n est strictement négative pour n assez grand.

(4) On considèrera $u \pm \epsilon$.

5 Unique ergodicité

Pour tout réel x , on note $[x]$ sa *partie entière* et $\{x\} = x - [x]$ sa *partie fractionnaire*.

Rappelons qu'une suite de réels $(x_n)_{n \in \mathbb{N}}$ est dite *équirépartie modulo 1* si, pour tout intervalle $[a, b]$ de $[0, 1]$, on a

$$\lim_{n \rightarrow \infty} \frac{1}{n} \text{Card } \{k \leq n / \{x_k\} \in [a, b]\} = b - a .$$

L'équirépartition modulo 1 de la suite des $P(n)$, où P est un polynôme non constant à coefficients réels et à coefficient dominant irrationnel, comme par exemple $P(X) = \alpha X$ pour α irrationnel, va servir de fil directeur à ce chapitre (ce résultat est dû à H. Weyl).

Nous démontrerons et interpréterons ce résultat à l'aide d'un système dynamique sur le tore $\mathbb{T}^n$ qui est uniquement ergodique, ce qui signifie que toutes les sommes de Birkhoff convergent vers la moyenne spatiale.

En appendice (voir paragraphe 5.4), nous donnerons une démonstration classique directe de cette équirépartition, qui repose sur un critère de Weyl et un lemme de van der Corput. Ce critère de Weyl dit qu'une suite de réels $(x_n)_{n \in \mathbb{N}}$ est équirépartie modulo 1 si et seulement si, pour tout m dans $\mathbb{Z} - \{0\}$, on a

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} e^{2i\pi m x_k} = 0 .$$

On en déduit facilement en particulier (voir paragraphe 5.4) que la suite $(n\alpha)_{n \in \mathbb{N}}$, pour α irrationnel, est équirépartie modulo 1.

Lorsque $x_n = n\alpha$, la moyenne $\frac{1}{n} \sum_{k=0}^{n-1} e^{2i\pi m x_k}$ est une moyenne orbitale de la fonction continue $e_m : t \mapsto e^{2i\pi m t}$ pour la transformation $\theta \mapsto \theta + \alpha \bmod 1$ de l'espace métrisable compact $\mathbb{R}/\mathbb{Z}$. Lorsque $m \neq 0$, la moyenne spatiale de e_m (pour la mesure de Lebesgue) est nulle. Compte tenu de la densité dans $C^0(\mathbb{R}/\mathbb{Z}, \mathbb{C})$ du sous-espace vectoriel engendré par la famille $(e_m)_{m \in \mathbb{Z}}$, ce critère de Weyl pour la suite $(n\alpha)_{n \in \mathbb{N}}$ dit précisément que cette suite est équirépartie modulo 1 si et seulement si les sommes de Birkhoff, pour la transformation $\theta \mapsto \theta + \alpha \bmod 1$, de toute fonction continue sur $\mathbb{R}/\mathbb{Z}$ convergent vers sa moyenne spatiale.

5.1 Unique ergodicité

Soit X un espace métrisable compact non vide et $f : X \rightarrow X$ une application continue.

On dit que f est *uniquement ergodique* si, pour tout φ dans $C(X) = C^0(X, \mathbb{C})$, et pour tout x dans X , les moyennes orbitales $S_n \varphi(x) = \frac{1}{n} \sum_{k=0}^{n-1} \varphi \circ f^k(x)$ convergent vers une limite $L(\varphi)$ qui ne dépend pas de x .

La limite $L(\varphi)$ est une forme linéaire positive sur $C(X)$, telle que $L(1) = 1$. Par le théorème de Riesz, il existe une probabilité borélienne μ sur X telle que $L(\varphi) = \int_X \varphi d\mu$. Par construction, la mesure μ est f -invariante. Lorsque f est sous-entendue, on dit aussi que cette mesure μ est *uniquement ergodique*.

La proposition suivante justifie la terminologie "unique ergodicité"

Proposition 5.1 *L'application f est uniquement ergodique si et seulement s'il existe une unique mesure de probabilité f -invariante μ . Cette mesure de probabilité est alors ergodique.*

Preuve. Supposons f uniquement ergodique, et soit μ' une mesure de probabilité f -invariante. Le théorème ergodique de Birkhoff, appliqué à la mesure μ' , assure que pour tout φ dans $C(X)$, la limite $\tilde{\varphi}(x)$ des sommes de Birkhoff existe μ' -presque partout et vérifie $\mu'(\tilde{\varphi}) = \mu'(\varphi)$. Par hypothèse, l'application $\tilde{\varphi}$ est constante, et égale à $\mu(\varphi)$. Donc $\mu(\varphi) = \mu'(\varphi)$ et $\mu = \mu'$.

Réciproquement, soit x un point de X tel que la suite $S_n\varphi(x)$ ne converge pas vers $\mu(\varphi)$. Comme cette suite est bornée, il existe une sous-suite $\tau(k)$ telle que la limite $\lim_{k \rightarrow \infty} S_{\tau(k)}\varphi(x)$ existe et est différente de $\mu(\varphi)$. La compacité de $\mathcal{M}(X)$ permet de supposer, quitte à extraire de nouveau une sous-suite, que la suite de mesures de probabilité

$$\mu_k = \frac{1}{\tau(k)} \sum_{i=0}^{\tau(k)-1} \delta_{f^i(x)}$$

converge faiblement vers une probabilité μ' . Par construction, μ' est f -invariante et

$$\mu'(\varphi) = \lim_{k \rightarrow \infty} S_{\tau(k)}\varphi(x) \neq \mu(\varphi) .$$

Donc $\mu' \neq \mu$.

La dernière assertion découle de la proposition 4.12. □

5.2 Unique ergodicité des translations sur le tore

L'équirépartition de la suite des $\{n\alpha\}$ s'interprète par l'unique ergodicité des "rotations irrationnelles" sur le cercle $\mathbb{R}/\mathbb{Z}$. Cela se généralise facilement en toutes dimensions. Pour cela, nous commençons par des rappels concernant le tore $\mathbb{T}^N$.

Le tore $\mathbb{T}^N$ est le quotient de $\mathbb{R}^N$ par $\mathbb{Z}^N$, c'est-à-dire l'ensemble des N -uplets $\theta = (\theta_1, \dots, \theta_N)$ où θ_i est un réel modulo 1. C'est un groupe topologique métrisable compact et commutatif.

Il existe sur $\mathbb{T}^N$ une unique mesure de probabilité borélienne invariante par translations : la mesure de Haar $d\theta$, définie par

$$\forall \varphi \in C^0(\mathbb{T}^N, \mathbb{C}), \quad \int_{\mathbb{T}^N} \varphi(\theta) d\theta = \int_{[0,1]^N} \varphi(x) dx$$

où $\dot{x}$ est la classe de x modulo $\mathbb{Z}^N$ et dx la mesure de Lebesgue sur $\mathbb{R}^N$.

Pour m dans $\mathbb{Z}^N$, on note e_m l'application dans $C^0(\mathbb{T}^N, \mathbb{C})$ donnée par

$$e_m(\theta) = e^{2i\pi\langle m, \theta \rangle}$$

où $\langle m, \theta \rangle = \sum_{j=1}^N m_j \theta_j \in \mathbb{R}/\mathbb{Z}$. Par le théorème de Stone-Weierstrass, la famille $(e_m)_{m \in \mathbb{Z}^N}$ engendre un sous-espace vectoriel dense de $C^0(\mathbb{T}^N, \mathbb{C})$.

Pour toute mesure borélienne μ sur $\mathbb{T}^N$, on note $(c_m(\mu))_{m \in \mathbb{Z}^N}$ ses coefficients de Fourier

$$c_m(\mu) = \int_{\mathbb{T}^N} e_{-m} d\mu = \int_{\mathbb{T}^N} e^{-2i\pi\langle m, \theta \rangle} d\mu(\theta) .$$

L'application $\mu \mapsto (c_m(\mu))_{m \in \mathbb{Z}^N}$ est injective. (Cela découle du résultat de densité ci-dessus.)

Proposition 5.2 Soient $a = (a_1, \dots, a_N) \in \mathbb{R}^N$, α la classe modulo $\mathbb{Z}^N$ de a et $\tau_\alpha : \mathbb{T}^N \rightarrow \mathbb{T}^N$ la translation $\theta \mapsto \theta + \alpha$. Les assertions suivantes sont équivalentes :

1. τ_α est uniquement ergodique;

2. τ_α est ergodique pour $d\theta$;
3. τ_α admet une orbite dense;
4. toutes les orbites de τ_α sont denses;
5. τ_α est minimal;
6. les réels $1, a_1, \dots, a_N$ sont linéairement indépendants sur $\mathbb{Q}$.

Preuve. (1) $\Rightarrow$ (2) La mesure $d\theta$ est invariante, donc ergodique par unicité.

(2) $\Rightarrow$ (3) Cela résulte de la proposition 4.2.

(3) $\Rightarrow$ (4) Si l'orbite d'un point θ est dense dans $\mathbb{T}^N$, il en est de même de l'orbite de tous les autres points θ' , car comme les translations τ_α et $\tau_{\theta'-\theta}$ commutent, l'application continue $\tau_{\theta'-\theta}$ envoie l'orbite de θ sur celle de θ' .

(4) $\Rightarrow$ (5) Ceci est général et a déjà été vu (voir la remarque précédent la proposition 3.2).

(5) $\Rightarrow$ (6) Sinon, il existerait un élément non nul m de $\mathbb{Z}^N$ tel que $\langle m, \alpha \rangle = 0$. L'ensemble des θ dans $\mathbb{T}^N$ tels que $\langle m, \theta \rangle = 0$ dans $\mathbb{R}/\mathbb{Z}$ serait donc un fermé non vide τ_α -invariant, contradiction.

(6) $\Rightarrow$ (1) Soit μ une mesure de probabilité borélienne τ_α -invariante. On a donc les égalités

$$c_m(\mu) = \mu(e_{-m}) = \mu(e_{-m} \circ \tau_\alpha) = \mu(e^{-2i\pi\langle m, \alpha \rangle} e_{-m}) = e^{-2i\pi\langle m, \alpha \rangle} c_m(\mu) .$$

Par hypothèse, lorsque $m \neq 0$, l'élément $\langle m, \alpha \rangle$ est non nul dans $\mathbb{R}/\mathbb{Z}$, donc $c_m(\mu) = 0$. On en déduit que, pour tout m , $c_m(\mu) = c_m(d\theta)$. L'injectivité de la décomposition en série de Fourier assure que $\mu = d\theta$. $\square$

5.3 Equirépartition modulo 1 de la suite $P(n)$

Proposition 5.3 Soit α un nombre irrationnel. L'homéomorphisme f du tore $\mathbb{T}^N$ donné par

$$f(\theta_1, \dots, \theta_N) = (\theta_1 + \alpha, \theta_2 + \theta_1, \dots, \theta_N + \theta_{N-1})$$

est uniquement ergodique.

Preuve. Soit μ une mesure de probabilité f -invariante sur $\mathbb{T}^N$. On veut montrer que $\mu = d\theta$. Pour tout m dans $\mathbb{Z}$, le coefficient de Fourier $c_m(\mu)$ vérifie, par invariance de μ , l'égalité $c_m(\mu) = \mu(e_{-m}) = \mu(e_{-m} \circ f)$. Donc on a

$$c_{(m_1, \dots, m_N)}(\mu) = e^{-2i\pi m_1 \alpha} c_{(m_1+m_2, m_2+m_3, \dots, m_{N-1}+m_N, m_N)}(\mu) . \quad (*)$$

Ces égalités à elles seules ne suffisent pas à montrer que $c_m(\mu) = 1$ si $m = 0$ et $c_m(\mu) = 0$ sinon. Nous devons exploiter le fait que les nombres $c_m(\mu)$ sont les coefficients de Fourier d'une mesure de probabilité. Notons

$$(\varphi, \psi)_{\mathbb{L}^2(\mu)} = \int_{\mathbb{T}^N} \overline{\varphi(\theta)} \psi(\theta) d\mu(\theta)$$

le produit scalaire de $\mathbb{L}^2(\mathbb{T}^N, \mu)$.

Montrons par récurrence sur d que si $m = (m_1, \dots, m_d, 0, \dots, 0) \neq 0$ alors $c_m(\mu) = 0$. Si $m = (m_1, 0, \dots, 0) \neq 0$, alors les égalités (*) assurent que $c_m(\mu) = 0$. Supposons le résultat vrai pour $d-1$, et soit $m = (m_1, \dots, m_d, 0, \dots, 0)$.

Fixons M dans $\mathbb{N} - \{0\}$. Pour $k = 1, \dots, M$, les fonctions $\varphi_k = e_m \circ f^k$ sont de la forme $\lambda_k e_{m(k)}$ avec λ_k un nombre complexe de module 1, $m(k) = (m'_1, m'_2, \dots, m'_d, 0, \dots, 0)$ et $m'_d = m_d$. Par hypothèse de récurrence, on a donc

$$(\varphi_k, \varphi_{k'})_{\mathbb{L}^2(\mu)} = \overline{\lambda_k} \lambda_{k'} \mu(e_{m(k')} - m(k)) = \overline{\lambda_k} \lambda_{k'} c_{m(k) - m(k')}(\mu) = \begin{cases} 1 & \text{si } k = k' \\ 0 & \text{sinon} \end{cases}$$

Autrement dit, la famille $(\varphi_k)_{k=1, \dots, M}$ est orthonormée. D'après le théorème de Pythagore, la fonction $\psi = e_0$ vérifie

$$\sum_{k=1}^M |(\varphi_k, \psi)_{\mathbb{L}^2(\mu)}|^2 \leq \|\psi\|_{\mathbb{L}^2(\mu)}^2 = 1.$$

On calcule facilement que $(\varphi_k, \psi)_{\mathbb{L}^2(\mu)} = \mu(e_{-m} \circ f^k) = \mu(e_{-m}) = c_m(\mu)$. Donc, on a $|c_m(\mu)|^2 \leq \frac{1}{M}$. Comme ceci est vrai pour tout M , on a $c_m(\mu) = 0$.

Ceci prouve que μ et $d\theta$ ont mêmes coefficients de Fourier. Donc $\mu = d\theta$. $\square$

Corollaire 5.4 *Soit P un polynôme non constant à coefficients réels dont le coefficient dominant est irrationnel. Alors la suite $n \mapsto P(n)$ est équirépartie modulo 1.*

En particulier, pour tout irrationnel α , la suite $(n\alpha)_{n \in \mathbb{N}}$ est équirépartie modulo 1.

Preuve. Notons N le degré de P et $\frac{1}{N!}\alpha$ le coefficient dominant de P . Posons $P_N = P$ et, par récurrence descendante, $P_j(X) = P_{j+1}(X+1) - P_{j+1}(X)$ pour $j = 0, \dots, N-1$. Le polynôme P_j est de degré j et $P_0 = \alpha$. Posons $\theta_n = (P_1(n), \dots, P_N(n)) \bmod \mathbb{Z}^N$. On a donc $f(\theta_n) = \theta_{n+1}$ pour tout entier n . Comme f est uniquement ergodique, les points $\theta_n = f^n(\theta_0)$ sont équirépartis sur $\mathbb{T}^N$, i.e. pour toute application continue φ sur $\mathbb{T}^N$, on a

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \varphi(\theta_k) = \int_{\mathbb{T}^N} \varphi d\theta.$$

En choisissant φ ne dépendant que de la dernière coordonnée, on en déduit que la suite $P(n)$ est équirépartie modulo 1. $\square$

5.4 Appendice: critère de Weyl et lemme de van der Corput

Dans ce paragraphe, nous donnons une démonstration “directe” de l'équirépartition des $P(n)$, où P est un polynôme non constant à coefficients réels et à coefficient dominant irrationnel.

Proposition 5.5 (Critère de Weyl) *Une suite de réels $(x_n)_{n \in \mathbb{N}}$ est équirépartie modulo 1 si et seulement si, pour tout m dans $\mathbb{Z} - \{0\}$, on a*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} e^{2i\pi m x_k} = 0.$$

Lemme 5.6 Soit $\mathcal{F}_b([0, 1], \mathbb{C})$ l'espace vectoriel des fonctions bornées sur $[0, 1]$, muni de la norme $\|\varphi\|_\infty = \sup_{t \in [0, 1]} |\varphi(t)|$, et E le sous-espace vectoriel formé des fonctions φ telles que $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \varphi(x_k) = \int_0^1 \varphi(t) dt$. Alors E est fermé dans $\mathcal{F}_b([0, 1], \mathbb{C})$.

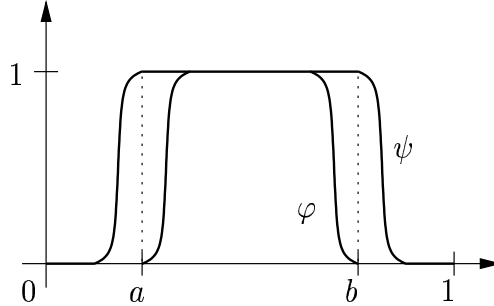
Preuve. Soit φ dans l'adhérence $\overline{E}$. Pour tout $\epsilon > 0$, il existe donc ψ dans E tel que $\|\varphi - \psi\|_\infty \leq \epsilon$. Notons $S_n \varphi = \frac{1}{n} \sum_{k=0}^{n-1} \varphi(x_k)$. Pour tout $n \geq 1$, on a donc $|S_n \varphi - S_n \psi| \leq \epsilon$ et $|\int_0^1 \varphi - \int_0^1 \psi| \leq \epsilon$. Comme $\lim_{n \rightarrow \infty} |S_n \psi - \int_0^1 \psi| = 0$, on a

$$\limsup_{n \rightarrow \infty} \left| S_n \varphi - \int_0^1 \varphi \right| \leq 2\epsilon .$$

Ceci est vrai pour tout ϵ . Donc $\lim_{n \rightarrow \infty} |S_n \varphi - \int_0^1 \varphi| = 0$. D'où φ est dans E , et E est fermé. $\square$

Preuve du critère de Weyl: Supposons tout d'abord que $(x_n)_{n \in \mathbb{N}}$ est équirépartie. Par hypothèse, les fonctions caractéristiques d'intervalles sont dans E . Comme toute fonction continue sur $[0, 1]$ est limite uniforme de fonctions en escalier, l'ensemble E contient $C^0([0, 1], \mathbb{C})$. En particulier, les fonctions $e_m : x \mapsto e^{2i\pi m x}$ sont dans E . On conclut en remarquant que $\int_0^1 e_m = 0$ pour $m \neq 0$.

Réciproquement, par hypothèse, pour tout m dans $\mathbb{Z}$, la fonction e_m est dans E . Comme toute fonction 1-périodique, de classe C^2 , est limite uniforme de sa série de Fourier, l'ensemble E contient les restrictions à $[0, 1]$ de ces fonctions. Soit $[a, b]$ un intervalle contenu dans $[0, 1]$. Pour tout $\epsilon > 0$, on peut donc trouver deux fonctions φ, ψ dans E telles que $\varphi \leq \mathbb{I}_{[a, b]} \leq \psi$ et $\int_0^1 \psi - \varphi \leq \epsilon$.



On a alors $S_n \varphi \leq S_n \mathbb{I}_{[a, b]} \leq S_n \psi$ et

$$b - a - \epsilon \leq \int_0^1 \varphi \leq \liminf_{n \rightarrow \infty} S_n \mathbb{I}_{[a, b]} \leq \limsup_{n \rightarrow \infty} S_n \mathbb{I}_{[a, b]} \leq \int_0^1 \psi \leq b - a + \epsilon .$$

Ceci est vrai pour tout ϵ . Donc $\lim_{n \rightarrow \infty} S_n \mathbb{I}_{[a, b]} = b - a$. $\square$

Corollaire 5.7 Pour tout irrationnel α , la suite $(n\alpha)_{n \in \mathbb{N}}$ est équirépartie modulo 1.

Preuve. Ceci résulte du critère de Weyl, car pour $m \neq 0$, on a $e^{2i\pi m \alpha} \neq 1$, et donc

$$\left| \frac{1}{n} \sum_{k=0}^{n-1} e^{2i\pi m k \alpha} \right| = \frac{1}{n} \left| \frac{1 - e^{2i\pi m n \alpha}}{1 - e^{2i\pi m \alpha}} \right| \leq \frac{2}{n |1 - e^{2i\pi m \alpha}|} \xrightarrow{n \rightarrow \infty} 0 . \square$$

Donnons une application amusante de ce corollaire. Considérons la suite 1, 2, 4, 8, 16, 32, 64, 128, 256, ... des puissances de 2, puis la suite 1, 2, 4, 8, 1, 3, 6, 1, 2, 5, 1, ... des premiers chiffres de ces nombres. La fréquence d'apparition d'un chiffre x dans cette suite est donnée par le corollaire suivant.

Corollaire 5.8 *Soit x dans $\{1, \dots, 9\}$. Alors*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \text{Card} \{k \leq n / \text{ le premier chiffre de } 2^k \text{ est } x\} = \log_{10}(1 + \frac{1}{x}) .$$

Preuve. Soit $\alpha = \log_{10} 2$. Alors le premier chiffre de 2^k est x si et seulement s'il existe p dans $\mathbb{N}$ tel que $x 10^p \leq 2^k < (x+1)10^p$, ce qui équivaut à fait que $\{k\alpha\}$ appartient à l'intervalle $[\log_{10} x, \log_{10}(x+1)[$. Comme α est irrationnel, la limite ci-dessus existe, et est égale à $\log_{10}(x+1) - \log_{10} x = \log_{10}(1 + \frac{1}{x})$. $\square$

Donnons maintenant une autre preuve du corollaire 5.4 d'équirépartition de la suite des $\{P(n)\}$.

Lemme 5.9 *Soit $(z_n)_{n \in \mathbb{N}}$ une suite de nombres complexes de module 1 telle que pour tout $m \geq 1$, $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} z_{k+m} z_k = 0$. Alors on a $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} z_k = 0$.*

Preuve. Sinon, comme cette suite est bornée, il existerait une sous-suite $\tau(n)$ telle que $\lim_{n \rightarrow \infty} \frac{1}{\tau(n)} \sum_{k=0}^{\tau(n)-1} z_k = c \neq 0$. Posons $y_n = z_n - c$, fixons un entier $M > \frac{1}{|c|^2}$ et étudions la suite positive

$$b_n = \frac{1}{\tau(n)} \sum_{k=0}^{\tau(n)-1} \left| \sum_{i=0}^{M-1} y_{k+i} \right|^2 .$$

C'est une somme de M^2 termes $b_n^{i,j}$ avec $b_n^{i,j} = \frac{1}{\tau(n)} \sum_{k=0}^{\tau(n)-1} y_{k+i} \overline{y_{k+j}}$, que l'on exprime à l'aide des z_n :

$$b_n^{i,j} = \frac{1}{\tau(n)} \sum_{k=0}^{\tau(n)-1} z_{k+i} \overline{z_{k+j}} - \frac{c}{\tau(n)} \sum_{k=0}^{\tau(n)-1} \overline{z_{k+j}} - \frac{\bar{c}}{\tau(n)} \sum_{k=0}^{\tau(n)-1} \overline{z_{k+j}} z_{k+i} + |c|^2 .$$

Le premier terme tend vers 0 si $i \neq j$ par hypothèse, et les deux termes suivants tendent vers $-|c|^2$. Donc la limite $\lim_{n \rightarrow \infty} b_n^{i,j}$ vaut $1 - |c|^2$ si $i = j$ et $-|c|^2$ sinon. En sommant les M^2 termes, on obtient $\lim_{n \rightarrow \infty} b_n = M - M^2|c|^2 < 0$, ce qui contredit la positivité de la suite b_n . $\square$

Lemme 5.10 (Lemme de van der Corput) *Soit $(a_n)_{n \in \mathbb{N}}$ une suite réelle telle que, pour tout $m \geq 1$, la suite $(a_{n+m} - a_n)_{n \in \mathbb{N}}$ est équirépartie modulo 1. Alors la suite $(a_n)_{n \in \mathbb{N}}$ est équirépartie modulo 1.*

Preuve. Ceci résulte du critère de Weyl et du lemme 5.9 ci-dessus appliqué aux suites $(z_n = e^{2i\pi m a_n})_{n \in \mathbb{N}}$ pour m un entier non nul. $\square$

Autre preuve du corollaire 5.4. Ce corollaire résulte d'un raisonnement par récurrence sur le degré d du polynôme P , à partir du cas $d = 1$ démontré dans le corollaire 5.7, et à l'aide du lemme de van der Corput. $\square$

5.5 Exercices

Exercice E.5.1 Soient f une transformation uniquement ergodique d'un espace métrique compact X , μ l'unique mesure de probabilité borélienne f -invariante sur X et S le support de μ .

- (1) Montrer que tout fermé non vide f -invariant de X contient S .
- (2) Lorsque f est bijective, a-t-on toujours l'égalité $S = X$?

5.6 Indications pour la résolution des exercices

Exercice E.5.1 (1) Soit F un tel fermé. Supposons par l'absurde qu'il existe un point x_0 de S hors de F . Soit $\epsilon = d(x_0, F) > 0$. La fonction positive $\varphi : x \mapsto \sup\{0, \epsilon - d(x, x_0)\}$ est continue, nulle sur F et strictement positive en x_0 . Par définition, les moyennes de Birkhoff $S_n\varphi(x) = \frac{1}{n} \sum_{k=0}^{n-1} \varphi(f^k(x))$ convergent vers $\int_X \varphi d\mu$. or, pour tout x dans F , on a $S_n\varphi(x) = 0$. Comme x_0 appartient à S , on a $\int_X \varphi d\mu > 0$. Contradiction.

(2) Non, par exemple, on peut prendre $X = \mathbb{R} \cup \{\infty\}$ et $f : X \rightarrow X$ l'application définie par $f(\infty) = \infty$ et $f(x) = x + 1$ pour x dans $\mathbb{R}$.

6 Mélange

Le mélange est une propriété d'un système dynamique plus forte que l'ergodicité. Nous étudierons dans ce chapitre trois exemples où cette propriété est vérifiée : les systèmes symboliques, certaines transformations linéaires du tore et le développement en fractions continues. Un autre exemple important, généralisant ces exemples, est celui de certains systèmes de Markov, sur lequel nous reviendrons aux chapitres 11 et 12.

6.1 Transformations mélangeantes

Soit $(X, \mathcal{B}, \mu)$ un espace de probabilité et $f : X \rightarrow X$ une application mesurable qui préserve μ .

On dit que f est *mélangeante* si

$$\forall A, B \in \mathcal{B}, \quad \lim_{n \rightarrow \infty} \mu(A \cap f^{-n}(B)) = \mu(A)\mu(B) .$$

Remarques. (1) Une application mélangeante est ergodique. En effet, si $A = f^{-1}(A)$, alors $\mu(A) = \mu(A \cap f^{-n}(A)) = \mu(A)^2$, et donc $\mu(A)$ vaut 0 ou 1.

(2) La propriété de mélange signifie que, pour n grand, les événements A et $f^{-n}(B)$ sont asymptotiquement indépendants.

(3) Certains auteurs disent *mélange fort* au lieu de mélange, le *mélange faible* signifiant que l'action diagonale de f sur $X \times X$ est ergodique pour $\mu \times \mu$.

(4) Lorsque f est sous-entendue, on dit aussi que la mesure μ est *mélangeante*.

L'interprétation $\mathbb{L}^2$ du mélange est la suivante.

Proposition 6.1 *L'application f est mélangeante si et seulement si*

$$\forall \varphi, \psi \in \mathbb{L}^2(X, \mu), \quad \lim_{n \rightarrow \infty} \int_X \varphi \cdot (\psi \circ f^n) d\mu = \left(\int_X \varphi d\mu \right) \left(\int_X \psi d\mu \right) . \quad (*)$$

En outre, il suffit de vérifier cette égalité () pour φ, ψ dans un sous-espace H' dense de $\mathbb{L}^2(X, \mu)$ pour qu'elle soit vraie pour tout φ, ψ dans $\mathbb{L}^2(X, \mu)$.*

Preuve. Il suffit de prendre $\varphi = \mathbb{I}_A$ et $\psi = \mathbb{I}_B$ pour montrer que la condition est suffisante.

Réciproquement, si f est mélangeante, alors l'égalité est vraie, par combinaison linéaire, pour des fonctions étagées. Par densité des fonctions étagées dans $\mathbb{L}^2(X, \mu)$, il suffit de montrer la dernière assertion de cette proposition. Pour cela, il suffit de montrer que pour tout φ dans $\mathbb{L}^2(X, \mu)$, le sous-espace H_φ des fonctions ψ dans $\mathbb{L}^2(X, \mu)$ telles que l'égalité (*) est vraie, est fermé dans $\mathbb{L}^2(X, \mu)$.

Comme l'égalité (*) est vraie pour φ constante (car f préserve μ), on peut supposer que $\int_X \varphi d\mu = 0$ et que φ n'est pas presque partout nulle. Soit ψ dans $\overline{H_\varphi}$. Pour tout $\epsilon > 0$, il existe ψ_1 dans H_φ tel que $\|\psi - \psi_1\|_{\mathbb{L}^2} \leq \frac{\epsilon}{2\|\varphi\|_{\mathbb{L}^2}}$. On a alors $|\int_X \varphi \cdot (\psi \circ f^n) d\mu| \leq T_1 + T_2$ avec

$$T_1 = \left| \int_X \varphi \cdot (\psi - \psi_1) \circ f^n d\mu \right| \leq \|\varphi\|_{\mathbb{L}^2} \|\psi - \psi_1\|_{\mathbb{L}^2} \leq \epsilon/2$$

et

$$T_2 = \left| \int_X \varphi \cdot (\psi_1 \circ f^n) d\mu \right| \leq \epsilon/2$$

pour n assez grand car $\int_X \varphi d\mu = 0$. Ceci prouve que $\lim_{n \rightarrow \infty} \int_X \varphi \cdot (\psi \circ f^n) d\mu = 0$. Donc ψ est dans H_φ , et H_φ est fermé. C'est ce que l'on voulait. $\square$

6.2 Transformations linéaires du tore

Les transformations linéaires du tore $\mathbb{T}^N$ sont des exemples particulièrement importants. D'une part, ils se prêtent bien aux calculs. D'autre part, la complexité de leur dynamique est assez représentative de situations plus générales. Commençons par quelques résultats préliminaires.

Soit $M = (m_{ij})_{1 \leq i, j \leq N}$ une matrice $N \times N$ à coefficients entiers. On note $f_M : \mathbb{T}^N \rightarrow \mathbb{T}^N$ l'application (bien définie car les coefficients de M sont entiers) qui envoie $\theta = (\theta_i)_{1 \leq i \leq N}$ sur

$$f_M(\theta) = \left(\sum_{j=1}^N m_{ij} \theta_j \right)_{1 \leq i \leq N}.$$

De sorte que le diagramme suivant commute, avec $p : x \mapsto \dot{x} = x \bmod \mathbb{Z}^N$:

$$\begin{array}{ccc} \mathbb{R}^N & \xrightarrow{M} & \mathbb{R}^N \\ p \downarrow & & \downarrow p \\ \mathbb{T}^N & \xrightarrow{f_M} & \mathbb{T}^N \end{array}.$$

Lemme 6.2 *On suppose que le déterminant de M est non nul. Alors*

1. f_M est surjective,
2. f_M préserve la mesure de Haar $\lambda = d\theta$ sur $\mathbb{T}^N$,
3. pour tout θ dans $\mathbb{T}^N$, on a $\text{Card } f_M^{-1}(\theta) = |\det M|$.

Preuve. (1) En effet, M et p sont surjectives.

(2) Rappelons que $\tau_\theta : x \mapsto x + \theta$ est la translation par θ sur $\mathbb{T}^N$. Comme on a l'égalité $f_M \circ \tau_\theta = \tau_{f_M(\theta)} \circ f_M$, pour tout θ dans $\mathbb{T}^N$, la mesure de probabilité image $(f_M)_* \lambda$ est invariante par $\tau_{f_M(\theta)}$. Comme f_M est surjective, et que λ est la seule mesure de probabilité invariante par translation sur $\mathbb{T}^N$ (voir section 5.2), on a $(f_M)_* \lambda = \lambda$.

(3) Ceci découle d'un argument immédiat de degré, mais nous donnons une preuve (certes plus longue) qui ne nécessite pas de prérequis sur les revêtements ni sur la géométrie différentielle.

L'application f_M est un morphisme de groupes. Comme M est injective, son noyau $f_M^{-1}(0)$ est un sous-groupe discret, donc fini car $\mathbb{T}^N$ est compact. On note $d = \text{Card } f_M^{-1}(0)$, $A_\epsilon =]-\epsilon, \epsilon[^N$ et $B_\epsilon = p(A_\epsilon)$ avec $0 < \epsilon \leq (2N \sup_{i,j} |m_{ij}|)^{-1}$. On a, par la commutativité du diagramme ci-dessus,

$$p^{-1} f_M^{-1} f_M p(A_\epsilon) = M^{-1} p^{-1} p M(A_\epsilon) = M^{-1}(M A_\epsilon + \mathbb{Z}^N) = A_\epsilon + M^{-1} \mathbb{Z}^N.$$

Donc

$$f_M^{-1}(f_M(B_\epsilon)) = \bigcup_{\theta \in f_M^{-1}(0)} (\theta + B_\epsilon)$$

et cette réunion est une réunion disjointe. On a alors

$$\lambda(f_M(B_\epsilon)) = \lambda(f_M^{-1}(f_M(B_\epsilon))) = d \lambda(B_\epsilon).$$

D'autre part, la formule de changement de variables dans $\mathbb{R}^N$ donne

$$\lambda(f_M(B_\epsilon)) = |\det M| \lambda(B_\epsilon) .$$

Donc $d = |\det M|$. □

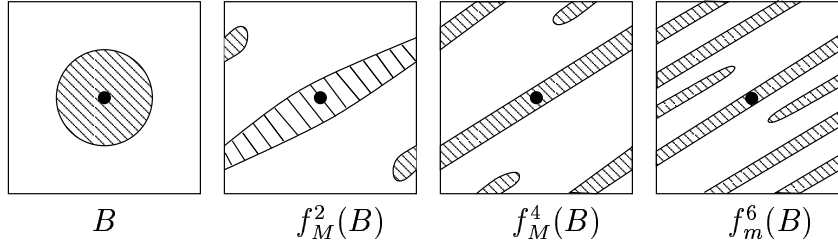
Proposition 6.3 *Soit $f_M : \mathbb{T}^N \rightarrow \mathbb{T}^N$ l'application donnée par une matrice M à coefficients entiers et telle que $\det M$ est non nul. Les assertions suivantes sont équivalentes :*

- (i) f_M est mélangeante (pour $d\theta$),
- (ii) f_M est ergodique (pour $d\theta$),
- (iii) aucune valeur propre de M n'est racine de l'unité.

L'assertion (iii) peut aussi s'écrire

$$\forall n \in \mathbb{N} - \{0\}, \quad \det (M^n - 1) \neq 0 .$$

Voici quelques dessins avec la matrice $M = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$ qui montrent bien le mélange.



Remarquons qu'en prenant $N = 1$, on retrouve le fait déjà vu au paragraphe 4.1 que l'application $g : \theta \rightarrow N\theta$ de $\mathbb{R}/\mathbb{Z}$ dans $\mathbb{R}/\mathbb{Z}$ est ergodique pour tout N dans $\mathbb{N} - \{0, 1\}$.

Preuve. On note toujours $e_m : \theta \mapsto e^{2i\pi\langle m, \theta \rangle}$. On a $e_m \circ f_M = e_{m'}$ avec $m' = {}^t M m$ l'image de m par la matrice transposée de M .

L'implication (i) $\Rightarrow$ (ii) résulte de la remarque (1) de la section 6.1.

Montrons la contraposée de l'implication (ii) $\Rightarrow$ (iii). Supposons donc qu'il existe $n \geq 1$ tel que $\det (M^n - 1) = 0$. Alors, soit m_0 un élément non nul de $\mathbb{Z}^N$ tel que ${}^t M^n m_0 = m_0$. Notons $m_i = {}^t M^i m_0$. La fonction continue $\varphi = e_{m_0} + e_{m_1} + \dots + e_{m_{n-1}}$ est alors f_M -invariante et non constante. Donc f_M n'est pas ergodique.

Pour l'implication (iii) $\Rightarrow$ (i), on veut montrer que

$$\forall \varphi, \psi \in \mathbb{L}^2(\mathbb{T}^N, d\theta), \quad \lim_{n \rightarrow \infty} \int_{\mathbb{T}^N} \varphi \cdot (\psi \circ f_M^n) d\theta = \left(\int_{\mathbb{T}^N} \varphi d\theta \right) \left(\int_{\mathbb{T}^N} \psi d\theta \right) .$$

Par la proposition 6.1, il suffit de le vérifier pour φ, ψ dans une partie dense de $\mathbb{L}^2(\mathbb{T}^N, d\theta)$. Par le théorème de Stone-Weierstrass, on peut donc supposer que $\varphi = e_\ell$ et $\psi = e_m$ avec ℓ, m dans $\mathbb{Z}$.

Si $m = 0$, l'assertion est évidente. Supposons donc $m \neq 0$. En particulier $\int_{\mathbb{T}^N} \psi d\theta = 0$. On a l'égalité $\psi \circ f_M^n = e_{m_n}$ avec $m_n = {}^t M^n m \in \mathbb{Z}^N$. Par hypothèse, ${}^t M$ n'a pas de point périodique dans $\mathbb{Z}^N - \{0\}$. Donc la suite m_n prend des valeurs deux à deux distinctes. En particulier, il existe n_0 dans $\mathbb{N}$ tel que pour $n \geq n_0$, on a $m_n \neq \ell$. On a alors $\int_{\mathbb{T}^N} \varphi \cdot (\psi \circ f_M^n) d\theta = 0$. Ce qui prouve le mélange. □

6.3 Fractions continues

6.3.1 Préliminaires

Il est facile d'approcher tout réel x par un rationnel $\frac{p}{q}$ de dénominateur $q = N$ avec une erreur d'au plus $\frac{1}{2N}$. Mieux, si on autorise un dénominateur $q \leq N$, on peut l'approcher avec une erreur d'au plus $\frac{1}{q^2}$ (c'est un *théorème de Dirichlet*). Cela résulte du lemme des tiroirs : on découpe l'intervalle $[0, 1]$ en N intervalles $I_k = [\frac{k}{N}, \frac{k+1}{N}]$; parmi les valeurs de $\{nx\}$ pour $0 \leq n \leq N$, deux d'entre elles $\{n_1x\}$ et $\{n_2x\}$ sont dans le même intervalle et $(n_2 - n_1)x$ diffère d'un entier p par au plus $\frac{1}{N}$. Donc (en supposant $n_2 > n_1$) avec $q = n_2 - n_1$, on a

$$|x - \frac{p}{q}| \leq \frac{1}{qN} \leq \frac{1}{q^2}.$$

Le développement en fraction continue est une méthode qui permet de trouver rapidement cette meilleure approximation. Voici cette méthode. Rappelons que si u est un nombre réel, alors on note $[u]$ la partie entière de u et $\{u\} = u - [u]$ sa partie fractionnaire.

Pour x dans $]0, 1]$, on pose $a_1(x) = [\frac{1}{x}]$, et

$$f(x) = \left\{ \frac{1}{x} \right\} = \frac{1}{x} - a_1(x).$$

Pour $n \geq 1$, on pose, lorsque c'est défini, $a_n = a_n(x) = a_1(f^{n-1}x)$. Remarquons que $a_n(x) \geq 1$. Comme $y = \frac{1}{a_1(y) + f(y)}$ pour tout y dans $]0, 1]$, on a l'égalité

$$x = \frac{1}{a_1 + \frac{1}{a_2 + \frac{1}{a_3 + \frac{1}{a_4 + \frac{1}{a_5 + \frac{1}{a_6 + \frac{1}{a_7 + \frac{1}{a_8 + \frac{1}{a_9 + \frac{1}{a_{10} + f^{10}(x)}}}}}}}}}}}}}}.$$

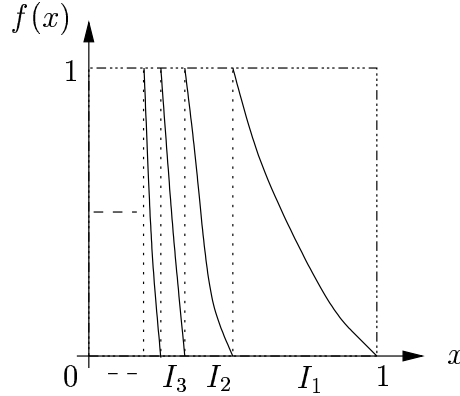
En réduisant au même dénominateur, on obtient une expression de la forme

$$x = \frac{p_n + p_{n-1} x_n}{q_n + q_{n-1} x_n}$$

avec $x_n = f^n(x) \in [0, 1]$. Les entiers p_n et q_n sont obtenus par les égalités

$$\begin{pmatrix} p_1 & q_1 \\ p_0 & q_0 \end{pmatrix} = \begin{pmatrix} 1 & a_1 \\ 0 & 1 \end{pmatrix} \quad \text{et} \quad \begin{cases} p_{n+1} = a_{n+1} p_n + p_{n-1} \\ q_{n+1} = a_{n+1} q_n + q_{n-1} \end{cases} \quad (*)$$

Tout ceci est bien défini tant que $f^n(x)$ ne s'annule pas.



Lemme 6.4 *i) Le développement en fraction continue $x \mapsto (a_1(x), a_2(x), \dots)$ s'arrête si et seulement si x est rationnel.*

ii) L'application $x \mapsto (a_1(x), a_2(x), \dots)$ est une bijection de $[0, 1] - \mathbb{Q}$ sur $(\mathbb{N} - \{0\})^{\mathbb{N} - \{0\}}$.

Preuve. L'assertion i) est évidente. Pour ii), nous ne vérifierons que l'injectivité, par des arguments qui nous réserveront plus tard. Le reste des affirmations est laissé en exercice.

Notons $I_\ell = a_1^{-1}(\ell) =]\frac{1}{\ell+1}, \frac{1}{\ell}]$ pour tout ℓ dans $\mathbb{N} - \{0\}$, et, pour $a = (a_1, \dots, a_n)$ avec a_i dans $\mathbb{N} - \{0\}$, notons

$$\Delta_a = I_{a_1} \cap f^{-1}(I_{a_2}) \cap \dots \cap f^{-n+1}(I_{a_n}) = \{x \in X / a_1(x) = a_1, \dots, a_n(x) = a_n\}.$$

D'après les calculs précédents, l'application f^n induit une bijection de Δ_a sur $[0, 1[$, dont l'inverse, noté $\psi_a : [0, 1[\rightarrow \Delta_a$ est donné par la formule

$$\psi_a(y) = \frac{p_n + p_{n-1}y}{q_n + q_{n-1}y}$$

où les p_n, q_n sont définis par les formules (*). On vérifie que q_n tend en croissant vers $+\infty$ et que

$$p_n q_{n-1} - p_{n-1} q_n = (-1)^{n-1}.$$

On en déduit que Δ_a est un intervalle, qui a pour extrémités $\frac{p_n}{q_n}$ et $\frac{p_n + p_{n-1}}{q_n + q_{n-1}}$, et donc que la longueur de Δ_a vérifie

$$\ell(\Delta_a) = \frac{1}{q_n(q_n + q_{n-1})} \leq \frac{1}{q_n^2} \xrightarrow{n \rightarrow \infty} 0.$$

D'où le résultat. $\square$

Soit X l'espace topologique $[0, 1] - \mathbb{Q}$. L'application $f : X \rightarrow X$, définie par $f(x) = \{\frac{1}{x}\}$, est appelée la *transformation de Gauss*. Pour tout irrationnel x , on a, en posant $a_n = a_n(x - [x])$,

$$x = \lim_{n \rightarrow +\infty} [x] + \frac{1}{a_1 + \frac{1}{a_2 + \frac{1}{a_3 + \dots + \frac{1}{a_{n-1} + \frac{1}{a_n}}}}}.$$

6.3.2 Ergodicité de la transformation de Gauss

L'étude des fractions continues est intimement liée à celle du système dynamique mesurable (X, f) .

Lemme 6.5 (Gauss) *La transformation de Gauss préserve la mesure de probabilité (dite mesure de Gauss)*

$$\mu = \frac{1}{\log 2} \left(\frac{dx}{1+x} \right) .$$

Preuve. Pour toute fonction positive mesurable φ sur $[0, 1]$, on a

$$\begin{aligned} \int_0^1 \varphi(f(t)) \frac{dt}{1+t} &= \sum_{n=1}^{\infty} \int_{\frac{1}{n+1}}^{\frac{1}{n}} \varphi\left(\frac{1}{t} - n\right) \frac{dt}{1+t} = \sum_{n=1}^{\infty} \int_0^1 \varphi(s) \frac{ds}{(s+n)^2(1+\frac{1}{s+n})} = \\ &= \int_0^1 \varphi(s) \sum_{n=1}^{\infty} \frac{1}{(s+n)(s+n+1)} ds = \int_0^1 \varphi(s) \frac{ds}{1+s} . \end{aligned}$$

□

Proposition 6.6 *La transformation de Gauss est ergodique pour la mesure de Gauss.*

On en déduit la propriété suivante sur la fréquence d'apparition d'un entier ℓ dans le développement en fractions continues d'un réel x .

Corollaire 6.7 *Pour presque tout x dans $[0, 1]$, pour tout ℓ dans $\mathbb{N} - \{0\}$, on a*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \text{Card} \{k \leq n / a_k(x) = \ell\} = \log_2 \left(1 + \frac{1}{\ell(\ell+2)} \right) .$$

Preuve. Le théorème ergodique de Birkhoff, appliqué à la transformation de Gauss f et à la fonction intégrable $\varphi = \mathbb{I}_{\{x / a_1(x) = \ell\}}$, affirme que, pour presque tout x , cette limite existe et est égale à

$$\int \varphi d\mu = \frac{1}{\log 2} \int_{\frac{1}{\ell+1}}^{\frac{1}{\ell}} \frac{dx}{1+x} = \frac{1}{\log 2} \left(\log\left(1 + \frac{1}{\ell}\right) - \log\left(1 + \frac{1}{\ell+1}\right) \right) = \log_2 \left(1 + \frac{1}{\ell(\ell+2)} \right) . \quad \square$$

Pour démontrer l'ergodicité de la transformation de Gauss, nous aurons besoin de l'inégalité de gauche dans le lemme suivant (l'inégalité de droite sera utile pour montrer le mélange).

Lemme 6.8 (Condition de Renyi) *Il existe des constantes $c, c' > 0$ telles que, pour tout borélien A de $[0, 1]$ et tout $a = (a_1, \dots, a_n)$ dans $(\mathbb{N} - \{0\})^n$, on a*

$$c \mu(\Delta_a) \mu(A) \leq \mu(\Delta_a \cap f^{-n}(A)) \leq c' \mu(\Delta_a) \mu(A) .$$

Preuve. Il suffit de montrer cette assertion avec la mesure de Lebesgue $\lambda = dx$ au lieu de la mesure de Gauss μ , car on a $\lambda \leq (2 \log 2)\mu \leq 2\lambda$. On rappelle que f^n est une bijection de Δ_a sur $[0, 1[$, d'inverse l'homographie ψ_a . On écrit alors

$$\lambda(\Delta_a \cap f^{-n}(A)) = \int_{\psi_a(A)} dx = \int_A |\psi'_a(y)| dy .$$

Comme $|\psi'_a(y)| = \frac{1}{(q_n + q_{n-1}y)^2}$, $\lambda(\Delta_a) = \frac{1}{q_n(q_n + q_{n-1})}$ et $q_{n-1} \leq q_n$, on a

$$\frac{1}{2}\lambda(\Delta_a) \leq |\psi'_a(y)| \leq 2\lambda(\Delta_a) .$$

D'où le résultat. $\square$

Preuve de la proposition 6.6. Soit A un borélien tel que $f^{-1}(A) = A$. D'après la condition de Renyi, la mesure borélienne $Y \mapsto \nu(Y) = \mu(A \cap Y) - c \mu(A)\mu(Y)$ prend des valeurs positives pour tout $Y = \Delta_a$. Elle prend donc, par approximation, des valeurs positives sur toutes les fonctions continues positives. Donc ν est une mesure positive. On a donc $\nu({}^cA) = -c \mu(A)\mu({}^cA) \geq 0$. Donc $\mu(A)$ vaut 0 ou 1, et la mesure de Gauss μ est ergodique. $\square$

6.3.3 Taux de croissance du développement en fractions continues

La proposition suivante assure que, pour presque tout x dans $[0, 1]$ et tout $\epsilon > 0$, la suite $a_n(x)$ est un $O(n^{1+\epsilon})$, mais n'est pas un $O(n \log n)$. En particulier, l'ensemble des x dans $[0, 1]$ dont le développement en fraction continue est borné est de mesure nulle.

Proposition 6.9 (Théorème de Borel-Bernstein) *Soient $(k_n)_{n \geq 1}$ une suite croissante d'entiers positifs non nuls et $E = \{x \in X / \forall n_0 \geq 1, \exists n \geq n_0, a_n(x) \geq k_n\}$.*

i) *Si $\sum_{n \geq 1} k_n^{-1} < \infty$, alors E est de mesure de Lebesgue nulle.*

ii) *Si $\sum_{n \geq 1} k_n^{-1} = \infty$, alors cE est de mesure de Lebesgue nulle.*

Preuve. C'est une adaptation du fameux lemme de Borel-Cantelli, voir par exemple [Spr]. On note

$$A_k = \{x \in X / a_1(x) \geq k\} =]0, \frac{1}{k}] - \mathbb{Q} .$$

On a donc $k^{-1} \leq (2 \log 2)\mu(A_k) \leq 2k^{-1}$. Remarquons que la mesure de Lebesgue et la mesure de Gauss sur $[0, 1]$ sont absolument continues l'une par rapport à l'autre, c'est-à-dire qu'elles ont les mêmes ensembles de mesure nulle.

Montrons l'assertion i). Comme $E = \bigcap_{n_0 \geq 1} \bigcup_{n \geq n_0} f^{-n+1}(A_{k_n})$, on a, pour tout $n_0 \geq 1$, puisque f préserve μ ,

$$\mu(E) \leq \sum_{n \geq n_0} \mu(f^{-n+1}(A_{k_n})) = \sum_{n \geq n_0} \mu(A_{k_n}) \leq \frac{1}{\log 2} \sum_{n \geq n_0} k_n^{-1} .$$

Or cette dernière expression tend vers 0 quand n_0 tend vers l'infini, car la série converge. Donc $\mu(E) = 0$.

Montrons maintenant l'assertion ii). On a besoin, comme dans le lemme de Borel-Cantelli, d'une propriété d'indépendance entre les parties $f^{-n+1}(A_{k_n})$. Cette propriété est donnée par la condition de Renyi (lemme 6.8), qui dit que l'on a une inégalité

$$\mu(\Delta \cap f^{-n}(A)) \geq c \mu(\Delta) \mu(A)$$

pour tout borélien A de X et toute partie Δ dans la σ -algèbre $\mathcal{B}_n$ sur X engendrée par les $\Delta_a \cap X$ avec $a = (a_1, \dots, a_n)$. (En effet, par récurrence sur n , deux $\Delta_a \cap X$ sont égaux ou disjoints, et le complémentaire d'un $\Delta_a \cap X$ dans X est une union dénombrable de $\Delta_a \cap X$. Donc $\mathcal{B}_n$ est l'ensemble des unions disjointes au plus dénombrables de $\Delta_a \cap X$. De plus, la formule ci-dessus est stable par union disjointe dénombrable de Δ .)

Fixons $n_0 \geq 1$, et posons

$$C_n = \bigcup_{j=n_0}^n f^{-j+1}(A_{k_j})$$

pour $n \geq n_0$. La suite $({}^c C_n)_{n \geq n_0}$ est décroissante pour l'inclusion, d'intersection F_{n_0} , avec ${}^c E = \bigcup_{n_0 \geq 1} F_{n_0}$. Pour montrer que $\mu({}^c E) = 0$, il suffit donc de voir que $\lim_{n \rightarrow \infty} \mu({}^c C_n) = 0$. Or on a

$$\mu({}^c C_{n-1} - {}^c C_n) = \mu({}^c C_{n-1} \cap f^{-n+1}(A_{k_n})) \geq c \mu({}^c C_{n-1}) \mu(A_{k_n})$$

car C_{n-1} est dans $\mathcal{B}_{n-1}$ (car $A_k = \bigsqcup_{\ell \geq k} I_\ell \cap X$, et par récurrence). On en déduit que $\mu({}^c C_n) \leq \mu({}^c C_{n-1})(1 - c \mu(A_{k_n}))$, puis que

$$\mu({}^c C_n) \leq \prod_{j=n_0+1}^n (1 - c \mu(A_{k_j})) \leq \prod_{j=n_0+1}^n \left(1 - \frac{c}{(2 \log 2) k_j}\right).$$

Or ce produit tend vers 0 quand n tend vers l'infini, car la série $\sum k_j^{-1}$ diverge. D'où $\lim_{n \rightarrow \infty} \mu({}^c C_n) = 0$. $\square$

6.3.4 Mélange de la transformation de Gauss

Avec un peu plus de travail, on peut préciser la proposition 6.6.

Proposition 6.10 *La transformation de Gauss est mélangeante pour la mesure de Gauss.*

Le lemme suivant nous donne un renseignement très proche du mélange.

Lemme 6.11 *Il existe des constantes $c, c' > 0$ telles que, pour tout boréliens A, B de $[0, 1]$, on a*

$$c \mu(A) \mu(B) \leq \liminf_{n \rightarrow \infty} \mu(A \cap f^{-n}(B)) \leq \limsup_{n \rightarrow \infty} \mu(A \cap f^{-n}(B)) \leq c' \mu(A) \mu(B). \quad (*)$$

Preuve. D'après la condition de Renyi, c'est vrai pour $A = \Delta_a$, donc aussi pour A dans la famille $\mathcal{B}' = \bigcup_{n \geq 1} \mathcal{B}_n$. Par un argument d'approximation, donné ci-dessous, cette assertion est donc vraie pour tout borélien.

Notons $\mathcal{A} \Delta \mathcal{A}' = (A - A') \cup (A' - A)$ la différence symétrique de deux parties A, A' . Comme la σ -algèbre $\mathcal{B}$ des boréliens de $[0, 1]$ est engendrée par l'algèbre de Boole $\mathcal{B}'$, pour tout A dans $\mathcal{B}$, on a

$$\forall \epsilon > 0, \quad \exists A' \in \mathcal{B}', \quad \mu(A \Delta A') \leq \epsilon.$$

(En effet, l'ensemble des boréliens A qui vérifient cette propriété est une σ -algèbre.) L'assertion (*) pour A' permet de montrer l'assertion pour A à ϵ près. Comme ϵ est arbitraire, l'assertion (*) pour A s'en déduit. $\square$

Preuve de la proposition 6.10. Posons $Y = [0, 1] \times [0, 1]$, $g : Y \rightarrow Y$ définie (presque partout) par $(x, x') \mapsto (f(x), f(x'))$, $\nu = \mu \times \mu$ la mesure produit sur Y , et $\nu_n = \mu \times (f^n)_* \mu$. Dire que f est mélangeante, c'est dire que ν_n converge faiblement vers ν (voir la proposition 6.1). Par compacité faible de l'espace $\mathcal{M}(Y)$ des mesures de probabilité sur Y , il suffit de montrer que si une sous-suite ν_{n_k} converge faiblement vers une mesure de probabilité ν_∞ , alors $\nu_\infty = \nu$.

Lemme 6.12 *Pour tous boréliens P, Q de Y , on a*

$$c^2 \nu(P) \nu(Q) \leq \liminf_{n \rightarrow \infty} \nu(P \cap g^{-n}(Q)) .$$

Preuve. Cela résulte du lemme précédent, lorsque P et Q sont des pavés $P = A_1 \times A_2$, $Q = B_1 \times B_2$, et donc aussi lorsque P et Q sont des réunions finies disjointes de pavés. Le cas général s'en déduit par approximation, comme ci-dessus. $\square$

Lemme 6.13 *La mesure ν est ergodique pour g .*

Preuve. Si Q est un borélien de Y tel que $g^{-1}(Q) = Q$, alors en prenant $P = {}^c Q$ dans le lemme précédent, on trouve $\nu(Q) \nu({}^c Q) = 0$. $\square$

Lemme 6.14 *Pour tout borélien P de Y , on a $\nu_\infty(P) \leq c' \nu(P)$.*

Preuve. Ceci résulte du lemme 6.11 lorsque P est un pavé $P = A \times B$ avec A et B fermés, car $\nu_\infty(P) \leq \limsup \nu_n(P) \leq c' \nu(P)$. Le cas général s'en déduit par approximation, comme ci-dessus. $\square$

Terminons la preuve de la proposition 6.10. Les mesures ν_∞ et ν sont des mesures de probabilité g -invariantes sur Y . La mesure ν_∞ est absolument continue par rapport à ν par le lemme précédent. Par le théorème de Radon-Nikodym, il existe φ dans $\mathbb{L}^1(Y, \nu)$ telle que $\nu_\infty = \varphi \nu$. La fonction φ est (ν -presque partout) g -invariante. Comme ν est ergodique, l'application φ est constante (ν -presque partout), et $\nu_\infty = \nu$. $\square$

6.4 Mélange des systèmes symboliques

Ce paragraphe est un cas particulier du chapitre 11. Nous montrerons aussi dans le chapitre 12 qu'une transformation hyperbolique du tore est conjuguée à un système symbolique, donc sa propriété de mélange découle aussi de celle des systèmes symboliques. De même, la transformation de Gauss est conjuguée à un système symbolique sur un alphabet dénombrable. Mais comme son entropie est finie (voir chapitre 9), un théorème d'Ornstein [Orn] assure que la transformation de Gauss est conjuguée à un système symbolique (sur un alphabet fini).

Proposition 6.15 *Un système de Bernoulli est mélangeant.*

Preuve. Soit $(Y, \mathcal{A}, \nu)$ un espace de probabilité, $(X, \mathcal{B}, \mu)$ son espace produit (par $\mathbb{N}$ ou $\mathbb{Z}$) et σ le décalage (voir paragraphe 2.3). D’après la proposition 6.1, comme l’espace vectoriel engendré par les fonctions caractéristiques des cylindres est dense dans $\mathbb{L}^2(X, \mu)$, il suffit de montrer que pour tous les cylindres A, B , on a

$$\lim_{n \rightarrow \infty} \mu(A \cap \sigma^{-n}(B)) = \mu(A)\mu(B)$$

Or si on décale suffisamment de fois le cylindre B , il sera de “support” disjoint de celui de A , et par définition de la mesure produit, le résultat en découle. $\square$

6.5 Exercices

Exercice E.6.1 Soit M une matrice $N \times N$ à coefficients entiers telle que $\det M \neq 0$. On suppose que $|\lambda_i| \neq 1$ pour tout i , où $\lambda_1, \dots, \lambda_N$ sont les valeurs propres de M . On note $\mathcal{N}(n, M)$ le nombre de points x de $\mathbb{T}^N$ tels que $M^n x = x$.

1. Montrer que

$$\mathcal{N}(n, M) = \prod_{j=1}^N |\lambda_j^n - 1|.$$

2. En déduire la valeur de $\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathcal{N}(n, M)$.

Exercice E.6.2 Soit $f : [0, 1] \rightarrow [0, 1]$ la transformation du boulanger définie par $f(x) = 2x$ si $x \leq \frac{1}{2}$ et $f(x) = 2(1 - x)$ si $x \geq \frac{1}{2}$.

1. Montrer que f préserve la mesure de Lebesgue $\lambda = dx$.
2. Montrer que f est ergodique pour λ .
3. Montrer que f est mélangeante pour λ .

Exercice E.6.3 Soit X un espace métrisable compact, $f : X \rightarrow X$ une application continue et μ une mesure de probabilité borélienne f -invariante sur X , de support X . On suppose que f est mélangeante pour μ . Montrer que f est *topologiquement mélangeante*, i.e. que, pour tous les ouverts U, V non vides de X , il existe n_0 dans $\mathbb{N}$ tel que, pour tout $n \geq n_0$, on a $f^n(U) \cap V \neq \emptyset$.

Exercice E.6.4 Soit $(X, \mathcal{B}, \mu)$ un espace de probabilité et $f : X \rightarrow X$ une application mesurable qui préserve μ . On suppose que f est mélangeante pour μ .

1. Montrer que l’application $f \times f : X \times X \rightarrow X \times X$ donnée par $(x, y) \mapsto (f(x), f(y))$ est mélangeante pour la mesure produit $\mu \times \mu$.
2. En déduire que f est faiblement mélangeante pour μ (i.e. que $f \times f$ est ergodique pour $\mu \times \mu$).

Exercice E.6.5 Est-ce qu’une rotation du cercle $\mathbb{R}/\mathbb{Z}$ est mélangeante?

Exercice E.6.6 Soient $(X, \mathcal{A}, \mu, f)$ et $(Y, \mathcal{B}, \nu, g)$ deux espaces de probabilité, munis d'une application mesurable et préservant la mesure.

(1) Montrer que f est ergodique si et seulement si

$$\forall \varphi, \psi \in \mathbb{L}^2(X, \mu), \quad \frac{1}{n} \sum_{k=0}^{n-1} \int_X \varphi(\psi \circ f^k) d\mu \xrightarrow{n \rightarrow \infty} \left(\int_X \varphi d\mu \right) \left(\int_X \psi d\mu \right).$$

(2) Montrer que si f est mélangeante et g ergodique, alors $f \times g : (x, y) \mapsto (f(x), g(y))$ est ergodique sur l'espace de probabilité produit $(X \times Y, \mathcal{A} \times \mathcal{B}, \mu \times \nu)$. (On rappelle que les combinaisons linéaires des fonctions de la forme $\phi(x, y) = \varphi(x)\varphi'(y)$, avec φ dans $\mathbb{L}^2(X, \mu)$ et φ' dans $\mathbb{L}^2(Y, \nu)$, sont denses dans $\mathbb{L}^2(X \times Y, \mu \times \nu)$.)

(3) Donner un exemple d'espace de probabilité $(X, \mathcal{A}, \mu)$ et de transformation ergodique $f : X \rightarrow X$, tels que $f \times f$ n'est pas ergodique sur l'espace de probabilité produit $(X \times X, \mathcal{A} \times \mathcal{A}, \mu \times \mu)$.

Exercice E.6.7 Soient $(X, \mathcal{A}, \mu)$ un espace de probabilité et $f : X \rightarrow X$ une application mesurable et préservant la mesure. Pour k dans $\mathbb{N} - \{0, 1\}$, on dit que f est *mélangeante d'ordre k* si pour toutes les parties mesurables $A_1, \dots, A_k$ de X , et toutes les applications $\tau_1, \dots, \tau_k$ de $\mathbb{N}$ dans $\mathbb{N}$ telles que $\lim_{n \rightarrow \infty} \tau_{i+1}(n) - \tau_i(n) = +\infty$ pour $i = 1, \dots, k-1$, on a

$$\mu \left(\bigcap_{i=1}^k f^{-\tau_i(n)}(A_i) \right) \xrightarrow{n \rightarrow \infty} \prod_{i=1}^k \mu(A_i).$$

(1) Montrer qu'un système dynamique symbolique est mélangeant d'ordre k pour tout k dans $\mathbb{N} - \{0, 1\}$.

(2) Soit N dans $\mathbb{N} - \{0, 1\}$. Montrer que l'application $f_N : \theta \mapsto N\theta$ sur le cercle $\mathbb{R}/\mathbb{Z}$, est mélangeante d'ordre k , pour tout k dans $\mathbb{N} - \{0, 1\}$, pour la mesure de Haar.

(3) Soit N dans $\mathbb{N} - \{0\}$ et M une matrice inversible $N \times N$ à coefficients entiers, sans valeur propre racine de l'unité. Montrer que l'application $f_M : x \bmod \mathbb{Z}^N \mapsto Mx \bmod \mathbb{Z}^N$ sur le tore $\mathbb{R}^N/\mathbb{Z}^N$, est mélangeante d'ordre k , pour tout k dans $\mathbb{N} - \{0, 1\}$, pour la mesure de Haar.

Exercice E.6.8 Soient p dans $]0, 1[$, $q = 1 - p$, $X = \mathbb{R}/\mathbb{Z}$ et $f : X \rightarrow X$ l'application $x \mapsto 2x$ de doublement de l'angle sur le cercle. On identifie l'ensemble X avec l'ensemble $[0, 1[$. Pour toute mesure de probabilité borélienne μ sur X , on note $\mu_1 = T(\mu)$ la mesure de probabilité borélienne sur X définie par

$$T(\mu)(A) = q\mu(f(A \cap [0, \frac{1}{2}[))) + p\mu(f(A \cap [\frac{1}{2}, 1[)))$$

pour tout borélien A de X . On note $\mu_n = T^n(\mu)$.

Pour tous les entiers k, d tels que $0 \leq k < 2^d$, on note $I_{k,d}$ l'intervalle $I_{k,d} = [\frac{k}{2^d}, \frac{k+1}{2^d}[$, on écrit k en base 2

$$k = \sum_{i=0}^{d-1} a_i 2^i$$

avec $a_i = 0$ ou $a_i = 1$, et on note $w_k = \sum_{i=0}^{d-1} a_i$.

A 1) Calculer $\mu_1([0, \frac{1}{2}[)$ et $\mu_1([\frac{1}{2}, 1[)$.

2) Montrer que l'on a $\mu_d(I_{k,d}) = p^{w_k} q^{d-w_k}$.

3) Montrer que, pour $n \geq d$, on a $\mu_n(I_{k,d}) = p^{w_k} q^{d-w_k}$.

4) Montrer que si une sous-suite $(\mu_{n_i})_{i \in \mathbb{N}}$ converge faiblement vers une mesure de probabilité borélienne μ_∞ , alors, pour tout $k < 2^d$, on a $\mu_\infty(I_{k,d}) = p^{w_k} q^{d-w_k}$. (On pourra montrer tout d'abord que $\mu_\infty(\frac{k}{2^d}) = 0$.)

5) Montrer qu'il existe une unique mesure de probabilité borélienne $\nu = \nu_p$ sur X telle que, pour tous k, d avec $k < 2^d$, on a $\nu(I_{k,d}) = p^{w_k} q^{d-w_k}$.

6) Montrer que, pour toute mesure de probabilité borélienne μ , la suite $(\mu_n)_{n \in \mathbb{N}}$ converge vers cette mesure de probabilité ν .

7) Déterminer ν lorsque $p = \frac{1}{2}$.

B 1) Montrer que ν est l'unique mesure de probabilité borélienne sur X telle que $T(\nu) = \nu$.

2) Montrer que ν est f -invariante.

3) Montrer que, pour tout borélien A de X , on a $\nu(f^{-d}(A) \cap I_{k,d}) = \nu(A)\nu(I_{k,d})$.

4) Montrer que f est mélangeante pour ν .

5) Pour tout x dans X (identifié avec $[0, 1[$), on note $x = 0, b_1 b_2 b_3 \dots$ le *développement dyadique* de x , autrement dit,

$$x = \sum_{i \in \mathbb{N} - \{0\}} b_i 2^{-i}$$

avec $b_i = 0$ ou $b_i = 1$. Calculer, pour ν -presque tout x dans X , la limite

$$\lim_{n \rightarrow \infty} \frac{1}{n} \text{Card} \{i \leq n / b_i = 1\}.$$

6) Montrer que, pour ν -presque tout x , l'orbite $\{f^n(x) / n \in \mathbb{N}\}$ est dense dans X .

7) Montrer que si $p \neq p'$, alors les mesures de probabilité ν_p et $\nu_{p'}$ sont étrangères (i.e. il existe une partie de mesure nulle pour l'une et de mesure pleine pour l'autre).

6.6 Indications pour la résolution des exercices

Exercice E.6.2 (2) On montrera que toute partie A telle que $f^{-1}(A) = A$ vérifie aussi $g^{-1}(A) = A$ pour les transformations g définies par $x \mapsto 1 - x$ et $x \mapsto \{2x\}$.

(3) On notera I l'un des intervalles $[\frac{k}{2^n}, \frac{k+1}{2^n}]$ pour $0 \leq k < 2^n$, et on montrera que, pour tout borélien A de $[0, 1]$, on a $\lambda(f^{-n}(A) \cap I) = \lambda(A)\lambda(I)$.

On peut aussi résoudre a la fois (2) et (3) en utilisant une conjugaison (au sens de la mesure) avec un système dynamiques symbolique.

Exercice E.6.5 Non. Soit f la rotation d'angle θ . Si $A = B = [0, \frac{1}{100}]$, si n_k est une suite d'entiers tels que $|(n_k \theta - \frac{1}{2}) \pmod{1}| \leq \frac{1}{100}$ (qui existe si θ est irrationnel, par densité de

l'orbite), alors comme f préserve la longueur, les intervalles A et $f^{n_k}(B)$ sont disjoints, donc $\mu(A \cap f^{-n_k}(B)) = 0$, alors que $\mu(A)\mu(B) = \frac{1}{10000}$.

Exercice E.6.6 On note $S_n\varphi = \frac{1}{n} \sum_{k=0}^{n-1} \varphi \circ f^k$.

(1) Si f est ergodique, alors par le théorème ergodique, $S_n\psi$ converge dans $\mathbb{L}^2(X, \mu)$ vers la fonction constante de valeur $\int_X \psi d\mu$, pour tout ψ dans $\mathbb{L}^2(X, \mu)$. Par continuité du produit scalaire de l'espace de Hilbert $\mathbb{L}^2(X, \mu)$, pour tout φ dans $\mathbb{L}^2(X, \mu)$, on en déduit donc que $\int_X \varphi S_n\psi d\mu$ converge vers $\int_X \varphi d\mu \int_X \psi d\mu$, ce qui montre le sens direct.

Réciproquement, pour tout ψ dans $\mathbb{L}^2(X, \mu)$, si la suite des $\int_X \varphi S_n\psi d\mu$ converge vers $\int_X \varphi d\mu \int_X \psi d\mu$ pour tout φ dans $\mathbb{L}^2(X, \mu)$, alors $S_n\psi - \int_X \psi d\mu$ converge (faiblement, donc fortement) dans $\mathbb{L}^2(X, \mu)$ vers la fonction nulle. Ceci implique que $S_n\psi - \int_X \psi d\mu$ converge presque partout vers la fonction nulle. Si ψ est (presque partout) invariante par f , il en découle que ψ est presque partout constante. Donc f est ergodique.

(2) Par le sens réciproque de la question (1) appliqué au produit, par densité des combinaisons linéaires de fonctions produits dans $\mathbb{L}^2(X \times Y, \mu \times \nu)$, et par bilinéarité, il suffit de montrer que, pour tout φ, φ' dans $\mathbb{L}^2(X, \mu)$ et ψ, ψ' dans $\mathbb{L}^2(Y, \nu)$, on a

$$\frac{1}{n} \sum_{k=0}^{n-1} \int_{X \times Y} \varphi(x) \varphi'(y) \psi(f^k(x)) \psi'(g^k(y)) d\mu(x) d\nu(y) \xrightarrow{n \rightarrow \infty} \int_{X \times Y} \varphi(x) \varphi'(y) d\mu(x) d\nu(y) \int_{X \times Y} \psi(x) \psi'(y) d\mu(x) d\nu(y) .$$

Ceci équivaut à

$$\frac{1}{n} \sum_{k=0}^{n-1} \left(\int_X \varphi(\psi \circ f^k) d\mu \right) \left(\int_Y \varphi'(\psi' \circ g^k) d\nu \right) \xrightarrow{n \rightarrow \infty} \left(\int_X \varphi d\mu \int_X \psi d\mu \right) \left(\int_Y \varphi' d\nu \int_Y \psi' d\nu \right) .$$

Comme f est mélangeante, on a $\int_X \varphi(\psi \circ f^k) d\mu \xrightarrow{n \rightarrow \infty} \int_X \varphi d\mu \int_X \psi d\mu$, et le résultat découle donc de l'ergodicité de g par le sens direct de la question (1).

(3) Soit f une rotation d'angle irrationnel sur le cercle $X = \mathbb{R}/\mathbb{Z}$, identifié avec $[0, 1[$. Alors f est ergodique pour la mesure de Lebesgue, mais $E = \{(\theta, \theta') \in X \times X / |\theta - \theta'| \leq 1/2\}$ est un borélien de $X \times X$, vérifiant $(f \times f)^{-1}(E) = E$, d'intérieur non vide, et de complémentaire d'intérieur non vide, donc de mesure non nulle et non totale pour la mesure de Haar produit. Donc $f \times f$ n'est pas ergodique.

Exercice E.6.7 Remarquons que f est mélangeant à l'ordre 2 si et seulement si f est mélangeante. La même méthode que dans le cours montre les deux questions.

(1) Soit Λ un alphabet fini, $X = \Lambda^E$ avec $E = \mathbb{N}$ ou $\mathbb{Z}$, et $f = \sigma$ le décalage à gauche sur X . Comme la σ -algèbre de X est engendrée par les cylindres, il suffit de vérifier l'assertion lorsque les A_i sont des cylindres. Soit E_i une partie finie de E , $A_{i,p}$ une partie de Λ pour tout $i = 1, \dots, k$ et p dans E , avec $A_{i,p} = \Lambda$ si $p \notin E_i$. Soit A_i le cylindre $\prod_{p \in E} A_{i,p}$. Alors si n est assez grand, l'ensemble $(E_{i+1} + \tau_{i+1}(n)) \cap ((E_1 + \tau_1(n)) \cup \dots \cup (E_i + \tau_i(n)))$ est

vide pour tout $i = 1, \dots, k-1$, donc, par définition du décalage et de la valeur de la mesure produit sur les cylindres,

$$\mu \left(\bigcap_{i=1}^k f^{-\tau_i(n)}(A_i) \right) = \prod_{i=1}^k \prod_{p \in E_i} \frac{\text{Card } A_{i,p}}{\text{Card } \Lambda} = \prod_{i=1}^k \mu(A_i)$$

ce qui montre le résultat.

(2) Une méthode rapide est de conjuguer mesurablement le système dynamique de cette question au système dynamique de la question précédente, en prenant les développements N -adiques. Voici une méthode qui s'adapte facilement pour traiter aussi le cas de la troisième question.

De la même manière que pour le mélange, étant donné un sous-espace dense $\mathcal{H}$ de $\mathbb{L}^2(X, \mu)$, on montre que f est mélangeante à l'ordre k si et seulement si, pour toutes les applications $\varphi_1, \dots, \varphi_k$ dans $\mathcal{H}$, et toutes les applications $\tau_1, \dots, \tau_k$ de $\mathbb{N}$ dans $\mathbb{N}$ telles que $\lim_{n \rightarrow \infty} \tau_{j+1}(n) - \tau_j(n) = +\infty$ pour $j = 1, \dots, k-1$, on a

$$\int_X \prod_{j=1}^k \varphi_j \circ f^{\tau_j(n)} d\mu \xrightarrow{n \rightarrow \infty} \prod_{j=1}^k \int_X \varphi_j d\mu . \quad (*)$$

Soit $X = \mathbb{R}/\mathbb{Z}$ et μ la mesure de Haar sur X . Pour m dans $\mathbb{Z}$, on note comme d'habitude $e_m : \theta \mapsto e^{2i\pi m\theta}$. Par multilinéarité, et par le théorème de Stone -Weierstrass, il suffit de montrer que pour tous $m_1, \dots, m_k$ dans $\mathbb{Z}$, on a la formule (*) pour $\varphi_j = e_{m_j}$. On raisonne par récurrence sur k . Comme remarqué ci-dessus, c'est vrai pour $k = 2$. Si $m_k = 0$, alors le résultat est vrai par l'hypothèse au rang $k-1$. On suppose donc m_k non nul. En particulier,

$$\int_X \varphi_k d\mu = 0 = \prod_{j=1}^k \int_X \varphi_j d\mu .$$

On a $\varphi_j \circ f_N^{\tau_j(n)} = e_{\ell_{j,n}}$ avec $\ell_{j,n} = N^{\tau_j(n)} m_j$. Comme $\tau_k(n) - \tau_j(n) \xrightarrow{n \rightarrow \infty} +\infty$ pour $j = 1, \dots, k-1$, et comme $N \geq 2$, pour n assez grand, on a

$$N^{\tau_k(n)} > \frac{|\sum_{j=1}^{k-1} N^{\tau_j(n)} m_j|}{|m_k|} .$$

Donc pour n assez grand, $\sum_{i=1}^k N^{\tau_i(n)} m_i$ est non nul, et par conséquent

$$\int_X \prod_{i=1}^k \varphi_i \circ f^{\tau_i(n)} d\mu = 0 ,$$

ce qu'il fallait montrer.

Exercice E.6.8 A 1) On a $\mu_1([0, \frac{1}{2}[) = q$ et $\mu_1([\frac{1}{2}, 1[) = p$.

2) Parmi les intervalles $I_{k,d}, f(I_{k,d}), \dots, f^{d-1}(I_{k,d})$, exactement w_k sont inclus dans $[\frac{1}{2}, 1[$ et les $d - w_k$ autres sont inclus dans $[0, \frac{1}{2}[$. Donc

$$\mu_d(I_{k,d}) = p^{w_k} q^{d-w_k} \mu([0, 1[) = p^{w_k} q^{d-w_k} .$$

3) Appliquons la question A 2) à la mesure $\mu' = \mu_{n-d}$. On a

$$\mu_n(I_{k,d}) = \mu'_d(I_{k,d}) = p^{w_k} q^{d-w_k} .$$

4) Soit x_0 un point de X qui peut sécrire $x_0 = \frac{k}{2^d}$. Choisissons une fonction positive continue φ_0 telle que $\varphi_0(x_0) = 1$ et $\varphi_0 \leq \mathbb{I}_{[\frac{k-1}{2^d}, \frac{k+1}{2^d}]}$. On a alors

$$\begin{aligned} \mu_\infty(\{x_0\}) &\leq \int_x \varphi_0 d\mu_\infty = \lim_{k \rightarrow \infty} \int_X \varphi_0 d\mu_{n_k} \leq \lim_{k \rightarrow \infty} \mu_{n_k} \left(\left[\frac{k-1}{2^d}, \frac{k+1}{2^d} \right] \right) \\ &\leq 2 \max\{p^d, q^d\} . \end{aligned}$$

Comme on peut choisir d aussi grand que l'on veut, on a $\mu_\infty(\{x_0\}) = 0$.

Comme $\mu_\infty(\{\frac{k}{2^d}\}) = \mu_\infty(\{\frac{k+1}{2^d}\}) = 0$, pour tout $\epsilon > 0$, on peut choisir des fonctions continues φ et ψ telles que $\varphi \leq \mathbb{I}_{I_{k,d}} \leq \psi$ et $\int_X (\psi - \varphi) d\mu_\infty \leq \epsilon$. On a alors

$$\begin{aligned} \mu_\infty(I_{k,d}) &\leq \int_X \psi d\mu_\infty \leq \int_X \varphi d\mu_\infty + \epsilon = \lim_{k \rightarrow \infty} \int_X \varphi d\mu_{n_k} + \epsilon \\ &\leq \lim_{k \rightarrow \infty} \mu_{n_k}(I_{k,d}) + \epsilon = p^{w_k} q^{d-w_k} + \epsilon . \end{aligned}$$

De même,

$$\begin{aligned} \mu_\infty(I_{k,d}) &\geq \int_X \varphi d\mu_\infty \geq \int_X \psi d\mu_\infty - \epsilon = \lim_{k \rightarrow \infty} \int_X \psi d\mu_{n_k} - \epsilon \\ &\geq \lim_{k \rightarrow \infty} \mu_{n_k}(I_{k,d}) - \epsilon = p^{w_k} q^{d-w_k} - \epsilon . \end{aligned}$$

On en déduit que $\mu_\infty(I_{k,d}) = p^{w_k} q^{d-w_k}$.

5) Par compacité faible de l'ensemble $\mathcal{M}(X)$ des mesures de probabilité boréliennes sur X , on peut trouver une suite extraite $(n_k)_{k \in \mathbb{N}}$ comme en 4). La mesure de probabilité $\nu = \mu_\infty$ convient. Pour montrer l'unicité, il suffit de remarquer que, comme toute fonction continue est limite uniforme de combinaisons linéaires d'applications de la forme $\mathbb{I}_{I_{k,d}}$, on connaît ν sur les fonctions continues.

6) Si μ_n ne convergeait pas vers ν , par compacité faible de $\mathcal{M}(X)$, il existerait une sous-suite qui convergerait vers une autre mesure de probabilité ν' . Ceci contredirait la question 4).

7) Lorsque $p = \frac{1}{2}$, la mesure ν est la mesure de Lebesgue dx .

B 1) En calculant comme en A 2), on a $T(\nu)(I_{k,d}) = p^{w_k} q^{d-w_k}$. Donc par A 5), on a $\nu = T(\nu)$.

2) Par construction, on a $f_*(T(\mu)) = \mu$. Donc $f_*(\nu) = \nu$.

3) On montre, comme en A 2), que

$$\mu_d(f^{-d}(A) \cap I_{k,d}) = \mu(A) p^{w_k} q^{d-w_k} .$$

Il suffit alors de prendre $\mu = \nu$.

4) Donc, si φ et ψ sont des fonctions de la forme $\mathbb{I}_{I_{k,d}}$, on a

$$\lim_{n \rightarrow \infty} \int_X (\varphi \circ f^n) \psi d\mu = \left(\int_X \varphi d\nu \right) \left(\int_X \psi d\nu \right) .$$

Comme ces fonctions engendrent un sous-espace dense de $\mathbb{L}^1(X, \nu)$, l'application f est mélangeante pour ν .

5) La mesure de probabilité ν est donc ergodique pour f . Le théorème ergodique de Birkhoff, appliqué à la fonction $\varphi = \mathbb{I}_{[\frac{1}{2}, 1[}$, prouve que, pour ν -presque tout x , la limite

$$\lim_{n \rightarrow \infty} \frac{1}{n} \text{Card } \{i \leq n / b_i = 1\} = \frac{1}{n} \lim_{n \rightarrow \infty} S_n \varphi(x)$$

existe et est égale à $\int_X \varphi d\nu = p$.

6) D'après A 5), le support de ν est égal à X . Il suffit d'appliquer la proposition 4.2.

7) En effet, les ensembles

$$E_p = \left\{ x \in X / \lim_{n \rightarrow \infty} \frac{1}{n} \text{Card } \{i \leq n / b_i = 1\} = p \right\}$$

sont disjoints et, d'après B 5), vérifient $\nu^p(E_p) = 1$.

7 Stabilité structurelle

Dans ce chapitre, nous montrons des phénomènes de “stabilité sous perturbation” pour des systèmes dynamiques qui vérifient des hypothèses “d’hyperbolicité”. Dit en terme de physiciens, il s’agit de montrer que “le chaos est stable par petite perturbation”. Nous montrerons trois tels résultats de stabilité par petite perturbation, un cas linéaire (la proposition 7.4), un cas local (le théorème de Grobman-Hartman 7.6) et un cas global (le théorème d’Anosov 7.11).

Les exemples principaux sont encore les automorphismes linéaires du tore, qui, lorsque que leur matrice entière associée n’a pas de valeur propre de module un, sont les archétypes des transformations “chaotiques”. Les démonstrations reposent sur le théorème du point fixe pour les applications contractantes.

7.1 Le théorème du point fixe

Le résultat suivant est bien classique.

Théorème 7.1 *Soient (X, d) un espace métrique complet, et $\varphi : X \rightarrow X$ une application δ -lipschitzienne avec $\delta < 1$. Alors φ admet un unique point fixe ℓ . En outre, pour tout x_0 dans X , on a*

$$d(x_0, \ell) \leq (1 - \delta)^{-1} d(x_0, \varphi(x_0)) .$$

Preuve. Pour montrer l’existence, considérons la suite de points définie par $x_n = \varphi^n(x_0)$. On a $d(x_n, x_{n+1}) \leq \delta^n d(x_0, \varphi(x_0))$. Donc la série

$$\sum_{n=0}^{\infty} d(x_n, x_{n+1}) \leq (1 - \delta)^{-1} d(x_0, \varphi(x_0))$$

converge. Donc la suite x_n est de Cauchy, et sa limite ℓ vérifie $\varphi(\ell) = \lim_{n \rightarrow \infty} \varphi(x_n) = \ell$ et $d(x_0, \ell) \leq (1 - \delta)^{-1} d(x_0, \varphi(x_0))$.

Pour montrer l’unicité, soient ℓ et ℓ' des points fixes. On a $d(\ell, \ell') = d(\varphi(\ell), \varphi(\ell')) \leq \delta d(\ell, \ell')$. Donc $\ell = \ell'$. $\square$

Voici une première application de ce théorème. On munit l’espace vectoriel $\mathbb{R}^N$ d’une norme $\|\cdot\|$. Pour toute application linéaire T de $\mathbb{R}^N$ dans $\mathbb{R}^N$, on note alors

$$\|T\| = \sup_{\|x\|=1} \|T(x)\|$$

la norme d’opérateur associée.

Lemme 7.2 *Soient T une bijection linéaire de $\mathbb{R}^N$ et $f : \mathbb{R}^N \rightarrow \mathbb{R}^N$ une application δ -lipschitzienne avec $\delta < \|T^{-1}\|^{-1}$. Alors l’application $F = T + f$ est un homéomorphisme bilipschitzien.*

Ceci signifie que F et F^{-1} sont lipschitziens. On note

$$\text{Lip}(f) = \sup_{x \neq x'} \frac{\|f(x) - f(x')\|}{\|x - x'\|}$$

la constante de Lipschitz de f .

Preuve. Comme $F = T \circ (id + T^{-1} \circ f)$ et que $\text{Lip}(T^{-1} \circ f) \leq \|T^{-1}\| \text{Lip}(f)$, il suffit de montrer ce lemme avec $T = id$ et $\delta < 1$, ce que nous supposons désormais.

Montrons que F est surjective. Soit y dans $\mathbb{R}^N$ et $\varphi : x \mapsto y - f(x)$. L'application φ est δ -lipschitzienne sur $\mathbb{R}^N$. Elle a donc un unique point fixe x . Celui-ci vérifie $F(x) = y$.

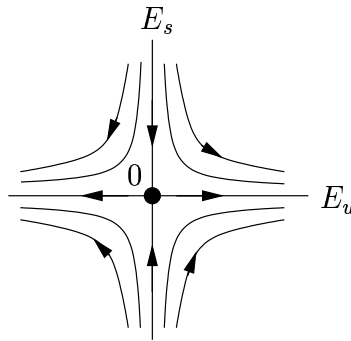
L'application F est clairement $(1 + \delta)$ -lipschitzienne. Pour tous x, x' dans $\mathbb{R}^N$, on a

$$\|F(x) - F(x')\| \geq \|x - x'\| - \|f(x) - f(x')\| \geq (1 - \delta)\|x - x'\|.$$

Donc F est injective, et F^{-1} est $(1 - \delta)^{-1}$ -lipschitzienne. $\square$

7.2 Automorphismes linéaires hyperboliques de $\mathbb{R}^N$

Une bijection linéaire T de l'espace vectoriel $E = \mathbb{R}^N$ est dite *hyperbolique* s'il existe une décomposition $E = E_s \oplus E_u$ en somme directe de sous-espaces T -invariants (i.e. $TE_s = E_s$ et $TE_u = E_u$) tels que, en notant $S = T|_{E_s}$ et $U = T|_{E_u}$, il existe $n \geq 1$ tel que $\|S^n\| < 1$ et $\|U^{-n}\| < 1$. (Les espaces E_s et E_u peuvent être réduits à $\{0\}$. Ce ne sera bien sûr pas le cas lorsque l'on suppose que T est de déterminant 1.)



Une telle transformation est l'archétype local d'un système dynamique hyperbolique (voir par exemple [KH]), et une familiarisation avec ce modèle linéaire est fondamental pour la théorie générale.

Remarques : (1) Cette définition ne dépend pas du choix de la norme sur E .

(2) La norme de E est dite *adaptée* à T si on a $\|S\| < 1$ et $\|U^{-1}\| < 1$ et si, pour tous x_s dans E_s et x_u dans E_u , on a $\|x_s + x_u\| = \max\{\|x_s\|, \|x_u\|\}$. On appelle alors *constante d'hyperbolicité* la constante

$$ch(T) = \max\{\|S\|, \|U^{-1}\|\} < 1.$$

Cette constante d'hyperbolicité dépend du choix de la norme adaptée.

Lemme 7.3 Soit T un automorphisme linéaire de $\mathbb{R}^N$. Alors T est hyperbolique si et seulement si T n'a pas de valeur propre de module 1. Dans ce cas, il existe une norme sur $\mathbb{R}^N$ adaptée à T .

Preuve. Si T est hyperbolique, les valeurs propres de S sont de module strictement inférieur à 1, celles de U^{-1} aussi.

Réciproquement, supposons que T n'a pas de valeur propre de module 1. On peut écrire $E = E_s \oplus E_u$ et $T = S \oplus U$ où les valeurs propres de S et U^{-1} sont de module strictement inférieur à 1. En mettant sous forme de Jordan la matrice de S , on vérifie que $\lim_{n \rightarrow \infty} \|S^n\| = 0$. Et donc, il existe un entier $n \geq 1$ tel que $\|S^n\| < 1$. De même avec U^{-1} .

Pour montrer la dernière affirmation, on part d'une norme quelconque $\|\cdot\|$ sur $\mathbb{R}^N$, on prend un entier n comme dans la définition et on pose

$$\|x_s + x_u\|' = \max \left\{ \sum_{k=0}^{n-1} \|S^k x_s\|, \sum_{k=0}^{n-1} \|U^{-k} x_u\| \right\}.$$

Cette application $\|\cdot\|'$ est une norme, adaptée à T . □

Les théorèmes de stabilité que nous montrerons se déduiront de la proposition suivante. Celle-ci implique en particulier qu'une petite perturbation (suffisamment régulière) d'un automorphisme hyperbolique de $\mathbb{R}^N$ lui est topologiquement conjuguée.

Proposition 7.4 *Soit T une bijection linéaire hyperbolique de $E = \mathbb{R}^N$. On fixe une norme $\|\cdot\|$ sur E adaptée à T . Soit $F = T + f$ où $f : E \rightarrow E$ est une application bornée et δ -lipschitzienne avec $\delta < \inf\{1 - ch(T), \|T^{-1}\|^{-1}\}$. Alors il existe un unique homéomorphisme $H = id + h$ de E tel que $h : E \rightarrow E$ est uniformément continue et bornée, avec*

$$H \circ T = F \circ H.$$

Remarques. (1) L'application F a donc un unique point fixe : le point $H(0)$.

(2) On notera $\|f\|_\infty = \sup_{x \in E} \|f(x)\|$. Il résultera de la démonstration que

$$\|h\|_\infty \leq (1 - \delta - ch(T))^{-1} \|f\|_\infty$$

et que h dépend continuellement de f .

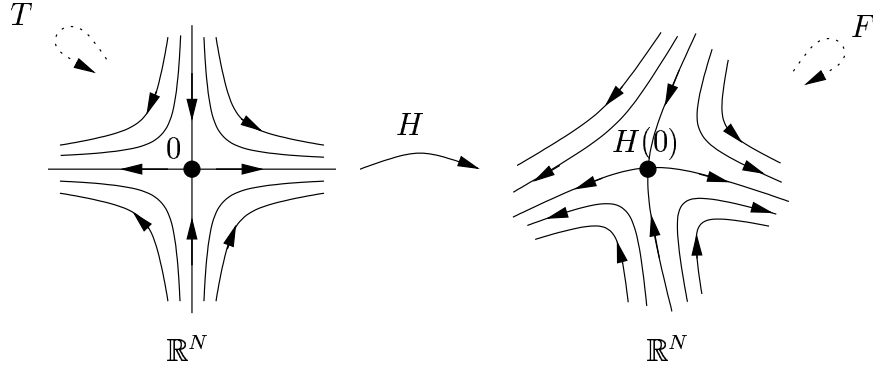
(3) Il n'est pas toujours possible de trouver une telle application H qui soit lipschitzienne. Par exemple, prendre $E = \mathbb{R}$, $T(x) = 2x$, $f(x) = -\epsilon \sin x$. En effet, dans cet exemple, on a $H(0) = 0$, et si on note $x_n = 2^{-n} \rightarrow 0$ et $y_n = H(x_n) \rightarrow 0$, alors on a

$$\lim_{n \rightarrow \infty} \frac{y_{n+1}}{y_n} = \lim_{n \rightarrow \infty} \frac{H(x_{n+1})}{F(H(x_{n+1}))} = \frac{1}{F'(0)} = \frac{1}{2 - \epsilon},$$

d'où $\lim_{n \rightarrow \infty} \frac{y_n}{x_n} = \infty$.

(4) C'est l'hypothèse " h bornée" qui force l'unicité de h . En effet, T commute à de nombreux éléments de $GL(E)$.

En particulier, cette proposition dit que la transformation linéaire hyperbolique T et sa petite perturbation F ont "la même dynamique" (voir section 2.1 ou le dessin ci-dessous).



Notons $\mathcal{H} = UCB(E, E)$ l'espace de Banach (vérifiez en exercice que c'en est bien un !) des applications uniformément continues et bornées de E dans E , muni de la norme $\|\cdot\|_\infty$. Voici une reformulation de la proposition précédente.

Proposition 7.5 *Soit T une bijection linéaire hyperbolique de $E = \mathbb{R}^N$. On fixe une norme $\|\cdot\|$ sur E adaptée à T . Soient $F = T + f$ et $G = T + g$ où $f, g : E \rightarrow E$ sont deux applications bornées et δ -lipschitzienne avec $\delta < \inf\{1 - ch(T), \|T^{-1}\|^{-1}\}$.*

(i) *Alors il existe un unique élément h de $\mathcal{H}$ tel que $H = id + h$ vérifie $F \circ H = H \circ G$. En outre, on a $\|h\|_\infty \leq (1 - \delta - ch(T))^{-1} \|f - g\|_\infty$.*

(ii) *L'application H est un homéomorphisme.*

Preuve. (i) Pour toute application $\gamma : E \rightarrow E$, on note $\gamma_s : E \rightarrow E_s$ et $\gamma_u : E \rightarrow E_u$ les applications telles que $\gamma = \gamma_s + \gamma_u$. En projetant l'égalité $F \circ H = H \circ G$ sur E_s et E_u , on obtient

$$\begin{cases} f_s \circ H + S \circ h_s &= g_s + h_s \circ G \\ f_u \circ H + U \circ h_u &= g_u + h_u \circ G \end{cases}.$$

Ce qui se traduit par l'égalité $h = \tilde{h}$, où

$$\begin{cases} \tilde{h}_s &= (f_s \circ H + S \circ h_s - g_s) \circ G^{-1} \\ \tilde{h}_u &= U^{-1} \circ (h_u \circ G + g_u - f_u \circ H) \end{cases}.$$

Notons que par le lemme 7.2 et le fait que $\delta < \|T^{-1}\|^{-1}$, l'application G est un homéomorphisme bilipschitzien de E .

L'application $\varphi : \mathcal{H} \rightarrow \mathcal{H}$ définie par $h \mapsto \tilde{h}$ est $(\delta + ch(T))$ -lipschitzienne. En effet, comme E est muni d'une norme adaptée à T , pour évaluer $\|\tilde{h} - \tilde{h}'\|_\infty$ avec h, h' dans $\mathcal{H}$, il suffit de calculer

$$\begin{aligned} \|\tilde{h}_s - \tilde{h}'_s\|_\infty &\leq \|f_s \circ H - f_s \circ H'\|_\infty + \|S\| \|h_s - h'_s\|_\infty \leq (\text{Lip}(f_s) + \|S\|) \|h - h'\|_\infty \leq \\ &(\delta + ch(T)) \|h - h'\|_\infty \end{aligned}$$

et

$$\begin{aligned} \|\tilde{h}_u - \tilde{h}'_u\|_\infty &\leq \|U^{-1}\| \|h_u \circ G - h'_u \circ G\|_\infty + \|U^{-1}\| \|f_u \circ H - f_u \circ H'\|_\infty \leq \\ &(\|U^{-1}\| + \text{Lip}(f_u)) \|h - h'\|_\infty \leq (\delta + ch(T)) \|h - h'\|_\infty, \end{aligned}$$

l'avant dernière inégalité étant vérifiée car $\|U^{-1}\| < 1$, puisque la norme est adaptée à T .

L'assertion (i) résulte alors du théorème du point fixe 7.1. La dernière affirmation s'obtient à l'aide de la dernière assertion du théorème 7.1, en prenant $x_0 = 0$. En effet,

$$d(0, \varphi(0)) = \|\tilde{0}\|_\infty \leq \max\{\|(f_s - g_s) \circ G^{-1}\|_\infty, \|U^{-1}\| \|g_u - f_u\|_\infty\} \leq \|g - f\|_\infty .$$

(ii) En échangeant les rôles de f et g , on trouve h' dans $\mathcal{H}$ tel que $H' = id + h'$ vérifie $H' \circ F = G \circ H'$. On a donc $(H \circ H') \circ F = F \circ (H \circ H')$. Comme $H \circ H' - id = h' + h \circ H'$ est dans $\mathcal{H}$, par unicité dans la partie (i), on a $H \circ H' = id$. De même, on a $H' \circ H = id$, et H est un homéomorphisme. $\square$

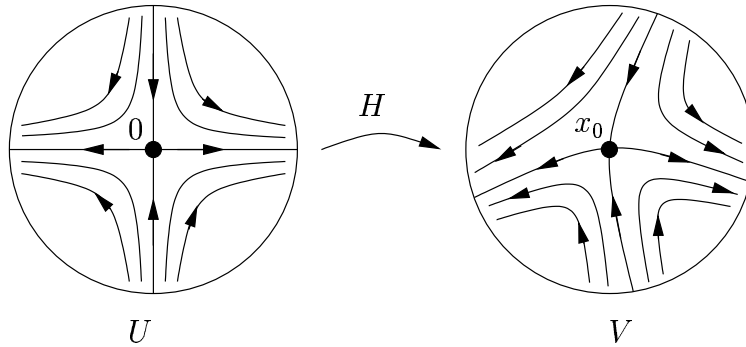
7.3 Le théorème de Grobman-Hartman

Le théorème suivant affirme que, dans certains cas, un difféomorphisme est localement topologiquement conjugué au voisinage d'un point fixe à sa différentielle. En particulier, la dynamique de ce difféomorphisme sera "la même" que celle de sa différentielle. Voir [KH] pour une autre preuve.

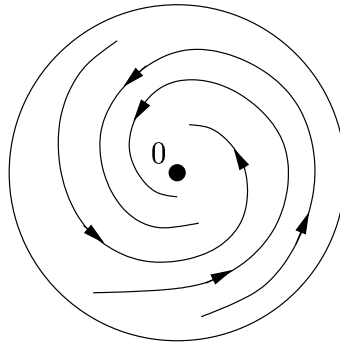
Théorème 7.6 *Soient Ω un ouvert de $\mathbb{R}^N$, $\varphi : \Omega \rightarrow \mathbb{R}^N$ un C^1 -difféomorphisme local et x_0 un point fixe hyperbolique de φ (i.e. x_0 est un point de Ω tel que $\varphi(x_0) = x_0$ et tel que la différentielle $T = D\varphi(x_0)$ est un automorphisme hyperbolique).*

Alors il existe des voisinages ouverts U de 0 dans $\mathbb{R}^N$ et V de x_0 dans Ω , et un homéomorphisme $H : U \rightarrow V$ tel que, pour tout x dans U avec $T(x)$ dans U , on a

$$H \circ T(x) = \varphi \circ H(x) .$$



Remarque. Si x_0 n'est pas hyperbolique, ceci n'est plus vrai. Par exemple, on peut considérer l'application $\varphi : \mathbb{C} \rightarrow \mathbb{C}$ définie par $z \mapsto e^{i\alpha}z(1 - |z|^2)$.



Preuve. Soit $\alpha : \mathbb{R}^N \rightarrow \mathbb{R}$ une fonction test de classe C^∞ telle que $\alpha(x) = 1$ si $\|x\| \leq 1$ et $\alpha(x) = 0$ si $\|x\| \geq 2$, et soit $\alpha_\epsilon(x) = \alpha(\frac{x}{\epsilon})$.

On peut supposer $x_0 = 0$. Écrivons alors $\varphi = T + f$ où $f : \Omega \rightarrow \mathbb{R}^N$ est de classe C^1 et vérifie $f(0) = 0$ et $Df(0) = 0$. Posons $F_\epsilon = T + f_\epsilon$ avec $f_\epsilon = \alpha_\epsilon f$. On prend ϵ assez petit. L'application $F_\epsilon : \mathbb{R}^N \rightarrow \mathbb{R}^N$ coïncide avec F sur $B(0, \epsilon)$ et avec T en dehors de $B(0, 2\epsilon)$. L'application f_ϵ est bornée et lipschitzienne, avec (par le théorème des accroissements finis)

$$\text{Lip}(f_\epsilon) = \sup_{x \in \mathbb{R}^N} \|Df_\epsilon(x)\|.$$

Or on a

$$Df_\epsilon(x) = \alpha(\frac{x}{\epsilon}) Df(x) + \frac{f(x)}{\epsilon} D\alpha(\frac{x}{\epsilon}).$$

Notons $M = \max \{ \|\alpha\|_\infty, \|x \mapsto \|D\alpha(x)\|_\infty \}$. On a alors

$$\text{Lip}(f_\epsilon) \leq M \left(\sup_{x \in B(0, 2\epsilon)} \|Df(x)\| + \sup_{x \in B(0, 2\epsilon)} \frac{\|f(x)\|}{\epsilon} \right) \leq 3M \sup_{x \in B(0, 2\epsilon)} \|Df(x)\|$$

(encore par le théorème des accroissements finis, car $f(0) = 0$ et $Df(0) = 0$). Donc $\text{Lip}(f_\epsilon)$ tend vers 0 avec ϵ , et on peut appliquer la proposition 7.4 à F_ϵ . Il existe donc un homéomorphisme H de $\mathbb{R}^N$ tel que $H \circ T = F_\epsilon \circ H$. On prend alors $U = B(0, \epsilon)$ et $V = H(B(0, \epsilon))$ et on remplace H par sa restriction à U . $\square$

7.4 Quelques lemmes de relèvement

Rassemblons ici quelques lemmes de relèvements dont nous aurons besoin. Notons $p : \mathbb{R}^N \rightarrow \mathbb{T}^N$ la projection canonique.

Lemme 7.7 (Relèvement des chemins) *Pour toute application continue $\varphi : [0, 1] \rightarrow \mathbb{T}^N$ et tout point x_0 tel que $p(x_0) = \varphi(0)$, il existe une unique application continue $\tilde{\varphi} : [0, 1] \rightarrow \mathbb{R}^N$ telle que $\tilde{\varphi}(0) = x_0$ et $p \circ \tilde{\varphi} = \varphi$.*

Lemme 7.8 (Relèvement des applications) *Pour toute application continue ψ de $\mathbb{R}^{N'}$ dans $\mathbb{T}^N$ et tout point x_0 de $\mathbb{R}^N$ tel que $p(x_0) = \psi(0)$, il existe une unique application continue $\tilde{\psi} : \mathbb{R}^{N'} \rightarrow \mathbb{R}^N$ telle que $\tilde{\psi}(0) = x_0$ et $p \circ \tilde{\psi} = \psi$.*

Corollaire 7.9 *Pour toute application continue $f : \mathbb{T}^N \rightarrow \mathbb{T}^N$ et tout point x_0 de $\mathbb{R}^N$ tel que $p(x_0) = f(x_0)$,*

1. *il existe une unique application continue $\tilde{f} : \mathbb{R}^N \rightarrow \mathbb{R}^N$, appelé relèvement de f telle que $\tilde{f}(0) = x_0$ et $p \circ \tilde{f} = f \circ p$,*
2. *il existe une matrice f_* à coefficients entiers telle que*

$$\forall x \in \mathbb{R}^N, \forall z \in \mathbb{Z}^N, \quad \tilde{f}(x + z) = \tilde{f}(x) + f_*(z),$$

3. *La matrice f_* est l'unique matrice M telle que $\tilde{f} - M$ est une application bornée sur $\mathbb{R}^N$.*

Remarque. Ces résultats pour $X = \mathbb{T}^N$ se généralisent à tout espace topologique X connexe par arcs et localement connexe par arcs, si on remplace $\mathbb{R}^N$ par un revêtement universel de X (s'il existe) et $\mathbb{Z}^N$ par son groupe fondamental. Voir par exemple les notes de cours [Pau], ainsi que [God, Hat, Spa].

Preuve du lemme 7.7 Pour montrer l'unicité, soient $\tilde{\varphi}, \tilde{\varphi}'$ deux relèvements de φ tels que $\tilde{\varphi}(0) = \tilde{\varphi}'(0)$. Comme $\tilde{\varphi} - \tilde{\varphi}'$ est continue et prend ses valeurs dans $\mathbb{Z}^N$, on a $\tilde{\varphi} = \tilde{\varphi}'$.

Pour montrer l'existence, on peut supposer $N = 1$ en prenant les coordonnées. Par compacité, on découpe l'intervalle $[0, 1]$ en un nombre fini d'intervalles $[t_i, t_{i+1}]$ tels que $\varphi([t_i, t_{i+1}]) \neq \mathbb{T}^1$. On relève alors successivement φ sur chaque intervalle. $\square$

Preuve du lemme 7.8 Pour x dans $\mathbb{R}^{N'}$, on note φ_x l'application de $[0, 1]$ dans $\mathbb{T}^N$ définie par $t \mapsto \psi(tx)$.

L'unicité découle du fait que l'on a forcément $\tilde{\psi}(x) = \widetilde{\varphi_x}(1)$.

Pour montrer l'existence, on pose $\tilde{\psi}(x) = \widetilde{\varphi_x}(1)$. Le procédé de construction dans la démonstration ci-dessus assure que $\tilde{\psi}$ est continue. $\square$

Preuve du corollaire 7.9 (1) On applique le lemme 7.8 à $\psi = f \circ p$.

(2) Soient z, z' dans $\mathbb{Z}^N$. L'application $x \mapsto \tilde{f}(x+z) - \tilde{f}(x)$ est une application continue à valeurs dans $\mathbb{Z}^N$. Elle est donc égale à une constante $f_*(z)$. On a

$$f_*(z) + f_*(z') = \left(\tilde{f}(x+z+z') - \tilde{f}(x+z') \right) + \left(\tilde{f}(x+z') - \tilde{f}(x) \right) = f_*(z+z').$$

Donc f_* est un morphisme de groupes de $\mathbb{Z}^N$ dans lui-même, et est donc donnée par une matrice à coefficients entiers.

(3) Remarquer que $\tilde{f} - f_*$ est continue et $\mathbb{Z}^N$ -périodique, donc bornée, et que toute application linéaire bornée sur $\mathbb{R}^N$ est nulle. $\square$

Munissons le tore $\mathbb{T}^N$ de la distance

$$d(\theta, \theta') = \inf_{p(x)=\theta, p(x')=\theta'} \|x - x'\|$$

où $\|\cdot\|$ est la norme euclidienne sur $\mathbb{R}^N$.

Lemme 7.10 Soient f, g deux applications continues de $\mathbb{T}^N$ dans $\mathbb{T}^N$ telles que, pour tout θ dans $\mathbb{T}^N$, on a $d(f(\theta), g(\theta)) < \frac{1}{2}$. Alors on a $f_* = g_*$.

Preuve. Notons $\tilde{f}, \tilde{g} : \mathbb{R}^N \rightarrow \mathbb{R}^N$ des relèvements de f et g . L'application continue $\tilde{f} - \tilde{g}$ prend ses valeurs dans $\bigcup_{z \in \mathbb{Z}^N} \overset{\circ}{B}(z, \frac{1}{2})$. Comme ces boules sont disjointes, l'application $\tilde{f} - \tilde{g}$ est bornée, donc par le corollaire 7.9, $f_* - g_*$ est bornée, et $f_* = g_*$. $\square$

On note $C^1(\mathbb{T}^N, \mathbb{T}^N)$ l'ensemble des applications $f : \mathbb{T}^N \rightarrow \mathbb{T}^N$ de classe C^1 , i.e. telles que $\tilde{f} : \mathbb{R}^N \rightarrow \mathbb{R}^N$ est de classe C^1 . Pour tout point $\theta = p(x)$ dans $\mathbb{T}^N$, on note alors $Df(\theta) = D\tilde{f}(x)$. Cette application linéaire ne dépend ni du choix de $\tilde{f}$ ni du choix de x . C'est, par définition la *différentielle* de f en θ . Voir un cours de géométrie différentielle (par exemple [Laf, Spi]) pour des compléments.

On munit l'ensemble $C^1(\mathbb{T}^N, \mathbb{T}^N)$ de la distance d_1 donnée par

$$d_1(f, g) = \sup_{\theta \in \mathbb{T}^N} (d(f(\theta), g(\theta)) + \|Df(\theta) - Dg(\theta)\|).$$

7.5 Stabilité structurelle des automorphismes hyperboliques du tore.

Soit M une matrice inversible à coefficients entiers. Nous avons vu dans les chapitres précédents que la dynamique de la transformation $f_M : \mathbb{T}^N \rightarrow \mathbb{T}^N$ peut être très chaotique : mélange, densité des orbites périodiques, ...

Le théorème suivant affirme que, lorsque M est hyperbolique, cette dynamique chaotique n'est pas modifiée par une petite perturbation de f_M .

Théorème 7.11 (Théorème d'Anosov) *Soit M une matrice inversible à coefficients entiers et hyperbolique. Alors il existe un voisinage U de f_M dans $C^1(\mathbb{T}^N, \mathbb{T}^N)$ tel que toute application f dans U est topologiquement conjuguée à f_M (i.e. il existe un homéomorphisme h de $\mathbb{T}^N$ tel que $h \circ f = f_M \circ h$).*

Remarque. On peut choisir l'homéomorphisme h vérifiant en plus l'égalité $h_* = id$. En outre, il existe alors un seul homéomorphisme h vérifiant ces deux conditions. Cela résultera de la démonstration.

Preuve. D'après le lemme 7.10, si on choisit U suffisamment petit, tous les éléments f de U vérifient $f_* = (f_M)_* = M$.

D'après le corollaire 7.9, les relèvements $\tilde{f}$ de f s'écrivent $\tilde{f} = M + \varphi$ avec φ une application bornée sur $\mathbb{R}^N$. Quitte à réduire U , on peut supposer que la constante de Lipschitz

$$\text{Lip}(\varphi) = \sup_{\theta \in \mathbb{T}^N} \|Df(\theta) - Df_M(\theta)\|$$

est aussi petite que l'on veut.

La proposition 7.4 nous donne alors un homéomorphisme H de $\mathbb{R}^N$ tel que $H \circ \tilde{f} = M \circ H$ et tel que l'application $H - id$ est bornée.

Lemme 7.12 *Soit x dans $\mathbb{R}^N$. Le point $H(x)$ est l'unique point y de $\mathbb{R}^N$ tel que*

$$\sup_{n \in \mathbb{Z}} \|M^n y - \tilde{f}^n(x)\| < \infty .$$

Preuve. Le point $H(x)$ vérifie bien cette condition, car on a $M^n H(x) - \tilde{f}^n(x) = H \tilde{f}^n(x) - \tilde{f}^n(x)$ et l'application $H - id$ est bornée.

Réciproquement, si un point $y = H(x) + v$ vérifie cette condition, alors la suite $(M^n v)_{n \in \mathbb{Z}}$ est bornée. Comme M est hyperbolique, on a forcément $v = 0$. $\square$

Pour terminer la preuve du théorème d'Anosov 7.11, montrons qu'il existe une application continue $h : \mathbb{T}^N \rightarrow \mathbb{T}^N$ telle que $p \circ H = h \circ p$. Pour cela, il suffit de voir que

$$\forall z \in \mathbb{Z}^N, \forall x \in \mathbb{R}^N, \quad H(x + z) = H(x) + z .$$

Mais cette égalité résulte du lemme ci-dessus, car la suite

$$M^n(H(x) + z) - \tilde{f}^n(x + z) = M^n H(x) - \tilde{f}^n(x)$$

est bornée.

L'application h vérifie $h \circ f = f_M \circ h$. Il reste à voir que h est un homéomorphisme. Or l'inverse $H' = H^{-1}$ vérifie aussi les égalités

$$\forall z \in \mathbb{Z}^N, \forall x \in \mathbb{R}^N, \quad H'(x + z) = H'(x) + z .$$

Il existe donc aussi une application continue h' telle que $p \circ H' = h' \circ p$. Par construction, h' est l'inverse de h . $\square$

7.6 Exercices

Exercice E.7.1 Soit Ω un ouvert de $\mathbb{R}^N$, $\varphi : \Omega \rightarrow \mathbb{R}^N$ un difféomorphisme local de classe C^1 , x_0 un point fixe de φ , $T = D\varphi(x_0)$ et $Sp(T)$ l'ensemble des valeurs propres de T .

1. Montrer que si $Sp(T)$ est contenu dans $\{z \in \mathbb{C} / |z| < 1\}$, alors x_0 est *asymptotiquement stable*, i.e. $\forall \epsilon > 0, \exists \delta > 0, \forall x \in B(x_0, \delta), \forall n \in \mathbb{N} - \{0\}$,

$$\|\varphi^n(x) - x_0\| \leq \epsilon \quad \text{et} \quad \lim_{n \rightarrow \infty} \varphi^n(x) = x_0$$

2. Donner un exemple de point fixe non asymptotiquement stable.

Exercice E.7.2 Donner un exemple de matrice M à coefficients entiers de déterminant 1 (i.e. $M \in SL_N(\mathbb{Z})$) qui n'admet pas de valeur propre racine de l'unité (de sorte que l'application $f_M : \mathbb{T}^N \rightarrow \mathbb{T}^N$ est mélangeante pour $d\theta$), mais qui n'est pas hyperbolique.

Exercice E.7.3 (Lemme de pistage) Soit T une bijection linéaire hyperbolique de $\mathbb{R}^N$, $\|\cdot\|$ une norme sur $\mathbb{R}^N$ adaptée à T , et $\epsilon > 0$. Montrer que, pour toute suite $(y_n)_{n \in \mathbb{Z}}$ de $\mathbb{R}^N$ telle que

$$\sup_{n \in \mathbb{Z}} \|y_{n+1} - T(y_n)\| \leq \epsilon,$$

il existe un unique point y de $\mathbb{R}^N$ tel que

$$\sup_{n \in \mathbb{Z}} \|y_n - T^n(y)\| \leq \frac{\epsilon}{1 - ch(T)}.$$

(Autrement dit, toute ϵ -pseudo-orbite de T est ϵ' -pistée par une (vraie) orbite avec $\epsilon' = \frac{\epsilon}{1 - ch(T)}$.)

Exercice E.7.4 (1) Soient $(X, \mathcal{B}, \mu)$ un espace de probabilité, et $f : X \rightarrow X$ une application mesurable qui préserve μ . On suppose f bijective, et son inverse mesurable. Soit $\varphi : X \rightarrow \mathbb{C}$ une application μ -sommable. On note, pour tout x dans X ,

$$S_n \varphi(x) = \frac{1}{n} \sum_{k=0}^{n-1} \varphi(f^k(x)) \quad \text{et} \quad S_{n,-} \varphi(x) = \frac{1}{n} \sum_{k=0}^{n-1} \varphi(f^{-k}(x)).$$

Montrer que, pour μ -presque tout x dans X , les limites

$$\tilde{\varphi}(x) = \lim_{n \rightarrow \infty} S_n \varphi(x) \quad \text{et} \quad \tilde{\varphi}_-(x) = \lim_{n \rightarrow \infty} S_{n,-} \varphi(x).$$

existent et sont égales.

(2) On prend désormais $X = \mathbb{T}^N$, $\mathcal{B}$ la σ -algèbre des boréliens et $\mu = d\theta$ la mesure de Haar. On note $E = \mathbb{R}^N$, $p : E \rightarrow X$ la projection canonique, $\|\cdot\|$ une norme sur E et d la distance quotient sur X , donnée par $d(\theta, \theta') = \inf\{\|v - v'\| / p(v) = \theta, p(v') = \theta'\}$. Soit $\psi : X \rightarrow \mathbb{C}$ une application lipschitzienne de constante de Lipschitz $\text{Lip}(\psi)$.

Montrer que, pour tout θ dans X et v dans E , on a

$$|\psi(\theta + p(v)) - \psi(\theta)| \leq \text{Lip}(\psi) \|v\|.$$

(3) Soit M une matrice $N \times N$ à coefficients entiers telle que $\det M = \pm 1$ et f_M la transformation de X telle que $p \circ M = f_M \circ p$.

a) Montrer que f_M est bijective. On prendra désormais $f = f_M$.

b) On note, pour θ dans X , par $\tau_\theta : X \rightarrow X$ la translation par θ . Montrer que $f \circ \tau_\theta = \tau_{f(\theta)} \circ f$.

(4) On suppose désormais que M est hyperbolique. On choisit pour norme $\|\cdot\|$ une norme adaptée à M , et on note $E = E_s \oplus E_u$ la décomposition de E en sous-espaces M -invariants telle que

$$ch(M) = \max\{\|M|_{E_s}\|, \|M^{-1}|_{E_u}\|\} < 1.$$

a) Montrer que, pour tout θ dans X et v dans E_s , on a

$$|S_n \psi(\theta + p(v)) - S_n \psi(\theta)| \leq \frac{\text{Lip}(\psi)}{n(1 - ch(M))} \|v\|.$$

b) Montrer que, pour tout v dans E_s , on a $\tilde{\psi} \circ \tau p(v) = \tilde{\psi}$ μ -presque partout.

c) Montrer que, pour tout v dans E_u , on a $\tilde{\psi} \circ \tau p(v) = \tilde{\psi}$ μ -presque partout.

d) Montrer que, pour tout v dans E , on a $\tilde{\psi} \circ \tau p(v) = \tilde{\psi}$ μ -presque partout.

e) Montrer que $\tilde{\psi}$ est constante μ -presque partout.

f) Montrer que $\tilde{\varphi}$ est constante μ -presque partout.

g) En déduire que f est ergodique pour μ .

7.7 Indications pour la résolution des exercices

Exercice E.7.2 On cherchera avec $N = 4$ et on considèrera le polynôme $X^4 + 2X^3 + X^2 + 2X + 1$.

Exercice E.7.3 Indication. On se ramènera au cas $\|T\| < 1$. On prendra alors $y = \lim_{n \rightarrow +\infty} T^n(y_{-n})$.

Preuve. En écrivant $y_n = (y_n^s, y_n^u)$ dans la décomposition en sous-espace stable et instable $E = E_s \oplus E_u$, qui est T -invariante, et puisque la norme est adaptée, on a $\sup_{n \in \mathbb{Z}} \|y_n^s - T(y_n^s)\| \leq \epsilon$ et $\sup_{n \in \mathbb{Z}} \|y_n^u - T(y_n^u)\| \leq \epsilon$. En se restreignant à E_s et E_u et en remplaçant f par f^{-1} sur E_u , on peut donc supposer que $E = E_s$ et $\|T\| < 1$.

Comme $\|T^{n+1}(y_{-(n+1)}) - T^n(y_{-n})\| \leq \|T\|^{n+1} \|T(y_{-n-1}) - y_{-n}\| \leq \epsilon \|T\|^n$, la suite des $T^n(y_{-n})$ est de Cauchy, donc converge vers un élément y de E .

Pour k dans $\mathbb{Z}$, pour $n \geq 1 - k$, on a $y_k - T^k(y) = \sum_{i=0}^{n+k-1} (T^i(y_{k-i}) - T^{i+1}(y_{k-i-1})) + T^{k+n}y_{-n} - T^k y$, donc

$$\|y_k - T^k(y)\| \leq \sum_{i=0}^{n+k-1} \|T\|^i \|y_{k-i} - T(y_{k-i-1})\| + \|T^k\| \|T^n(y_{-n}) - y\|,$$

par conséquent $\|y_k - T^k(y)\| \leq \frac{\epsilon}{1 - \|T\|}$.

Exercice E.7.4 (1) On posera $A = \{x \in X / \tilde{\varphi}(x) > \tilde{\varphi}_-(x)\}$ et on calculera $\int_A (\tilde{\varphi}(x) - \tilde{\varphi}_-(x)) d\mu$.

8 Représentations unitaires

Nous avons étudié, dans les chapitres précédents, divers exemples de systèmes dynamiques sur le tore $\mathbb{T}^N$, qui est le quotient du groupe $G = \mathbb{R}^N$ par le sous-groupe discret $\Gamma = \mathbb{Z}^N$, en particulier pour l'action par translations de $\mathbb{R}^n$ sur $\mathbb{T}^N$. Dans ce chapitre, nous étendons cette étude à des groupes G non commutatifs comme $G = SL_N(\mathbb{R})$.

L'exemple le plus simple de tel système dynamique sur $\Gamma \backslash G$ s'interprète géométriquement comme le "flot géodésique" sur un quotient du "plan hyperbolique", voir paragraphe 8.6.

Le quotient $\Gamma \backslash G$ n'est plus un groupe, mais un espace homogène sous l'action de G . Les séries de Fourier, qui jouaient un rôle si important pour le tore $\mathbb{T}^N$, seront remplacées ici par les représentations unitaires de G .

En effet, si $G = \mathbb{R}^N$, $\Gamma = \mathbb{Z}^N$ alors les fonctions trigonométriques e_m forment une base hilbertienne de l'espace de Hilbert $\mathcal{H} = \mathbb{L}^2(\mathbb{Z}^N \backslash \mathbb{R}^N, d\theta)$, sur lequel le groupe $\mathbb{R}^N$ agit par $(t, \varphi) \mapsto (\pi(t)\varphi : x \bmod \mathbb{T}^N \mapsto \varphi(x + t \bmod \mathbb{T}^N))$. Cette action est une représentation unitaire de $G = \mathbb{R}^N$ sur $\mathcal{H}$ (voir paragraphe 8.2 pour leur définition). La preuve de l'ergodicité des translations par un élément t de $\mathbb{R}^n$ totalement irrationnel (voir paragraphe 4.1) passait par l'étude des coefficients matriciels (voir paragraphe 8.2 pour leur définition) $\langle \pi(nt)e_m, e_{m'} \rangle$, et en particulier leur annulation pour n assez grand. C'est ce résultat d'annulation qui formera la trame de tout ce chapitre.

Nous renvoyons par exemple aux références [Car, BHV] pour des généralités sur les représentations unitaires, et à [BM, HT] pour le contenu du paragraphe 8.4.

Dans tout ce chapitre, on fixe un entier $N \geq 2$.

8.1 Sur $SL_N(\mathbb{R})$

Soit G le groupe topologique localement compact

$$SL_N(\mathbb{R}) = \{g \in \mathcal{M}(N, \mathbb{R}) \mid \det g = 1\}.$$

Pour g dans G , on note λ_g et ρ_g les homéomorphismes de G définies par $\lambda_g : x \mapsto gx$ (*translation à gauche*) et $\rho_g : x \mapsto xg^{-1}$ (*translation à droite*).

Lemme 8.1 1. Il existe une mesure borélienne ν sur G , finie sur les compacts, telle que, pour tout g dans G , on a $(\lambda_g)_*\nu = \nu$.

2. Pour tout g dans G , on a aussi $(\rho_g)_*\nu = \nu$.

Remarque. L'assertion (1) est vraie pour tout groupe topologique localement compact G . La mesure ν est unique à un scalaire multiplicatif près, et est appelée *mesure de Haar* (à gauche). Lorsque l'assertion (2) est vérifiée, on dit que G est *unimodulaire*.

Preuve. Il est facile de construire explicitement ν pour $G = SL_N(\mathbb{R})$. On note λ la mesure de Lebesgue sur $\mathcal{M}(N, \mathbb{R}) \simeq \mathbb{R}^{N^2}$, et on pose, pour tout borélien A de G ,

$$\nu(A) = \lambda \left(\left\{ x \in \mathcal{M}(N, \mathbb{R}) \mid 1 \leq \det x \leq 2 \text{ et } \frac{1}{(\det x)^{\frac{1}{N}}} x \in A \right\} \right).$$

On a bien $(\lambda_g)_*\nu = (\rho_g)_*\nu = \nu$, car les applications linéaires de $\mathcal{M}(N, \mathbb{R})$ définies par $x \mapsto gx$ et $x \mapsto xg^{-1}$ sont de déterminant 1, pour tout g dans G . $\square$

Il existe sur G (au moins) une distance d_G invariante à gauche, i.e. telle que pour tous g, x, y dans G , on a

$$d_G(gx, gy) = d_G(x, y) .$$

On peut, par exemple, en notant e la matrice identité et $\| \cdot \|$ une *norme matricielle* (i.e. une norme sur $\mathcal{M}_N(\mathbb{R})$ vérifiant $\|xy\| \leq \|x\|\|y\|$, comme la norme d'opérateur en tant qu'application linéaire de l'espace euclidien $\mathbb{R}^N$, voir paragraphe 7.1), prendre la distance

$$d_G(x, y) = \log(1 + \|x^{-1}y - e\|) + \log(1 + \|y^{-1}x - e\|) .$$

(L'inégalité triangulaire découle par exemple des inégalités

$$\begin{aligned} 1 + \|x^{-1}z - e\| &= 1 + \|x^{-1}y - e\| + \|y^{-1}z - e\| + (\|x^{-1}z - e\| - \|x^{-1}y - e\| - \|y^{-1}z - e\|) \leq \\ &1 + \|x^{-1}y - e\| + \|y^{-1}z - e\| + \|(x^{-1}z - e) - (x^{-1}y - e) - (y^{-1}z - e)\| = \\ 1 + \|x^{-1}y - e\| + \|y^{-1}z - e\| + \|(x^{-1}y - e)(y^{-1}z - e)\| &\leq (1 + \|x^{-1}y - e\|)(1 + \|y^{-1}z - e\|) . \end{aligned}$$

Soit Γ un sous-groupe discret de G . Il existe donc $\epsilon_0 > 0$ tel que $B_G(e, 2\epsilon_0) \cap \Gamma = \{e\}$. Par définition, le quotient $X = \Gamma \backslash G$ est l'ensemble des classes à gauches $x = \Gamma g$. On le munit de la distance (qui induit la topologie quotient)

$$d(\Gamma g, \Gamma g') = \inf_{\gamma \in \Gamma} d_G(g, \gamma g') ,$$

et on note p la projection canonique $p : G \rightarrow X$ définie par $g \mapsto \Gamma g$. On a, comme pour le tore $\mathbb{T}^N$, le résultat suivant, dont la preuve est laissée au lecteur.

Lemme 8.2 *Pour tout g dans G , l'application p induit un homéomorphisme de $B_G(g, \epsilon_0)$ sur $B(p(g), \epsilon_0)$. De plus, $p^{-1}(B(p(g), \epsilon_0))$ est la réunion disjointe des boules $B_G(\gamma g, \epsilon_0)$ pour γ dans Γ .* $\square$

Pour g dans G et x dans X , on note encore ρ l'action de G sur X par translation à droite sur les classes à gauches i.e. $\rho(g) : x \mapsto xg^{-1}$.

Lemme 8.3 *Il existe une mesure borélienne μ sur X telle que*

$$\forall g \in G, \quad \rho(g)_* \mu = \mu .$$

Remarque. Ce lemme est encore vrai pour tous les groupes topologiques localement compacts G et leurs sous-groupes discrets Γ . La mesure μ est unique à scalaire multiplicatif près.

Preuve. On définit d'abord μ sur les boules $B(p(g), \frac{\epsilon_0}{2})$: pour tout borélien A contenu dans $B(p(g), \frac{\epsilon_0}{2})$, on écrit $A = p(\tilde{A})$ avec $\tilde{A}$ un borélien contenu dans $B(g, \frac{\epsilon_0}{2})$ et on pose $\mu(A) = \nu(\tilde{A})$. Cela ne dépend pas des choix, car ν est invariante à gauche.

On définit alors μ sur X : pour tout borélien A , on écrit $A = \bigsqcup_{i \in \mathbb{N}} A_i$ avec les A_i des boréliens deux à deux disjoints, de diamètre au plus $\frac{\epsilon_0}{2}$, et on pose $\mu(A) = \sum_{i \in \mathbb{N}} \mu(A_i)$. Cela ne dépend pas non plus des choix. Par construction, et comme G est unimodulaire, pour tout g dans G , on a $\mu(Ag) = \mu(A)$. $\square$

Un sous-groupe discret Γ de G est appelé un *réseau* si $\mu(\Gamma \backslash G) < \infty$. Il est dit *cocompact* (ou *uniforme*) si $\Gamma \backslash G$ est compact.

Exemples : (1) Le sous-groupe $\mathrm{SL}_N(\mathbb{Z})$ est un réseau non cocompact de $\mathrm{SL}_N(\mathbb{R})$.

(2) Le groupe topologique $\mathrm{SL}_N(\mathbb{R})$ contient aussi des réseaux cocompacts.

Nous ne montrerons pas ces assertions, voir par exemple [Bor, BHC]. Ces exemples sont construits de façon arithmétique et sont le point de départ d’une interaction riche “systèmes dynamiques $\longleftrightarrow$ arithmétique”, voir par exemple [Zim, Mar].

8.2 Représentations unitaires

Lorsque Γ est un réseau de G , l’action des éléments g de G sur $X = \Gamma \backslash G$ est un système dynamique, qui préserve la mesure finie μ . Pour étudier les propriétés ergodiques de ce système dynamique, nous aurons besoin de la définition suivante. Si $\mathcal{H}$ est un espace de Hilbert, on note $U(\mathcal{H})$ le *groupe unitaire* de $\mathcal{H}$, i.e. le groupe des bijections linéaires continues u dans $\mathcal{L}(\mathcal{H})$ telles que $uu^* = u^*u = 1$. En particulier, l’application $G \times \mathcal{H} \rightarrow \mathcal{H}$ définie par $(g, v) \mapsto \pi(g)v$ est une action du groupe G sur l’ensemble $\mathcal{H}$.

Une *représentation unitaire* de G dans un espace de Hilbert $\mathcal{H}$ est un morphisme de groupes $\pi : G \rightarrow U(\mathcal{H})$ tel que, pour tout v dans $\mathcal{H}$, l’application $G \rightarrow \mathcal{H}$ définie par $g \mapsto \pi(g)v$ est continue.

Exemple : Pour φ dans $\mathbb{L}^2(X, \mu)$, on note $\pi(g)\varphi$ l’élément de $\mathbb{L}^2(X, \mu)$ donné par

$$(\pi(g)\varphi)(x) = \varphi(xg) .$$

Comme μ est invariante par l’action par translation à droite de G , l’application linéaire $\pi(g)$ est dans $U(\mathbb{L}^2(X, \mu))$. Le morphisme de groupes $\pi : G \rightarrow U(\mathbb{L}^2(X, \mu))$ ainsi défini est une représentation unitaire dans l’espace de Hilbert $\mathbb{L}^2(X, \mu)$.

Preuve. Le seul point non trivial est la continuité de l’application $g \mapsto \pi(g)\varphi$. C’est vrai si φ est dans $C_c(X)$, l’ensemble des fonctions continues à support compact sur X . On conclut par densité de $C_c(X)$ dans $\mathbb{L}^2(X)$: soient g_n une suite dans G qui converge vers g , φ dans $\mathbb{L}^2(X)$ et $\epsilon > 0$; il existe φ_0 dans $C_c(X)$ tel que $\|\varphi - \varphi_0\|_{\mathbb{L}^2} \leq \epsilon$; alors

$$\|\pi(g_n)\varphi - \pi(g)\varphi\|_{\mathbb{L}^2} \leq 2\|\varphi - \varphi_0\|_{\mathbb{L}^2} + \|\pi(g_n)\varphi_0 - \pi(g)\varphi_0\|_{\mathbb{L}^2} \leq 3\epsilon$$

pour n assez grand. □

On note $\mathcal{H}^G = \{v \in \mathcal{H} / \forall g \in G, \pi(g)v = v\}$.

Exemple : Soit

$$\mathbb{L}_0^2(X, \mu) = \{\varphi \in \mathbb{L}^2(X, \mu) / \int_X \varphi d\mu = 0\} .$$

C’est un sous-espace de Hilbert de $\mathbb{L}^2(X, \mu)$, qui est G -invariant. On a $\mathbb{L}_0^2(X, \mu)^G = \{0\}$ (rappelons que si un groupe G agit sur un ensemble E , on note E^G l’ensemble des points de E fixés par tous les éléments de G).

Preuve. Soit $\varphi \in \mathbb{L}_0^2(X, \mu)$. Pour tout g dans G , on a $\varphi(xg) = \varphi(x)$ pour μ -presque tout x . Donc, par Fubini, pour μ -presque tout x , on a $\varphi(xg) = \varphi(x)$ pour ν -presque tout g . On fixe un tel point x . Lorsque g décrit le complémentaire d’un ensemble ν -négligeable, les

points xg décrivent le complémentaire d'un ensemble μ -négligeable. Donc φ est constante μ -presque tout. Comme φ est d'intégrale nulle, elle est nulle μ -presque tout. $\square$

Remarque. Cette preuve montre que $\mathbb{L}^2(X, \mu)^G$ est réduit aux fonctions μ -presque partout constantes.

Pour v, v' dans $\mathcal{H}$, on note $c_{v,v'}$ l'application de G dans $\mathbb{C}$ définie par $c_{v,v'}(g) = \langle (\pi(g)v, v') \rangle$, appelée un *coefficient matriciel* de π . C'est une fonction continue sur G . Le coefficient matriciel $c_{v,v'}$ a comme propriété remarquable d'être invariant à droite par le fixateur de v et invariant à gauche par le fixateur de v' :

$$\forall g, h, h' \in G, \quad \text{si } \pi(h)v = v \text{ et } \pi(h')v' = v', \quad \text{alors } c_{v,v'}(h'gh) = c_{v,v'}(g)$$

8.3 Ergodicité des quotients de $SL_2(\mathbb{R})$ par un réseau

Pour $t > 0$ et $s \in \mathbb{R}$, on note $a_t = \begin{pmatrix} t & 0 \\ 0 & t^{-1} \end{pmatrix}$, $u_s = \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix}$, et $u_s^- = \begin{pmatrix} 1 & 0 \\ s & 1 \end{pmatrix}$.

Théorème 8.4 *Soit Γ un réseau de $G = SL_2(\mathbb{R})$ et $t \neq 1$ (respectivement $s \neq 0$). Alors l'action de a_t (respectivement de u_s ou de u_s^-) est ergodique pour μ .*

D'après la proposition 4.1, ce théorème est une conséquence du résultat suivant, appliqué à $\mathcal{H} = \mathbb{L}^2(X, \mu)$, en utilisant la dernière remarque du paragraphe 8.2.

Proposition 8.5 *Soit $(\mathcal{H}, \pi)$ une représentation unitaire de G et $t \neq 1$ (respectivement $s \neq 0$). Alors tout vecteur v de $\mathcal{H}$ qui est a_t -invariant (respectivement u_s -invariant ou u_s^- -invariant) est G -invariant.*

Cette proposition repose sur les lemmes suivants.

Lemme 8.6 (Phénomène de Mautner) *Pour tout v dans $\mathcal{H}$, on note $G_v = \{g \in G / \pi(g)v = v\}$ le fixateur de v .*

- a) *Le fixateur G_v est un sous-groupe fermé de G .*
- b) *On a l'équivalence $g \in G_v \Leftrightarrow c_{v,v}(g) = \|v\|^2$.*
- c) *Soit g dans G tel qu'il existe des suites g_n dans G et s_n, s'_n dans G_v avec*

$$\lim_{n \rightarrow \infty} g_n = g \quad \text{et} \quad \lim_{n \rightarrow \infty} s_n g_n s'_n = e.$$

Alors g est dans G_v .

Preuve. L'assertion a) est claire, et b) résulte de l'égalité

$$\|\pi(g)v - v\|^2 = 2(\|v\|^2 - \operatorname{Re}(c_{v,v}(g))).$$

Pour montrer l'assertion c), remarquons que $c_{v,v}(g_n) = c_{v,v}(s_n g_n s'_n)$ car $\pi(s'_n)v = v$ et $\pi(s_n^{-1})v = v$. Par continuité, on en déduit que $c_{v,v}(g) = c_{v,v}(e) = \|v\|^2$. Il résulte de l'assertion b) que g est dans G_v . $\square$

Le phénomène de Mautner permet de montrer les deux lemmes suivants.

Lemme 8.7 *Soit v dans $\mathcal{H}$ et $t \neq 1$. Si v est a_t -invariant, alors, pour tout s dans $\mathbb{R}$, le vecteur v est u_s -invariant et u_s^- -invariant.*

Preuve. On peut supposer $t > 1$. Pour montrer que v est u_s -invariant, on applique le lemme 8.6 c) avec $g_n = g = u_s$, $s_n = a_t^{-n}$ et $s'_n = a_t^n$. On a juste à vérifier que

$$s_n g_n s'_n = \begin{pmatrix} 1 & t^{-2n} s \\ 0 & 1 \end{pmatrix} \xrightarrow{n \rightarrow \infty} e ,$$

ce qui est clair. On montre de même que v est u_s^- -invariant. $\square$

Lemme 8.8 *Soit v dans $\mathcal{H}$ et $s \neq 0$. Si v est u_s -invariant (respectivement u_s^- -invariant), alors pour tout $t > 0$, le vecteur v est a_t -invariant.*

Preuve. Supposons que v est u_s -invariant. Pour montrer que v est a_t -invariant, il suffit de le prouver lorsque t est un rationnel $t = \frac{p}{q}$. On applique alors le lemme 8.6 c) avec

$$g = a_t = \begin{pmatrix} \frac{p}{q} & 0 \\ 0 & \frac{q}{p} \end{pmatrix}, \quad g_n = \begin{pmatrix} \frac{p}{t-1} & 0 \\ \frac{t-1}{snp} & \frac{q}{p} \end{pmatrix},$$

$$s_n = u_s^{-np} = \begin{pmatrix} 1 & -snp \\ 0 & 1 \end{pmatrix} \quad \text{et} \quad s_n = u_s^{np} = \begin{pmatrix} 1 & snp \\ 0 & 1 \end{pmatrix} .$$

On a juste à vérifier que

$$s_n g_n s'_n = \begin{pmatrix} 1 & 0 \\ \frac{t-1}{snp} & 1 \end{pmatrix} \xrightarrow{n \rightarrow \infty} e ,$$

ce qui est clair. On procède de même façon lorsque v est u_s^- -invariant. $\square$

Preuve de la proposition 8.5. Soit v un vecteur a_t , u_s ou u_s^- -invariant de $\mathcal{H}$ pour un $t \neq 1$ ou $s \neq 0$. Il résulte des lemmes 8.7 et 8.8 que le groupe G_v fixant v contient tous les éléments de la forme $a_{t'}$, $u_{s'}$ ou $u_{s'}^-$. Comme ces éléments engendrent G , on a $G_v = G$. $\square$

8.4 Décroissance des coefficients et mélange

L'étude précédente pour $G = SL_2(\mathbb{R})$ va nous permettre de montrer le résultat plus général suivant.

Théorème 8.9 *Soit Γ un réseau de $G = SL_N(\mathbb{R})$ et g un élément de G tel que $\lim_{n \rightarrow \infty} \|g^n\| = \infty$. Alors l'action de g sur $\Gamma \backslash G$ est mélangeante pour μ .*

Preuve. D'après la proposition 6.1, il suffit d'appliquer la proposition suivante avec $\mathcal{H} = \mathbb{L}_0^2(\Gamma \backslash G, \mu)$, car tout élément G -invariant de $\mathcal{H}$ est nul (voir la fin du paragraphe 8.2). $\square$

Proposition 8.10 (Théorème de Howe-Moore) *Soit $G = SL_N(\mathbb{R})$, π une représentation unitaire de G dans un espace de Hilbert $\mathcal{H}$, et v, w des vecteurs de $\mathcal{H}$. Si $\mathcal{H}^G = \{0\}$, alors*

$$\lim_{\|g\| \rightarrow \infty} \langle \pi(g)v, w \rangle = 0 .$$

Pour montrer ce théorème, nous aurons besoin de deux résultats bien classiques (voir par exemple [MT], ainsi que la proposition 4.10 pour le second). Notons $K = SO_N(\mathbb{R})$ et

$$A^+ = \left\{ g = \begin{pmatrix} t_1 & & 0 \\ & \ddots & \\ 0 & & t_N \end{pmatrix} / t_1 \geq \dots \geq t_N > 0 \text{ et } \prod_{i=1}^N t_i = 1 \right\}.$$

Lemme 8.11 (Décomposition de Cartan) *On a $G = KA^+K$.*

Autrement dit, tout élément g de G peut s'écrire $g = k_1 a k_2$ avec k_1, k_2 dans K et a dans A^+ .

Preuve. La matrice ${}^t g g$ est symétrique, définie positive. Il existe donc une matrice symétrique définie positive p telle que ${}^t g g = p^2$. On a donc $g = k' p$ avec $k' = g p^{-1}$ dans K . On diagonalise p dans une base orthonormée en ordonnant ses valeurs propres, d'où $p = k a k^{-1}$ avec k dans K et a dans A^+ . On a alors $g = k' k a k^{-1}$. C'est ce que l'on voulait. $\square$

On dit qu'une suite v_n dans un espace de Hilbert $\mathcal{H}$ converge faiblement vers un vecteur v dans $\mathcal{H}$ si, pour tout w dans $\mathcal{H}$, on a

$$\lim_{n \rightarrow \infty} \langle v_n, w \rangle = \langle v, w \rangle.$$

Le produit scalaire de $\mathcal{H}$ étant non dégénéré, ce vecteur v est alors unique.

Lemme 8.12 (Compacité faible) *Toute suite bornée v_n dans un espace de Hilbert séparable admet une sous-suite $v_{\tau(n)}$ qui converge faiblement.*

Preuve. Soit $(w_i)_{i \in \mathbb{N}}$ une partie dénombrable dense dans $\mathcal{H}$. Par un procédé d'extraction diagonale, on peut supposer que pour tout i , la suite bornée $n \mapsto \langle v_n, w_i \rangle$ converge. On en déduit, par approximation, que pour tout w dans $\mathcal{H}$, la suite $\langle v_n, w \rangle$ est de Cauchy, donc converge vers un réel $f(w)$. La majoration $|f(w)| \leq \|w\| \max_{n \in \mathbb{N}} \|v_n\|$ montre que f est une forme linéaire continue, donc par le théorème de Riesz (le dual de $\mathcal{H}$ est $\mathcal{H}$ pour la dualité donnée par le produit scalaire, voir par exemple [Rud]), de la forme $f(w) = \langle v, w \rangle$. $\square$

Preuve du théorème de Howe-Moore. Comme G est séparable, quitte à se restreindre au plus petit sous-espace fermé G -invariant contenant v, w , on peut supposer que $\mathcal{H}$ est séparable.

Supposons par l'absurde qu'il existe $\epsilon > 0$, v', w' dans $\mathcal{H}$ et une suite g_n dans G tels que

$$\forall n \in \mathbb{N}, \quad |\langle \pi(g_n) v', w' \rangle| \geq \epsilon \quad \text{et} \quad \lim_{n \rightarrow \infty} \|g_n\| = \infty.$$

Quitte à extraire une sous-suite, on peut supposer que $\pi(g_n) v'$ converge faiblement vers un vecteur v'_0 . Notre hypothèse assure que v'_0 est non nul.

Par la décomposition de Cartan, écrivons $g_n = k_{1,n} a_n k_{2,n}$ avec $k_{1,n}, k_{2,n}$ dans K et a_n dans A^+ . Comme K est compact, quitte à extraire, on peut supposer que les limites $k_1 = \lim_{n \rightarrow \infty} k_{1,n}$ et $k_2 = \lim_{n \rightarrow \infty} k_{2,n}$ existent. On pose $v = \pi(k_2) v'$ et $v_0 = \pi(k_1^{-1}) v'_0$.

Lemme 8.13 *Avec ces notations, la suite a_n de A^+ vérifie*

$$\lim_{n \rightarrow \infty} \|a_n\| = \infty$$

et la suite $\pi(a_n) v$ converge faiblement vers le vecteur non nul v_0 .

Preuve. La première assertion est claire. Pour montrer la seconde, on calcule, pour tout w dans $\mathcal{H}$, en posant $w'' = \pi(k_1)w$,

$$|\langle \pi(a_n)v - v_0, w \rangle| = |\langle \pi(k_1 a_n k_2)v' - v'_0, w'' \rangle| \leq T_1 + T_2 + T_3$$

avec

$$T_1 = |\langle \pi(k_1 a_n k_2)v' - \pi(k_1 a_n k_{2,n})v', w'' \rangle| \leq \|\pi(k_2)v' - \pi(k_{2,n})v'\| \|w''\| \xrightarrow{n \rightarrow \infty} 0 ,$$

$$T_2 = |\langle \pi(k_1 a_n k_{2,n})v' - \pi(k_{1,n} a_n k_{2,n})v', w'' \rangle| \leq \|v'\| \|\pi(k_1^{-1})w'' - \pi(k_{1,n}^{-1})w''\| \xrightarrow{n \rightarrow \infty} 0 ,$$

et

$$T_3 = |\langle \pi(g_n)v' - v'_0, w'' \rangle| \xrightarrow{n \rightarrow \infty} 0 . \quad \square$$

Notons E_{ij} la matrice $N \times N$ dont les coefficients sont tous nuls sauf celui à la place (i, j) qui vaut 1. Pour $i \neq j$, on note U_{ij}, A_{ij}, S_{ij} les sous-groupes suivants de G :

$$U_{ij} = \{id + sE_{ij} / s \in \mathbb{R}\}$$

$$A_{ij} = \{id + (t - 1)E_{ii} + (t^{-1} - 1)E_{jj} / t > 0\}$$

$$S_{ij} = \{id + (a - 1)E_{ii} + bE_{ij} + cE_{ji} + (d - 1)E_{jj} / ad - bc = 1\}$$

Les groupes S_{ij} sont isomorphes au groupe $SL_2(\mathbb{R})$. On pourra leur appliquer la proposition 8.5.

Lemme 8.14 *Avec ces notations, il existe $k \in \{1, \dots, N - 1\}$ tel que, pour tout i, j avec $1 \leq i \leq k < j \leq N$, le vecteur v_0 est U_{ij} -invariant.*

Preuve. Les matrices a_n sont diagonales avec des coefficients diagonaux $(t_{i,n})_{1 \leq i \leq N}$ décroissants. Comme $\lim_{n \rightarrow \infty} \|a_n\| = \infty$ et $\det a_n = 1$, on peut trouver un entier k tel que $\lim_{n \rightarrow \infty} t_{k,n}/t_{k+1,n} = \infty$.

Soit u dans U_{ij} avec $1 \leq i \leq k < j \leq N$. On veut montrer que $\pi(u)v_0 = v_0$.

Rappelons que si X est une matrice diagonale inversible, de coefficients diagonaux $x_1, \dots, x_N$, et si $Y = (y_{ij})_{1 \leq i, j \leq N}$, alors la matrice XYX^{-1} est égale à $(\frac{x_i}{x_j}y_{ij})_{1 \leq i, j \leq N}$.

Le choix de k assure donc que l'on a $\lim_{n \rightarrow \infty} a_n^{-1}u a_n = e$. Par un calcul analogue à celui du lemme 8.13, on a alors (les limites étant prises au sens de la convergence faible sur $\mathcal{H}$)

$$\pi(u)v_0 = \lim_{n \rightarrow \infty} \pi(u)\pi(a_n)v = \lim_{n \rightarrow \infty} \pi(a_n)\pi(a_n^{-1}u a_n)v = \lim_{n \rightarrow \infty} \pi(a_n)v = v_0 .$$

C'est ce que l'on voulait. $\square$

Lemme 8.15 *Le groupe $G = SL_N(\mathbb{R})$ est engendré par les sous-groupes A_{ij} et U_{ij} avec $1 \leq i, j \leq N$ et $i \neq j$.*

Preuve. Pour g dans G , la matrice $(1 + sE_{ij})g$ est obtenue à partir de g par “une opération élémentaire sur les lignes” qui consiste à ajouter à la i -ème ligne de g un multiple de sa j -ème ligne. De même, la matrice $g(1 + sE_{ij})$ est obtenue à partir de g par “une opération élémentaire sur les colonnes”. Partant de n’importe quelle matrice g , on se ramène, par une succession d’opérations élémentaires sur les lignes et les colonnes à la matrice identité. $\square$

Terminons maintenant la preuve du théorème de Howe-Moore. Pour résumer ce qui précède, en raisonnant par l’absurde, nous avons construit un entier $k \leq N - 1$ et un vecteur non nul v_0 de $\mathcal{H}$, qui est invariant par tous les groupes U_{ij} avec $1 \leq i \leq k < j \leq N$. D’après la proposition 8.5, ce vecteur v_0 est S_{ij} -invariant. Il est en particulier A_{ij} -invariant et donc invariant par toutes les matrices $A_{i'j'}$ avec $1 \leq i', j' \leq N$ et $i' \neq j'$. De nouveau, par la proposition 8.5, ce vecteur est $U_{i'j'}$ -invariant. Le lemme 8.15 ci-dessus prouve que le vecteur non nul v_0 est G -invariant, contradiction. $\square$

8.5 Une construction de réseaux uniformes de $SL_2(\mathbb{R})$

Il existe plusieurs façons de construire des réseaux cocompacts de $SL_2(\mathbb{R})$ par l’analyse, la géométrie ou l’arithmétique. La méthode ci-dessous est arithmétique, voir par exemple [Kat]. Elle a l’avantage d’être très explicite et de se généraliser facilement à $SL_N(\mathbb{R})$, voir par exemple [BHC]. (Il suffit de disposer d’un corps gauche (i.e. non commutatif) $A_{\mathbb{Q}}$, de dimension N^2 sur $\mathbb{Q}$, tel que l’algèbre $A_{\mathbb{Q}} \otimes_{\mathbb{Q}} \mathbb{R}$ soit isomorphe à l’algèbre des matrices $\mathcal{M}(N, \mathbb{R})$. De tels corps gauches existent, voir par exemple [CohP].)

Fixons deux entiers a, b au moins égaux à 1, tels que l’équation $x^2 - ay^2 - bz^2$ n’ait pas de solution entières non nulles, par exemple $a = 2$ et $b = 3$. On note

$$g_v = \begin{pmatrix} x + y\sqrt{a} & \sqrt{b}(z - t\sqrt{a}) \\ \sqrt{b}(z + t\sqrt{a}) & x - y\sqrt{a} \end{pmatrix}$$

pour tout $v = (x, y, z, t)$ dans $\mathbb{R}^4$.

Proposition 8.16 *L’ensemble $\Gamma = \{g_v / v \in \mathbb{Z}^4 \text{ et } \det(g_v) = 1\}$ est un sous-groupe discret cocompact du groupe $G = SL_2(\mathbb{R})$.*

Preuve. Vérifions tout d’abord que Γ est un groupe. Pour v dans $\mathbb{R}^4$, on a $g_v = x \mathbf{1} + y I + z J + t K$ avec

$$\mathbf{1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad I = \begin{pmatrix} \sqrt{a} & 0 \\ 0 & -\sqrt{a} \end{pmatrix}, \quad J = \begin{pmatrix} 0 & \sqrt{b} \\ \sqrt{b} & 0 \end{pmatrix} \text{ et } K = \begin{pmatrix} 0 & -\sqrt{ab} \\ \sqrt{ab} & 0 \end{pmatrix}.$$

Les égalités $I^2 = a \mathbf{1}$, $J^2 = b \mathbf{1}$, $IJ = -JI = -K$ assurent que $A_{\mathbb{Q}} = \{g_v / v \in \mathbb{Q}^4\}$ est une $\mathbb{Q}$ -algèbre et que $A_{\mathbb{Z}} = \{g_v / v \in \mathbb{Z}^4\}$ est un anneau. On appelle $A_{\mathbb{Q}}$ (respectivement $A_{\mathbb{Z}}$) une *algèbre de quaternions* sur $\mathbb{Q}$ (respectivement *anneau des quaternions entiers*). On appelle *conjugué* de g_v l’élément $g_{\bar{v}}$ avec $\bar{v} = (x, -y, -z, -t)$. On a l’égalité

$$g_v g_{\bar{v}} = g_{\bar{v}} g_v = \det(g_v) \mathbf{1}.$$

Ce qui prouve que Γ est un groupe. Il est clair que Γ est discret dans $G = SL_2(\mathbb{R}) \subset \mathbb{R}^4$.

Lemme 8.17 *Tout élément non nul de $A_{\mathbb{Q}}$ est inversible (donc $A_{\mathbb{Q}}$ est un corps gauche).*

Preuve. Cela résulte de notre hypothèse sur a, b . En effet, si $\det(g_v) = 0$, on a $x^2 - ay^2 = b(z^2 - at^2)$. En multipliant par $z^2 - at^2$, on en déduit que $X^2 - aY^2 - bZ^2 = 0$ avec $X = xz + ayt$, $Y = xt + yz$ et $Z = z^2 - at^2$. Les solutions entières de cette équation sont nulles. Il en est de même des solutions rationnelles. Donc $X = Y = Z = 0$. L'hypothèse sur a, b et l'égalité $Z = 0$ impliquent que $z = t = 0$. Donc $x = y = z = t = 0$, contradiction. $\square$

Notons que l'algèbre $A_{\mathbb{R}} = \{g_v \mid v \in \mathbb{R}^4\}$ est égale à $\mathcal{M}(2, \mathbb{R})$, donc contient des éléments non nuls non inversibles.

Nous aurons besoin de deux lemmes classiques.

Lemme 8.18 (Minkovski) *Soient $(e_1, \dots, e_N)$ une base de $\mathbb{R}^N$, $\Delta = |\det(e_1, \dots, e_N)|$, $L = \mathbb{Z}e_1 + \dots + \mathbb{Z}e_N$ et C un convexe fermé de $\mathbb{R}^N$ qui est symétrique (i.e. $(v \in C) \Leftrightarrow (-v \in C)$) et tel que $\text{vol}(C) > 2^n \Delta$. Alors C contient un vecteur non nul de L .*

C'est encore vrai sous l'hypothèse $\text{vol}(C) \geq 2^n \Delta$, par un argument de passage à la limite.

Preuve. La mesure de Lebesgue sur $\mathbb{R}^N$ induit sur le tore $\mathbb{R}^N/L$ une mesure μ telle que $\mu(\mathbb{R}^N/L) = \Delta$. Comme $\text{vol}(\frac{1}{2}C) > \Delta$, la projection canonique $p : \mathbb{R}^N \rightarrow \mathbb{R}^N/L$, restreinte à $\frac{1}{2}C$ n'est pas injective. Il existe donc des points x, y dans $\frac{1}{2}C$ tel que $x - y$ est dans $L - \{0\}$. Par symétrie de C , les points $2x$ et $-2y$ sont dans C . Donc, par convexité de C , leur milieu $x - y$ est aussi dans C . Le résultat en découle. $\square$

Pour montrer que Γ est cocompact dans G , on doit construire un compact K de G tel que, pour tout g dans G , il existe g_v dans Γ tel que $g_v g$ est dans K . On va tout d'abord chercher g_v dans $A_{\mathbb{Z}} - \{0\}$.

Corollaire 8.19 *Il existe un compact C de $\mathcal{M}(2, \mathbb{R})$ tel que, pour tout g dans G , il existe g_v dans $A_{\mathbb{Z}} - \{0\}$ tel que $g_v g$ soit dans C .*

Preuve. L'espace vectoriel $\mathcal{M}(2, \mathbb{R})$ s'identifie à $\mathbb{R}^4$. Les images e_1, e_2, e_3, e_4 des éléments $1, I, J, K$ par l'application $g_v \mapsto g_v g$ forment une base, dont la valeur absolue du déterminant $\Delta = 4ab$ ne dépend pas de g . Il suffit donc d'appliquer le lemme de Minkovski à une boule euclidienne centrée en 0 et de rayon suffisamment grand dans $\mathbb{R}^4$. $\square$

Pour pouvoir choisir g_v dans Γ , nous aurons besoin du lemme suivant.

Lemme 8.20 *Soit n un entier non nul et $B_n = \{g \in \mathcal{M}(N, \mathbb{Z}) \mid \det g = n\}$. Alors il existe une partie finie G_n de B_n telle que $B_n = G_n SL_N(\mathbb{Z})$.*

Preuve. Multiplier à gauche une matrice g par une matrice de la forme $Id + sE_{ij}$ avec s dans $\mathbb{Z}$ permet d'ajouter à la j -ème colonne de g un multiple entier de la i -ème colonne, pour tout $i \neq j$. Par une succession de telles opérations, on peut remplacer toute matrice g de B_n par une matrice

$$g' = \begin{pmatrix} a_{11} & & 0 \\ & \ddots & \\ a_{N1} & & a_{NN} \end{pmatrix}$$

qui est triangulaire, avec $|a_{ij}| < |a_{ii}|$ et $\prod_{1 \leq i \leq N} a_{ii} = n$.

(En effet, on met la matrice g sous forme triangulaire inférieure, en annulant les coefficients strictement au-dessus de la diagonale par récurrence sur les lignes, et pour la première ligne, on raisonne par récurrence sur l'entier $\alpha = \sum_{j=1}^n |a_{1j}|$. Si $\alpha = 0$, c'est impossible, car n est non nul. Sinon, soit j_0 tel que $|a_{1j_0}| = \max\{|a_{1j}| / j = 1, \dots, n\}$. Si $a_{1j} = 0$ pour tout $j \neq j_0$, alors si $j_0 = 1$, c'est fini, sinon on rajoute à la première colonne la j_0 -ème, puis on enlève la première à la j_0 -ème et c'est fini. Sinon, soit j_1 différent de j_0 tel que $a_{1j_1} \neq 0$. Alors en ajoutant ou enlevant la j_1 -ème colonne à la j_0 -ème, on diminue strictement α .)

Or il n'y a qu'un nombre fini de telles matrices g' . □

Corollaire 8.21 *Soit n un entier non nul et $A_n = \{g \in A_{\mathbb{Z}} / \det g = n\}$. Alors il existe une partie finie F_n de A_n telle que $A_n = F_n \Gamma$.*

Preuve. La base $\mathbf{1}, I, J, K$ de $A_{\mathbb{Q}}$ permet d'identifier $\text{End}_{\mathbb{Q}}(A_{\mathbb{A}})$ avec $\mathcal{M}(4, \mathbb{Q})$. Pour tout élément g de $A_{\mathbb{Q}}$, on note a_g la matrice 4-4 donnée par la multiplication à gauche par g dans cette base. On a l'égalité (exercice) :

$$\det a_g = (\det g)^2 .$$

D'après le lemme précédent, il existe donc une partie F_n de A_n telle que, pour tout g dans A_n , il existe f dans F_n tel que la matrice $(a_f)^{-1} \circ a_g$ est dans $SL_4(\mathbb{Z})$. (En effet, soit $F' = \{f_1, \dots, f_m\}$ une partie minimale de B_{n^2} telle que $a_g \in F' SL_4(\mathbb{Z})$ pour tout g dans $A_{\mathbb{Z}}$. Soient $g_1, \dots, g_m$ dans $A_{\mathbb{Z}}$ tels que $a_{g_i} \in f_i SL_4(\mathbb{Z})$. Alors $F_n = \{g_1, \dots, g_m\}$ convient.) Comme la matrice $a_{f^{-1}g} = (a_f)^{-1} \circ a_g$ est à coefficient entiers, l'élément $f^{-1}g$ est dans $A_{\mathbb{Z}}$. Comme $\det(f^{-1}g) = \det(g)/\det(f) = 1$, l'élément $f^{-1}g$ est dans Γ . □

Fin de la preuve de la proposition 8.16. D'après le corollaire 8.19, il existe un compact C de $\mathcal{M}(2, \mathbb{R})$ tel que, pour tout g dans G , il existe g_v dans $A_{\mathbb{Z}} - \{0\}$ tel que $g_v g$ est dans C . D'après le lemme 8.17, on a $\det(g_v) \neq 0$. D'autre part, l'entier $n = \det(g_v) = \det(g_v g)$ reste borné, car $g_v g$ reste dans C . Cet entier n ne peut prendre qu'un nombre fini de valeurs. D'après le corollaire 8.21, on peut donc trouver une partie finie F de $A_{\mathbb{Z}} - \{0\}$ telle que tous ces éléments g_v s'écrivent $g_v = f\gamma$ avec f dans F et γ dans Γ . L'élément γg appartient alors au compact $K = \bigcup_{f \in F} f^{-1}C$. Donc Γ est cocompact dans G . □

8.6 Interprétation géométrique

Il existe une façon géométrique très simple de visualiser le groupe $SL_2(\mathbb{R})$, ou plus précisément le groupe quotient $G = PSL_2(\mathbb{R})$ de $SL_2(\mathbb{R})$ par son centre $\{\pm Id\}$. Un élément de G est, par définition une matrice $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ de déterminant 1 définie au signe près.

Le groupe G agit sur le *demi-plan supérieur* $\mathbb{H} = \{z = x + iy \in \mathbb{C} / y = \text{Im } z > 0\}$ par la formule

$$g \cdot z = \frac{az + b}{cz + d} .$$

En effet, cette formule ne dépend pas du signe de g , et un petit calcul montre que l'on a bien $g \cdot (g' \cdot z) = (gg') \cdot z$. Notons que $z \mapsto g \cdot z$ est une application holomorphe sur l'ouvert $\mathbb{H}$, dont la différentielle est $\xi \mapsto \frac{\xi}{(cz+d)^2}$.

Pour $v = (z, \xi)$ dans $\mathbb{H} \times \mathbb{C}$, on note

$$|v|_z = \frac{|\xi|}{y} ,$$

et on appelle cette quantité la *norme hyperbolique* de v . [Ceci définit une *structure riemannienne* sur $\mathbb{H}$, voir par exemple [GHL]] On appelle *plan hyperbolique* l'espace $\mathbb{H}$ muni de la norme hyperbolique. On note $T^1\mathbb{H} = \{v = (z, \xi) \in \mathbb{H} \times \mathbb{C} / |v|_z = 1\}$, que l'on appelle le *fibré unitaire tangent* du plan hyperbolique. Le groupe G agit aussi sur $T^1\mathbb{H}$ par la formule

$$g \cdot (z, \xi) = \left(\frac{az + b}{cz + d}, \frac{\xi}{(cz + d)^2} \right),$$

comme on le vérifie facilement. [Comme $|g \cdot v|_{g \cdot z} = |v|_z$, ceci traduit le fait que G est un groupe d'isométries de la structure riemannienne.] Le lemme suivant est immédiat.

Lemme 8.22 – *L'application $\Phi : G \rightarrow T^1\mathbb{H}$ définie par $g \mapsto g \cdot (i, i)$ est une bijection (et même un homéomorphisme).*

– *Pour tout g, g' dans G , on a $\Phi(gg') = g\Phi(g')$.* □

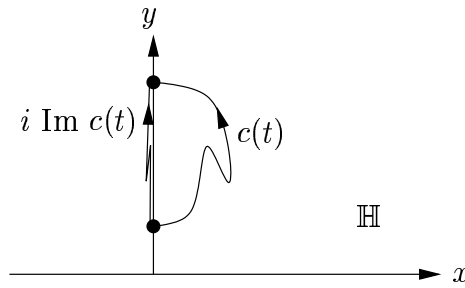
On définit la *longueur (hyperbolique)* $\ell(c)$ d'une courbe $c : [a, b] \rightarrow \mathbb{H}$, de classe C^1 par morceaux, par la formule

$$\ell(c) = \int_a^b |c'(t)|_{c(t)} dt$$

et la *distance (hyperbolique)* $d(w, z)$ entre deux points de $\mathbb{H}$ comme la borne inférieure des longueurs des courbes C^1 par morceaux entre x et y . Il est facile de vérifier que d est bien une distance, invariante par l'action de G sur $\mathbb{H}$. On appelle *géodésique (hyperbolique)* toute courbe $c : I \rightarrow \mathbb{H}$, avec I un intervalle de $\mathbb{R}$, telle que pour tous $s < t$ dans I , on a $d(c(s), c(t)) = t - s$.

Lemme 8.23 *Les géodésiques décrivent les arcs de demi-droites et demi-cercles dans $\mathbb{H}$ orthogonaux à l'axe réel.*

Preuve. Comme $|c'(t)| \geq |(\text{Im } c)'(t)|$, on a $\ell(c) \geq \ell(i \text{Im } c)$ pour tout chemin c de classe C^1 par morceaux entre deux points de l'axe imaginaire, avec inégalité stricte si la courbe c n'est pas contenue dans l'axe imaginaire.

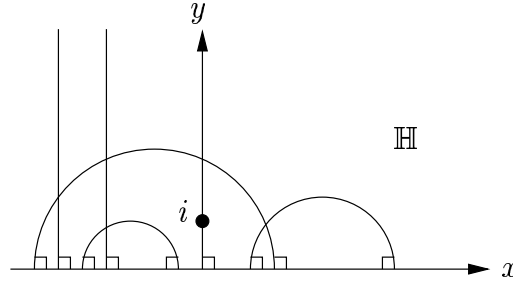


De plus, tout chemin sur l'axe imaginaire faisant un aller-retour non trivial est de longueur (hyperbolique) strictement supérieure à celle de ce chemin privé de cet aller-retour. Donc, pour tout $y > 1$,

$$d(i, yi) = \int_1^y \frac{ds}{s} = \log y,$$

et la courbe $c : [0, \log y] \rightarrow \mathbb{H}$ définie par $t \mapsto e^t i$ est l'unique géodésique C^1 par morceaux joignant i à $y i$. Par approximation, on montre facilement que c'est aussi la seule géodésique (continue) joignant i à $y i$ (et elle est donc C^∞). Comme tout élément g de G est une isométrie hyperbolique, la courbe $t \mapsto g \cdot (e^t i)$ est l'unique géodésique joignant $g \cdot i$ à $g \cdot (y i)$.

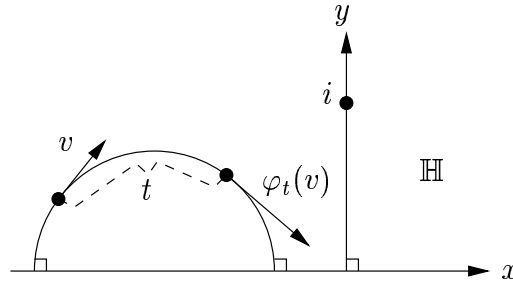
Il est bien connu (voir cours d'analyse complexe) que les homographies $z \mapsto \frac{az+b}{cz+d}$ avec a, b, c, d dans $\mathbb{C}$ et $ad - bc = 1$ préservent la famille des cercles-droites, ainsi que la relation d'orthogonalité, qu'elles préservent le demi-plan supérieur si a, b, c, d sont réels, et qu'elles agissent de manière transitive sur la famille des cercles-droites.



Le résultat en découle. $\square$

Pour tout $v = (z, \xi)$ dans $T^1\mathbb{H}$, il existe donc une unique géodésique $t \mapsto c(t)$, définie sur $\mathbb{R}$, telle que $v = (c(0), c'(0))$. On note alors

$$\varphi_t(v) = (c(t), c'(t)) .$$



Lemme 8.24 *La famille $(\varphi_t)_{t \in \mathbb{R}}$ est un groupe à un paramètre de difféomorphismes de $T^1\mathbb{H}$, appelé flot géodésique sur $\mathbb{H}$, qui commute avec l'action de G sur $T^1\mathbb{H}$ (i.e. $\varphi_t(g \cdot v) = g \cdot \varphi_t(v)$). De plus, si $v = \Phi(g)$, alors*

$$\varphi_t(v) = \Phi(ga_{\frac{t}{2}})$$

où $a_t = \begin{pmatrix} e^t & 0 \\ 0 & e^{-t} \end{pmatrix} \bmod \pm 1$.

Preuve. La première assertion est immédiate. Pour la seconde, il suffit, par équivariance, de vérifier que $\varphi_t(i, i) = \Phi(a_{t/2})$, ce qui est clair car ces deux termes valent $(e^t i, e^t i)$. $\square$

Pour tout sous-groupe discret Γ de G , le flot φ_t sur $T^1\mathbb{H}$ induit par passage au quotient un flot sur $\Gamma \backslash T^1\mathbb{H}$, appelé le *flot géodésique* sur $\Gamma \backslash \mathbb{H}$, et que l'on note encore φ_t . L'application Φ induit aussi par passage au quotient une application $\Gamma \backslash G \rightarrow \Gamma \backslash T^1\mathbb{H}$, que l'on note encore Φ .

En conclusion, le flot étudié au paragraphe 8.3 défini par la multiplication à droite par a_t n'est rien d'autre que le flot géodésique sur les quotients de volume fini $\Gamma \backslash \mathbb{H}$, i.e. le diagramme suivant commute :

$$\begin{array}{ccc} \Gamma \backslash G & \xrightarrow{a_{t/2}} & \Gamma \backslash G \\ \Phi \downarrow & & \downarrow \Phi \\ \Gamma \backslash T^1 \mathbb{H} & \xrightarrow{\varphi_t} & \Gamma \backslash T^1 \mathbb{H} . \end{array}$$

9 Entropie métrique

L'entropie ou *entropie métrique* est un invariant des systèmes dynamiques mesurables. L'entropie $h_\mu(f)$ d'une transformation f d'un espace mesurable X préservant une mesure de probabilité μ est un nombre réel positif ou nul, éventuellement infini, qui mesure le "désordre créé par f ". Esquissons sa définition. On définit tout d'abord l'entropie $H_\mu(\alpha)$ d'une partition α de X comme l'information moyenne issue de la connaissance de la partie de α dans laquelle se trouve un point de X . L'entropie $h_\mu(f, \alpha)$ de f relativement à α est alors le taux de croissance en n de l'information moyenne issue de la connaissance de la suite des parties de α dans lesquelles se trouvent les n premiers points des orbites. L'entropie $h_\mu(f)$ est alors la borne supérieure des entropies de f relativement aux partitions de X .

Nous verrons que cette borne supérieure est parfois atteinte, ce qui facilite beaucoup le calcul de $h_\mu(f)$.

9.1 Information et entropie d'une partition : définitions

Commençons par une présentation heuristique de ces concepts. Lorsqu'on découpe un espace en $p = 2^n$ parties de tailles égales (par un procédé de dichotomie par exemple), savoir dans quelle partie un point de X se trouve est une information qui peut s'exprimer à l'aide de n chiffres 0 ou 1. On dira que l'information de cette partition est cet entier $n = \log_2(p)$, ou, plutôt, le réel $\log p$ qui lui est proportionnel. Lorsque le découpage n'est plus en parts égales, l'information est une fonction de x égale à la valeur absolue du logarithme de la taille de la partie dans laquelle se trouve x . La valeur moyenne de cette information est alors l'entropie de la partition.

Nous introduirons aussi l'information relative d'une partition par rapport à une autre, qui reflète l'information supplémentaire qu'ajoute la première à la seconde. L'entropie relative sera la valeur moyenne de l'information relative.

Soit $(X, \mathcal{B}, \mu)$ un espace de probabilité. Nous appellerons *partition mesurable* une famille finie $\alpha = \{A_1, \dots, A_p\}$ de parties mesurables non vides telles que, pour $i \neq j$, on a $A_i \cap A_j = \emptyset$ et $A_1 \cup \dots \cup A_p = X$. Dans la suite, les symboles $\alpha, \beta, \gamma, \dots$ désigneront des partitions mesurables.

Remarque. On peut aussi admettre des familles dénombrables, avec des intersections deux à deux de mesure nulle, et de réunion de mesure totale dans X . C'est parfois plus pratique pour des calculs explicites. Cela ne change pas les définitions et résultats qui suivent (voir par exemple [PY, chap. 8] [KH, Part 1, 4.3]).

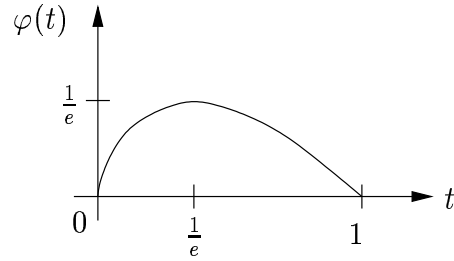
L'*information* de α est la fonction $I_\alpha : X \rightarrow [0, +\infty]$ donnée par

$$I_\alpha(x) = \sum_{A \in \alpha} (-\log \mu(A)) \mathbb{I}_A(x) .$$

L'*entropie* de α est la moyenne de I_α :

$$H(\alpha) = H_\mu(\alpha) = \int_X I_\alpha(x) d\mu(x) = \sum_{A \in \alpha} -\mu(A) \log \mu(A) .$$

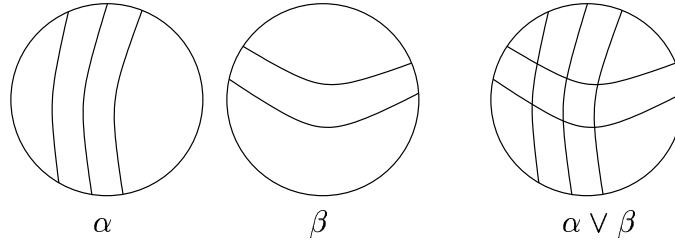
Il est pratique de noter $\varphi(t) = -t \log t$. C'est une application positive concave sur $[0, 1]$. On a $H(\alpha) = \sum_{A \in \alpha} \varphi(\mu(A))$. On rappelle la convention $0 \times \infty = 0$.



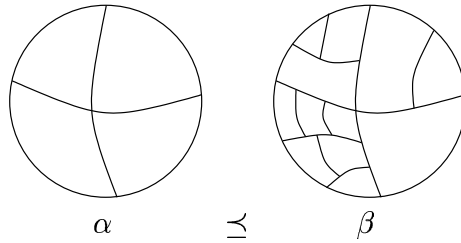
On note $\alpha \vee \beta$ le *joint* de α et β , c'est-à-dire la partition mesurable formée des intersections

$$\alpha \vee \beta = \{A \cap B \mid A \in \alpha, B \in \beta, A \cap B \neq \emptyset\}.$$

On note $\bigvee_{i=1}^n \alpha_i$ pour $\alpha_1 \vee \dots \vee \alpha_n$.



On dit que α est *moins fine* que β , en on note $\alpha \preceq \beta$, si tout élément de α est réunion d'éléments de β . On dit aussi que β est un *raffinement* de α . La relation $\preceq$ est clairement une relation d'ordre sur l'ensemble des partitions mesurables de X .



Remarque. On a l'équivalence $\alpha \preceq \beta \Leftrightarrow \alpha \vee \beta = \beta$.

L'*information relative* de α par rapport à β est la fonction

$$I_{\alpha|\beta}(x) = \sum_{A \in \alpha, B \in \beta} -\log \left(\frac{\mu(A \cap B)}{\mu(B)} \right) \mathbb{I}_{A \cap B}(x).$$

L'*entropie relative* de α par rapport à β est sa moyenne

$$H(\alpha|\beta) = \int_X I_{\alpha|\beta}(x) d\mu(x) = \sum_{A \in \alpha, B \in \beta} \varphi \left(\frac{\mu(A \cap B)}{\mu(B)} \right) \mu(B).$$

Remarques. (1) L'information relative est une fonction positive et donc $H(\alpha|\beta) \geq 0$. On a $I_\alpha = I_{\alpha|\{X\}}$ et $H(\alpha) = H(\alpha|\{X\})$.

(2) Si $\text{Card}(\alpha) = p$, alors $H(\alpha) \leq \log p$, avec égalité si et seulement si, pour tout A dans α , on a $\mu(A) = \frac{1}{p}$.

(3) On a $H(\alpha) \in [-\log(\sup_{A \in \alpha} \mu(A)), -\log(\inf_{A \in \alpha} \mu(A))]$.

La seconde remarque résulte de la stricte concavité de φ . En effet,

$$\frac{1}{p}H(\alpha) = \frac{1}{p} \sum_{A \in \alpha} \varphi(\mu(A)) \leq \varphi\left(\frac{1}{p} \sum_{A \in \alpha} \mu(A)\right) = \varphi\left(\frac{1}{p}\right) = \frac{1}{p} \log p.$$

9.2 Information et entropie d'une partition : propriétés

Proposition 9.1 (Additivité de l'information et de l'entropie) *On a*

1. $I_{\alpha \vee \beta | \gamma} = I_{\alpha | \beta \vee \gamma} + I_{\beta | \gamma}$,
2. $H(\alpha \vee \beta | \gamma) = H(\alpha | \beta \vee \gamma) + H(\beta | \gamma)$.

Preuve. La première assertion découle, en se ramenant sur $A \cap B \cap C$, de l'égalité

$$\log\left(\frac{\mu(A \cap B \cap C)}{\mu(C)}\right) = \log\left(\frac{\mu(A \cap B \cap C)}{\mu(B \cap C)}\right) + \log\left(\frac{\mu(B \cap C)}{\mu(C)}\right).$$

La seconde assertion s'en déduit par intégration. □

On en déduit, en prenant $\gamma = \{X\}$, le résultat suivant.

Corollaire 9.2 1. $I_{\alpha \vee \beta} = I_{\alpha | \beta} + I_{\beta}$,

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

Proposition 9.3 (Monotonie de l'entropie)

1. Si $\alpha \preceq \beta$, alors $H(\alpha | \gamma) \leq H(\beta | \gamma)$.
2. Si $\gamma \preceq \beta$, alors $H(\alpha | \beta) \leq H(\alpha | \gamma)$.

Preuve. (1) Pour A dans α , B dans β et C dans γ , si B est contenu dans A , alors

$$-\log\left(\frac{\mu(A \cap C)}{\mu(C)}\right) \leq -\log\left(\frac{\mu(B \cap C)}{\mu(C)}\right).$$

L'inégalité $I_{\alpha | \gamma} \leq I_{\beta | \gamma}$ s'en déduit. On conclut en intégrant sur X .

(2) Comme γ est moins fine que β , la partition β induit une partition β_C de tout élément C de γ . Par concavité de φ , on a

$$\sum_{B \in \beta_C} \varphi\left(\frac{\mu(A \cap B)}{\mu(B)}\right) \mu(B) \leq \varphi\left(\frac{\mu(A \cap C)}{\mu(C)}\right) \mu(C).$$

En sommant cette inégalité pour A dans α et C dans γ , on trouve bien

$$\sum_{A \in \alpha, B \in \beta} \varphi\left(\frac{\mu(A \cap B)}{\mu(B)}\right) \mu(B) \leq \sum_{A \in \alpha, C \in \gamma} \varphi\left(\frac{\mu(A \cap C)}{\mu(C)}\right) \mu(C).$$

□

Corollaire 9.4 1. Si $\alpha \preceq \beta$, alors $H(\alpha) \leq H(\beta)$.

2. $H(\alpha|\beta) \leq H(\beta)$.

3. $H(\alpha) - H(\beta) \leq H(\alpha|\beta)$.

Preuve. Les deux premières assertions découlent de la proposition en prenant $\gamma = \{X\}$. La dernière découle alors du (1) et du corollaire 9.2 (2), car $\alpha \preceq \alpha \vee \beta$. $\square$

Proposition 9.5 (Sous-additivité de l'entropie)

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

Preuve. Ceci résulte de la proposition 9.1 (2) et 9.3 (2), car $\gamma \preceq \beta \vee \gamma$. $\square$

Corollaire 9.6 $H(\alpha \vee \beta) \leq H(\alpha) + H(\beta)$. $\square$

Proposition 9.7 (Exhaustion de l'information) Soit $\beta_0 \succeq \beta_1 \succeq \dots \succeq \beta_n \succeq \dots$ une suite de plus en plus fine de partitions mesurables de X telle que la réunion $\bigcup_{n \in \mathbb{N}} \beta_n$ engendre la σ -algèbre $\mathcal{B}$. Alors $H(\alpha|\beta_n)$ tend vers 0 quand $n \rightarrow \infty$.

Remarque. Ceci signifie que $I_{\alpha|\beta_n}$ tend vers 0 dans $\mathbb{L}^1(X, \mu)$. On peut aussi montrer que $I_{\alpha|\beta_n}$ tend vers 0 μ -presque sûrement (c'est une conséquence du théorème de convergence des martingales).

Lemme 9.8 Soit $(X, \mathcal{B}, \mu)$ un espace de probabilité et $\mathcal{C}$ une sous-algèbre de Boole de $\mathcal{B}$ qui engendre la σ -algèbre $\mathcal{B}$. Alors, pour tout B dans $\mathcal{B}$ et $\epsilon > 0$, il existe C dans $\mathcal{C}$ telle que $\mu(B \Delta C) \leq \epsilon$.

On a noté ci-dessus $B \Delta C = (B \cup C) - (B \cap C)$ la *différence symétrique* de B et C . On a $\mu(B \Delta C) = \|\mathbb{I}_B - \mathbb{I}_C\|_{\mathbb{L}^1}$.

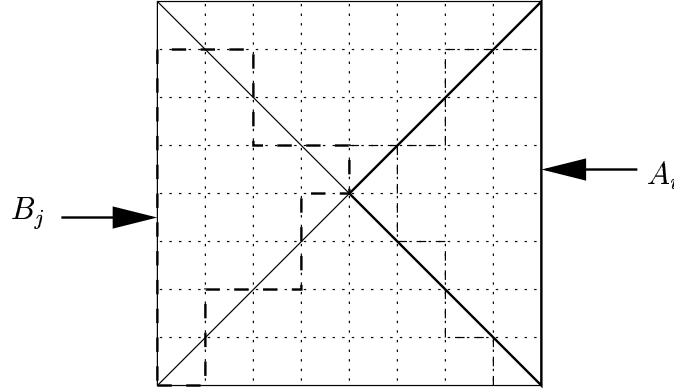
Preuve. Soit $\mathcal{B}' = \{B \in \mathcal{B} / \forall \epsilon > 0, \exists C \in \mathcal{C}, \mu(B \Delta C) \leq \epsilon\}$. Comme $\mathcal{B}'$ contient $\mathcal{C}$, il suffit de montrer que $\mathcal{B}'$ est une σ -algèbre. Or $\mathcal{B}$ est stable par

- passage au complémentaire, car ${}^c B \Delta {}^c C = B \Delta C$,
- union finie, car $(A_1 \cup A_2) \Delta (C_1 \cup C_2) \subset (A_1 \Delta C_1) \cup (A_2 \Delta C_2)$,
- union dénombrable croissante, car si B est une réunion croissante de B_n , alors $\mu(B \Delta B_n)$ tend vers 0.

$\square$

Preuve de la proposition 9.7 : Écrivons $\alpha = \{A_1, \dots, A_p\}$. On peut supposer que $\mu(A_i)$ est non nul pour tout i . On applique le lemme précédent avec $\mathcal{C}$ l'ensemble des unions finies d'éléments des β_n , qui est une algèbre de Boole, car les β_n sont des partitions de

plus en plus fine. Pour tout $\epsilon > 0$, on peut trouver $n_0 \geq 1$ et des unions (finies) $B_1, \dots, B_p$ d'éléments de β_{n_0} telles que $\mu(A_i \Delta B_i) \leq \epsilon$.



On a alors $\mu(B_i \cap B_j) \leq 2\epsilon$ pour $i \neq j$, car

$$B_i \cap B_j = (B_i \cap B_j \cap^c A_i) \cup (B_i \cap B_j \cap^c A_j) \subset (B_i \Delta A_i) \cup (B_j \Delta A_j) ,$$

et $\mu(A - \bigcup_{i=1}^p B_i) \leq p\epsilon$, car

$$A - \bigcup_{i=1}^p B_i = \bigcup_{j=1}^p (A_j - \bigcup_{i=1}^p B_i) \subset \bigcup_{j=1}^p (A_j - B_j) \subset \bigcup_{j=1}^p (A_j \Delta B_j) .$$

Quitte à changer de B_i et de ϵ , on peut supposer que $\beta = \{B_1, \dots, B_p\}$ est une partition de X . (Voir le dessin ci-dessus, avec les traits pointillés représentant la partition β_{n_0} , les traits en tirets la partition β et les traits pleins la partition α .) De nouveau, quitte à changer de ϵ , on peut supposer $\epsilon \leq \frac{1}{e}$ et $\mu(A_i \Delta B_i) \leq \frac{1}{2}\mu(A_i)\epsilon \leq \mu(B_i)\epsilon$. On a alors

$$\forall i, \quad \frac{\mu(A_i \cap B_i)}{\mu(B_i)} \geq 1 - \epsilon \quad \text{et} \quad \forall i \neq j, \quad \frac{\mu(A_i \cap B_j)}{\mu(B_j)} \leq \epsilon .$$

La proposition 9.3 donne, pour $n \geq n_0$, comme $\beta_n \succeq \beta_{n_0} \succeq \beta$,

$$H(\alpha|\beta_n) \leq H(\alpha|\beta_{n_0}) \leq H(\alpha|\beta) .$$

Or, comme φ est croissante sur $[0, \epsilon]$ et décroissante sur $[1 - \epsilon, 1]$, on a

$$H(\alpha|\beta) = \sum_{i,j} \varphi\left(\frac{\mu(A_i \cap B_j)}{\mu(B_j)}\right) \mu(B_j) \leq p^2 \max\{\varphi(\epsilon), \varphi(1 - \epsilon)\} .$$

Ce majorant tend vers 0 avec ϵ . Donc $\lim_{n \rightarrow \infty} H(\alpha|\beta_n) = 0$. □

La fin de cette démonstration prouve aussi le lemme suivant, qui sera utilisé plus loin.

Lemme 9.9 *Soit $(X, \mathcal{B}, \mu)$ un espace de probabilité, et α une partition mesurable de X . Alors, pour tout $\eta > 0$, il existe $\epsilon > 0$ tel que, si β est une partition mesurable de X telle que, pour tout A dans α , il existe B dans β avec $\mu(A \Delta B) \leq \epsilon$, alors on a $H(\alpha|\beta) \leq \eta$. □*

9.3 Entropie d'un système dynamique mesurable : définitions

Un *système dynamique probabilisé* $(X, \mathcal{B}, \mu, f)$ est la donnée d'un espace de probabilité $(X, \mathcal{B}, \mu)$ et d'une application $f : X \rightarrow X$ mesurable qui préserve μ . Le système est dit *inversible* si f est bijective et si f^{-1} préserve aussi μ .

Le lemme suivant affirme que les fonctions sous-additives ont un taux de croissance.

Lemme 9.10 *Soit $(a_n)_{n \geq 1}$ une suite réelle, telle que, pour tous $m, n \geq 1$, on a $a_{m+n} \leq a_m + a_n$. Alors $\lim_{n \rightarrow \infty} \frac{a_n}{n}$ existe et vaut $\inf_{n \geq 1} \frac{a_n}{n}$.*

Preuve. Soit $\ell = \inf_{n \geq 1} \frac{a_n}{n}$. Supposons tout d'abord $\ell > -\infty$. Alors la suite positive $b_n = a_n - \ell n$ est encore sous-additive, et pour tout $\epsilon > 0$, il existe $n_0 \geq 1$ tel que $\frac{b_{n_0}}{n_0} \leq \epsilon$. Écrivons $n = qn_0 + r$ avec $r < n_0$. Par sous-additivité, pour $n \geq n_0$, on a alors $\frac{b_n}{n} \leq \frac{qb_{n_0} + rb_1}{n} \leq \frac{b_{n_0}}{n_0} + \frac{n_0 b_1}{n}$. Donc $\frac{b_n}{n} \leq 2\epsilon$ pour n grand. De même, si $\ell = -\infty$, on montre que $\frac{a_n}{n}$ converge vers $-\infty$. $\square$

Soit $(X, \mathcal{B}, \mu, f)$ un système dynamique probabilisé. Soit α une partition mesurable de X . On note $f^{-1}\alpha$ la partition $\{f^{-1}(A) \mid A \in \alpha\}$. On a $f^{-1}(\alpha \vee \beta) = (f^{-1}\alpha) \vee (f^{-1}\beta)$. Puisque f préserve la mesure, on a $H(f^{-1}\alpha \mid f^{-1}\beta) = H(\alpha \mid \beta)$, et donc $H(f^{-1}\alpha) = H(\alpha)$.

Lemme 9.11 *La limite $\lim_{n \rightarrow \infty} \frac{1}{n} H\left(\bigvee_{i=0}^{n-1} f^{-i}\alpha\right)$ existe.*

On l'appelle l'*entropie* de f pour la partition α , et on la note $h(f, \alpha) = h_\mu(f, \alpha)$.

Preuve. On applique le lemme 9.10 à la suite $a_n = H(\bigvee_{i=0}^{n-1} f^{-i}\alpha)$. En effet, par le corollaire 9.6, on a $a_{m+n} \leq H(\bigvee_{i=0}^{n-1} f^{-i}\alpha) + H(\bigvee_{i=n}^{m+n-1} f^{-i}\alpha) = a_n + a_m$ car f préserve la mesure. $\square$

Remarque. Si $\alpha \preceq \beta$, alors $h(f, \alpha) \leq h(f, \beta)$ d'après le corollaire 9.4.

Le nombre $h(f) = h_\mu(f) = \sup_\alpha h(f, \alpha)$ dans $[0, +\infty]$ est appelé l'*entropie* de f .

Le résultat suivant dit que l'entropie métrique est un invariant des systèmes dynamiques probabilisés, i.e. qu'elle est invariante par conjugaison (au sens de la mesure).

Proposition 9.12 *Si $(X, \mathcal{B}, \mu, f)$ et $(X', \mathcal{B}', \mu', f')$ sont deux systèmes dynamiques probabilisés conjugués (au sens de la mesure), alors*

$$h_\mu(f) = h_{\mu'}(f') .$$

Preuve. Soit $\psi : X \rightarrow X'$ une bijection (quitte à enlever des ensembles de mesure nulle) mesurable et préservant la mesure, ainsi que son inverse, telle que $\psi \circ f = f' \circ \psi$. Si α' est une partition mesurable de X' , alors $\psi^{-1}\alpha' = \{\psi^{-1}(A) \mid A \in \alpha'\}$ est une partition mesurable de X . Les propriétés de ψ impliquent que

$$H\left(\bigvee_{i=0}^{n-1} (f')^{-i}\alpha'\right) = H\left(\bigvee_{i=0}^{n-1} f^{-i}\psi^{-1}\alpha'\right) .$$

En divisant par n et en prenant la limite quand n tend $+\infty$, on en déduit que $h_\mu(f, \psi^{-1}\alpha) = h_{\mu'}(f', \alpha)$. En prenant la borne supérieure sur α , on obtient $h_{\mu'}(f') \leq h_\mu(f)$. Par symétrie, la proposition en résulte. $\square$

Le lemme suivant exprime intuitivement que l'entropie de f est l'information moyenne supplémentaire de savoir où est un point x de X quand on sait où sont les images $f^i(x)$ pour $i \geq 1$. L'entropie est bien une quantité qui mesure le désordre créé par f .

Lemme 9.13 *La suite $c_n = H(\alpha | \bigvee_{i=1}^n f^{-i}\alpha)$ est décroissante et converge vers $h(f, \alpha)$.*

Preuve. La décroissance de c_n résulte de la proposition 9.3 (2). Notons c sa limite. Le corollaire 9.2 (2) donne

$$a_n = H\left(\bigvee_{k=0}^{n-1} f^{-k}\alpha\right) = \sum_{k=0}^{n-1} H\left(f^{-k}\alpha \middle| \bigvee_{i=k+1}^{n-1} f^{-i}\alpha\right) + H(f^{-n+1}\alpha) = H(\alpha) + \sum_{k=0}^{n-1} c_{n-k-1}.$$

Par le lemme de Césaro, la suite a_n/n converge donc vers c et on a bien $h(f, \alpha) = c$. $\square$

9.4 Entropie d'un système dynamique mesuré : propriétés

Proposition 9.14 *Soient $(X, \mathcal{B}, \mu, f)$ un système dynamique probabilisé, et α, β des partitions mesurables de X . Alors*

1. $h(f, \alpha \vee \beta) \leq h(f, \alpha) + h(f, \beta)$ (sous-additivité)
2. $h(f, \alpha) - h(f, \beta) \leq H(\alpha | \beta)$ (inégalité de Rokhlin))
3. pour tout $k \geq 1$, on a $h(f, \alpha) = h(f, \bigvee_{i=0}^k f^{-i}\alpha)$. De plus, si f est inversible, alors $h(f, \alpha) = h(f, \bigvee_{i=-k}^k f^{-i}\alpha)$.

Preuve. (1) Ceci résulte du corollaire 9.6 de sous-additivité de l'entropie et de la définition de $h(f, \alpha)$.

(2) En utilisant les propositions 9.1, 9.3 et 9.5 et leurs corollaires, on a

$$\begin{aligned} H\left(\bigvee_{i=0}^{n-1} f^{-i}\alpha\right) - H\left(\bigvee_{i=0}^{n-1} f^{-i}\beta\right) &\leq H\left(\bigvee_{i=0}^{n-1} f^{-i}\alpha \middle| \bigvee_{i=0}^{n-1} f^{-i}\beta\right) \leq \sum_{i=0}^{n-1} H\left(f^{-i}\alpha \middle| \bigvee_{i=0}^{n-1} f^{-i}\beta\right) \\ &\leq \sum_{i=0}^{n-1} H(f^{-i}\alpha | f^{-i}\beta) = nH(\alpha | \beta). \end{aligned}$$

D'où le résultat en divisant par n et en passant à la limite.

(3) Notons $a_n(\alpha) = H(\bigvee_{i=0}^{n-1} f^{-i}\alpha)$ et $\alpha_k = \bigvee_{i=0}^k f^{-i}\alpha$. Comme on a l'égalité

$$\bigvee_{i=0}^{n-1} f^{-i}\alpha_k = \bigvee_{i=0}^{n+k-1} f^{-i}\alpha,$$

on a $a_n(\alpha_k) = a_{n+k}(\alpha)$. Donc les suites $\frac{1}{n}a_n(\alpha_k)$ et $\frac{1}{n}a_n(\alpha)$ ont la même limite. On procède de même lorsque f est inversible, avec $\alpha_k = \bigvee_{i=-k}^k f^{-i}\alpha$. $\square$

Proposition 9.15 *Pour tout $k \geq 1$, on a $h_\mu(f^k) = k h_\mu(f)$.*

Preuve. Pour tout α , en utilisant la remarque suivant le lemme 9.11 pour la première inégalité, on a

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

Ceci prouve que $h(f^k, \alpha) \leq k h(f, \alpha)$, mais aussi que $k h(f, \alpha) = h\left(f^k, \bigvee_{i=0}^{k-1} f^{-i}\alpha\right) \leq h(f^k)$, par définition de $h(\cdot)$ comme borne supérieure. Donc en prenant la borne supérieure sur α , on a $h(f^k) = k h(f)$. $\square$

Le calcul de l'entropie d'une transformation f est en général compliqué. Le théorème suivant est l'un des moyens le plus pratique pour calculer une entropie (voir paragraphe 12.5), car il évite de prendre la borne supérieure sur toutes les partitions mesurables de X .

Une partition mesurable α de X est dite *génératrice* pour f si la suite des partitions $\alpha_n = \bigvee_{i=0}^n f^{-i}\alpha$ engendre la σ -algèbre $\mathcal{B}$, ou si f est inversible et si la suite des partitions $\alpha_n = \bigvee_{i=-n}^n f^{-i}\alpha$ engendre la σ -algèbre $\mathcal{B}$.

Théorème 9.16 (Théorème de Kolmogorov-Sinaï) *Si α est génératrice pour f , alors*

$$h_\mu(f) = h_\mu(f, \alpha) .$$

Preuve. Fixons $\epsilon > 0$. Il existe une partition mesurable β telle que $h(f, \beta) \geq h(f) - \epsilon$. D'après la proposition 9.7, il existe $n \geq 1$ tel que $H(\beta|\alpha_n) \leq \epsilon$. Mais alors, par la proposition 9.14 (3) puis (2), on a

$$h(f, \alpha) = h(f, \alpha_n) \geq h(f, \beta) - H(\beta|\alpha_n) \geq h(f) - 2\epsilon .$$

Or $h(f) \geq h(f, \alpha)$ et ceci est vrai pour tout $\epsilon > 0$, donc $h(f) = h(f, \alpha)$. $\square$

Exemple : Soit Λ un alphabet fini, ν une mesure de probabilité sur Λ , $X = \Lambda^{\mathbb{N}}$ ou $X = \Lambda^{\mathbb{Z}}$, μ la probabilité produit sur X et $f = \sigma$ le décalage sur X . Alors, en posant $p_\lambda = \nu(\{\lambda\})$ pour tout λ dans Λ , l'entropie métrique du décalage est

$$h_\mu(\sigma) = - \sum_{\lambda \in \Lambda} p_\lambda \log p_\lambda .$$

En effet, pour λ dans Λ , posons $A_\lambda = \{(x_i)_i \in X / x_0 = \lambda\}$. Soit $\alpha = \{A_\lambda / \lambda \in \Lambda\}$. Alors α est une partition mesurable génératrice, car pour tout $n \geq k$, la partition $\bigvee_{i=k}^n \sigma^{-i}\alpha$ est la partition par les cylindres $C_{\lambda_k, \dots, \lambda_n} = \{(x_i)_i \in X / x_k = \lambda_k, \dots, x_n = \lambda_n\}$, et que la σ -algèbre de X est engendrée par l'ensemble de tous ces cylindres. On calcule alors facilement

$$\begin{aligned} H\left(\bigvee_{i=0}^{n-1} \sigma^{-i}\alpha\right) &= - \sum_{\lambda_0, \dots, \lambda_{n-1} \in \Lambda} \mu(C_{\lambda_0, \dots, \lambda_{n-1}}) \log \mu(C_{\lambda_0, \dots, \lambda_{n-1}}) \\ &= - \sum_{\lambda_0, \dots, \lambda_{n-1} \in \Lambda} (p_{\lambda_0} \dots p_{\lambda_{n-1}}) \log(p_{\lambda_0} \dots p_{\lambda_{n-1}}) \\ &= -n \sum_{\lambda \in \Lambda} p_\lambda \log p_\lambda , \end{aligned}$$

en utilisant que $\sum_{\lambda_i \in \Lambda} p_{\lambda_i} = 0$. Le résultat s'en déduit.

Et en particulier, si ν est l'équiprobabilité sur Λ et si p est le cardinal de Λ , alors

$$h_\mu(\sigma) = \log p .$$

Nous reviendrons sur cet exemple (pour le généraliser) au paragraphe 11.2. Pour d'autres calculs, voir les exercices.

9.5 L'entropie comme fonction de la mesure

Lemme 9.17 *Soient $(X, \mathcal{B}, \mu)$ un espace de probabilité, α une partition mesurable de X , ν une mesure de probabilité sur $(X, \mathcal{B})$ et $s, t \geq 0$ avec $s + t = 1$. Alors*

$$sH_\mu(\alpha) + tH_\nu(\alpha) \leq H_{s\mu+t\nu}(\alpha) \leq sH_\mu(\alpha) + tH_\nu(\alpha) + \log 2 .$$

Preuve. Pour A dans α , on note $\Delta_A = \varphi(s\mu(A) + t\nu(A)) - s\varphi(\mu(A)) - t\varphi(\nu(A))$. Par concavité de φ , on a l'inégalité $\Delta_A \geq 0$. D'autre part, on a

$$\Delta_A = -s\mu(A) (\log(s\mu(A) + t\nu(A)) - \log(s\mu(A)))$$

$$-t\nu(A) (\log(s\mu(A) + t\nu(A)) - \log(t\nu(A))) - s\mu(A) \log s - t\nu(A) \log t .$$

Les deux premiers termes sont négatifs, donc on a $\Delta_A \leq -s\mu(A) \log s - t\nu(A) \log t$. On en déduit que

$$0 \leq \sum_{A \in \alpha} \Delta_A \leq -s \log s - t \log t \leq \log 2 .$$

C'est ce que l'on voulait. □

La proposition suivante signifie que l'entropie $h_\mu(f)$ dépend de façon affine de μ . Mais attention, elle ne dépend en général pas de façon continue de μ quand on munit l'espace $\mathcal{M}(X)$ des mesures de probabilités sur X de la topologie faible (au sens du paragraphe 4.3).

Proposition 9.18 *Soient $(X, \mathcal{B}, \mu, f)$ un système dynamique probabilisé, α une partition mesurable de X , ν une mesure de probabilité f -invariante sur $(X, \mathcal{B})$ et $s, t \geq 0$ avec $s + t = 1$. Alors on a*

$$1. \ h_{s\mu+t\nu}(f, \alpha) = sh_\mu(f, \alpha) + th_\nu(f, \alpha),$$

$$2. \ h_{s\mu+t\nu}(f) = sh_\mu(f) + th_\nu(f).$$

Preuve. (1) On applique le lemme à la partition $\bigvee_{i=0}^{n-1} f^{-i}\alpha$, on divise par n et on prend la limite pour n tendant vers $+\infty$. Le terme $\frac{1}{n} \log 2$ tend vers 0, ce qui donne l'égalité cherchée.

(2) On note $\alpha_n, \beta_n, \gamma_n$ des partitions mesurables de X telles que

$$\lim_{n \rightarrow \infty} h_\mu(f, \alpha_n) = h_\mu(f), \quad \lim_{n \rightarrow \infty} h_\nu(f, \beta_n) = h_\nu(f), \quad \lim_{n \rightarrow \infty} h_{s\mu+t\nu}(f, \gamma_n) = h_{s\mu+t\nu}(f) .$$

On peut, sans changer ces égalités, remplacer ces partitions par des partitions plus fines comme $\alpha_n \vee \beta_n \vee \gamma_n$. On peut donc supposer que $\alpha_n = \beta_n = \gamma_n$. L'assertion se déduit alors par passage à la limite de celle de (1). □

9.6 Exercices

Exercice E.9.1 Soit $(X, \mathcal{B}, \mu)$ un espace de probabilité, et $g : X \rightarrow X$ une application mesurable. Montrer que $H_{g_*\mu}(\xi) = H_\mu(g^{-1}\xi)$ pour toute partition mesurable ξ de X .

Exercice E.9.2 Pour N dans $\mathbb{N}$, on considère la transformation $f_N : \theta \mapsto N\theta$ du cercle $\mathbb{S}^1 = \mathbb{R}/\mathbb{Z}$. Calculer l'entropie métrique de f_N pour la mesure de Lebesgue.

Exercice E.9.3 On considère la rotation $f : \theta \mapsto \theta + \kappa$ du cercle $\mathbb{S}^1 = \mathbb{R}/\mathbb{Z}$, d'angle κ . Calculer l'entropie métrique de f pour la mesure de Lebesgue.

Exercice E.9.4 Soient $(X, \mathcal{B}, \mu)$ un espace de probabilité et α, β deux partitions mesurables de X . Montrer que les assertions suivantes sont équivalentes. On dit alors que α et β sont *indépendantes*.

1. $\forall A \in \alpha, \forall B \in \beta, \quad \mu(A \cap B) = \mu(A)\mu(B).$
2. $I_{\alpha \vee \beta} = I_\alpha + I_\beta.$
3. $H(\alpha \vee \beta) = H(\alpha) + H(\beta).$
4. $I_{\alpha|\beta} = I_\alpha.$
5. $H(\alpha|\beta) = H(\alpha).$

Exercice E.9.5 Soient $(X, \mathcal{B}, \mu)$ un espace de probabilité et α, β deux partitions mesurables de X . On dit que $\alpha = \beta$ μ -presque sûrement si chaque partie de α de mesure non nulle est égale μ -presque partout à une partie de β . On dit que $\alpha \preceq \beta$ μ -presque sûrement si chaque partie de α de mesure non nulle est égale μ -presque partout à une réunion de parties de β .

Montrer que l'égalité presque sûre est une relation d'équivalence sur l'ensemble des partitions mesurables de X .

Montrer les équivalences

$$\alpha \preceq \beta \quad \mu\text{-ps} \Leftrightarrow \alpha \vee \beta = \beta \quad \mu\text{-ps} \Leftrightarrow I_{\alpha|\beta} = 0 \quad \mu\text{-ps} \Leftrightarrow H(\alpha|\beta) = 0 .$$

Exercice E.9.6 Soient $(X, \mathcal{B}, \mu, f)$ un système dynamique probabilisé inversible et α une partition mesurable de X . Montrer que $h_\mu(f^{-1}, \alpha) = h_\mu(f, \alpha)$. En déduire que

$$h_\mu(f^{-1}) = h_\mu(f) .$$

Exercice E.9.7 Soient $(X, \mathcal{B}, \mu, f), (Y, \mathcal{C}, \nu, g)$ deux systèmes dynamiques probabilisés et $(X \times Y, \mathcal{B} \otimes \mathcal{C}, \mu \otimes \nu, f \times g)$ le système dynamique probabilisé produit. Montrer que

$$h_{\mu \otimes \nu}(f \times g) = h_\mu(f) + h_\nu(g) .$$

Exercice E.9.8 Soient $(X, \mathcal{B}, \mu, f)$ un système dynamique probabilisé inversible et α une partition mesurable de X , telle que la suite de partitions $\alpha_n = \bigvee_{i=0}^n f^{-i}\alpha$ engendre la σ -algèbre $\mathcal{B}$. Montrer que

$$h_\mu(f) = 0 .$$

9.7 Indications pour la résolution des exercices

Exercice E.9.2 Si $N \geq 2$, en considérant l'écriture N -adique, le système dynamique probabilisé $(\mathbb{S}^1, d\theta, f_N)$ est conjugué (au sens de la mesure) au décalage sur $\Lambda^{\mathbb{N}}$ pour la mesure produit, avec $\Lambda = \{0, \dots, N-1\}$ muni de l'équiprobabilité (voir exercice E.2.5). Par invariance de l'entropie métrique, et l'exemple à la fin du paragraphe 9.4, on a

$$h_{d\theta}(f_N) = \log N .$$

Exercice E.9.3 Si $\kappa = \frac{p}{q}$ est rationnel, alors, pour toute partition mesurable β , on a

$$h(f, \beta) = \lim_{n \rightarrow \infty} \frac{1}{qn} H \left(\bigvee_{i=0}^{nq-1} f^{-i} \beta \right) = \lim_{n \rightarrow \infty} \frac{1}{qn} H \left(\bigvee_{i=0}^{q-1} f^{-i} \beta \right) = 0 .$$

Donc $h_{d\theta}(f) = 0$.

Si κ est irrationnel, alors montrer que la partition mesurable $\alpha = \{[0, \frac{1}{2}[, [\frac{1}{2}, 0[\}$ est génératrice et même que la suite de partitions $\bigvee_{i=1}^n f^{-i} \alpha$ engendre la σ -algèbre des boréliens du cercle, et utiliser le lemme 9.13, pour montrer que $h_{d\theta}(f) = h_{d\theta}(f, \alpha) = 0$.

Exercice E.9.4 Pour montrer que l'assertion (5) implique l'assertion (1), on étudiera le cas d'égalité dans la démonstration de la proposition 9.3 (2) de monotonie de l'entropie.

Exercice E.9.7 Pour toute partition mesurable α, β de X, Y respectivement, on notera $\alpha \times \beta = \{A \times B \mid A \in \alpha, B \in \beta\}$ et on vérifiera que $H_{\mu \otimes \nu}(\alpha \times \beta) = H_{\mu}(\alpha) + H_{\nu}(\beta)$.

Exercice E.9.8 On remarquera que la suite de partitions $\bigvee_{i=1}^n f^{-i} \alpha$ engendre aussi $\mathcal{B}$, et on étudiera la suite $H(\alpha \mid \bigvee_{i=1}^n f^{-i} \alpha)$.

10 Entropie topologique

L'entropie topologique mesure la complexité d'un système dynamique topologique. Si on compte le nombre minimum d'orbites de longueur n qui permet de les "pister" toutes à ϵ -près, on trouve une fonction de n . L'entropie topologique est, grosso modo, le taux de croissance exponentiel de cette fonction de n lorsque ϵ est très petit.

Le but de ce chapitre est de définir précisément l'entropie topologique et de montrer qu'elle est égale à la borne supérieure des entropies métriques pour toutes les mesures de probabilités boréliennes invariantes.

Nous calculerons l'entropie topologique de quelques exemples dans les deux chapitres 11 et 12 suivants.

10.1 Recouvrements ouverts

Dans la cadre topologique, les partitions mesurables sont remplacées par les recouvrements ouverts.

Soit X un espace métrisable compact. Pour tout recouvrement ouvert $\mathcal{U}$ de X , on note $N(\mathcal{U})$ le cardinal minimum des sous-recouvrements finis de $\mathcal{U}$. (L'existence de cet entier $N(\mathcal{U})$ résulte de la compacité de X .) Pour toute application continue $f : X \rightarrow X$, on note $f^{-1}\mathcal{U}$ le recouvrement ouvert $\{f^{-1}(U)\}_{U \in \mathcal{U}}$ de X .

Soit $\mathcal{V}$ un (autre) recouvrement ouvert de X . On note $\mathcal{U} \vee \mathcal{V}$ le *joint* de $\mathcal{U}$ et $\mathcal{V}$

$$\mathcal{U} \vee \mathcal{V} = \{U \cap V / U \in \mathcal{U}, V \in \mathcal{V}\}.$$

On note $\bigvee_{i=1}^n \mathcal{U}_i$ pour $\mathcal{U}_1 \vee \dots \vee \mathcal{U}_n$. On dit que $\mathcal{U}$ est *moins fin* que $\mathcal{V}$, et on note $\mathcal{U} \preceq \mathcal{V}$, si tout ouvert de $\mathcal{V}$ est contenu dans un ouvert de $\mathcal{U}$.

Voici quelques propriétés de cet entier $N(\mathcal{U})$.

Lemme 10.1 *i) On a $N(f^{-1}\mathcal{U}) \leq N(\mathcal{U})$ avec égalité lorsque f est surjective.*

ii) $N(\mathcal{U} \vee \mathcal{V}) \leq N(\mathcal{U})N(\mathcal{V})$.

iii) Si $\mathcal{U} \preceq \mathcal{V}$, alors $N(\mathcal{U}) \leq N(\mathcal{V})$.

Preuve. i) Si les ouverts $U_1, \dots, U_n$ de $\mathcal{U}$ recouvrent X , alors $f^{-1}(U_1), \dots, f^{-1}(U_n)$ recouvrent aussi X . La réciproque est vraie lorsque f est surjective.

ii) Si les ouverts $U_1, \dots, U_n$ de $\mathcal{U}$ recouvrent X et si les ouverts $V_1, \dots, V_p$ de $\mathcal{V}$ recouvrent X , alors les ouverts $U_i \cap V_j$ pour $i = 1, \dots, n$ et $j = 1, \dots, p$ recouvrent aussi X .

iii) Si les ouverts $V_1, \dots, V_p$ de $\mathcal{V}$ recouvrent X , on choisit pour tout i un ouvert U_i de $\mathcal{U}$ contenant V_i . Alors les ouverts $U_1, \dots, U_p$ de $\mathcal{U}$ recouvrent aussi X □

Soit maintenant (X, d) un espace métrique compact. L'exemple le plus simple de recouvrement de X est le recouvrement $\mathcal{O}_\epsilon$ par les boules

$$B(x, \epsilon) = \{y \in X / d(x, y) < \epsilon\}$$

où x décrit X . On dit qu'une partie F de X est ϵ -dense (pour d) si on a

$$X = \bigcup_{x \in F} B(x, \epsilon).$$

On dit que F est ϵ -séparé (pour d) si

$$\inf\{d(x, y) \mid x, y \in F, x \neq y\} \geq \epsilon .$$

On note $N_d(\epsilon)$ la borne inférieure des cardinaux des parties ϵ -denses de X , et $N'_d(\epsilon)$ la borne supérieure des cardinaux des parties ϵ -séparées de X . Ces deux nombres sont finis, par compacité de X . On a bien sûr $N_d(\epsilon) = N(\mathcal{O}_\epsilon)$.

Lemme 10.2 *Pour tout $\epsilon > 0$, on a les inégalités*

$$N'_d(2\epsilon) \leq N_d(\epsilon) \leq N'_d(\epsilon) .$$

Preuve. En effet, d'une part les points d'une partie 2ϵ -séparée sont dans des boules de rayon ϵ distinctes. D'autre part, une partie ϵ -séparée de cardinal maximal est forcément ϵ -dense. $\square$

10.2 Entropie topologique : définition

Soit X un espace topologique compact et $f : X \rightarrow X$ une application continue.

Lemme 10.3 *Soit $\mathcal{U}$ un recouvrement ouvert de X . Alors la suite des $\frac{1}{n} \log N(\bigvee_{i=0}^{n-1} f^{-i}\mathcal{U})$ admet une limite finie quand n tends vers $+\infty$.*

On appelle cette limite l'*entropie* de f pour le recouvrement ouvert $\mathcal{U}$, et on la note

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

Preuve. La suite $a_n = \log N(\bigvee_{i=0}^{n-1} f^{-i}\mathcal{U})$ est sous-additive, car par le lemme 10.1, on a

$$a_{n+m} \leq \log \left(\bigvee_{i=0}^{n-1} f^{-i}\mathcal{U} \right) + \log \left(\bigvee_{i=n}^{n+m-1} f^{-i}\mathcal{U} \right) \leq a_n + a_m .$$

$\square$

Remarque. Si $\mathcal{U} \preceq \mathcal{V}$, alors $h(f, \mathcal{U}) \leq h(f, \mathcal{V})$. Ceci motive la définition suivante.

On appelle *entropie topologique* de f , et on note $h_{\text{top}}(f)$ l'élément de $[0, +\infty]$ suivant :

$$h_{\text{top}}(f) = \sup_{\mathcal{U}} h(f, \mathcal{U}) ,$$

où la borne supérieure est prise sur tous les recouvrements ouverts $\mathcal{U}$ de X .

La proposition suivante dit que l'entropie topologique est un "invariant" des systèmes dynamiques topologiques.

Proposition 10.4 *Soient $(X, f), (Y, g)$ des systèmes dynamiques topologiques (topologiquement) conjugués, avec X, Y métrisables compacts. Alors*

$$h_{\text{top}}(f) = h_{\text{top}}(g) .$$

Preuve. Soit $h : X \rightarrow Y$ un homéomorphisme tel que $h \circ f = g \circ h$. Pour tout recouvrement ouvert $\mathcal{V}$ de Y , le recouvrement ouvert $h^{-1}\mathcal{V} = \{h^{-1}(V) \mid V \in \mathcal{V}\}$ vérifie $N(h^{-1}\mathcal{V}) = N(\mathcal{V})$. Donc $h(g, \mathcal{V}) = h(f, h^{-1}\mathcal{V})$. D'où $h_{\text{top}}(g) \leq h_{\text{top}}(f)$, et par symétrie, le résultat en découle. $\square$

Soit d une distance sur X compatible avec la topologie de X . La proposition suivante fournit une définition plus intuitive de l'entropie topologique de f : elle affirme que l'on peut se contenter des recouvrements $\mathcal{O}_\epsilon$.

Notons d_n^f la distance sur X définie par

$$d_n^f(x, y) = \sup_{i=0, \dots, n-1} d(f^i(x), f^i(y)) .$$

L'entier $N_{d_n^f}(\epsilon)$ est "le nombre minimum d'orbites de longueur n permettant de suivre ϵ près n'importe quelle orbite de longueur n ". L'entier $N'_{d_n^f}(\epsilon)$ est "le nombre maximum d'orbites de longueur n telles que deux d'entre elles ne se suivent pas à ϵ près". On a l'égalité

$$N_{d_n^f}(\epsilon) = N \left(\bigvee_{i=0}^{n-1} f^{-i} \mathcal{O}_\epsilon \right) .$$

Proposition 10.5 *On a*

$$h_{\text{top}}(f) = \lim_{\epsilon \rightarrow 0} \left(\lim_{n \rightarrow \infty} \frac{1}{n} \log N_{d_n^f}(\epsilon) \right) = \lim_{\epsilon \rightarrow 0} \left(\limsup_{n \rightarrow \infty} \frac{1}{n} \log N'_{d_n^f}(\epsilon) \right) .$$

Remarque. En particulier, ces limites restent inchangées si on remplace la distance d par une autre distance qui induit la même topologie.

Preuve. La première égalité résulte du lemme suivant, et la seconde du lemme 10.2. $\square$

Lemme 10.6 (Lemme de Lebesgue) *Pour tout recouvrement ouvert $\mathcal{U}$ d'un espace métrique compact X , il existe $\epsilon > 0$ tel que $\mathcal{U}$ est moins fin que $\mathcal{O}_\epsilon$.*

Preuve. Pour tout x dans X , il existe une boule $B(x, 2\epsilon_x)$ contenue dans l'un des ouverts de $\mathcal{U}$. On extrait du recouvrement par les boules $B(x, \epsilon_x)$ un sous-recouvrement fini $\{B(x_i, \epsilon_{x_i})\}_i$, et on prend $\epsilon = \inf_i \epsilon_{x_i}$. $\square$

Si $\mathcal{U}$ est un recouvrement ouvert de X , notons $\text{diam}(\mathcal{U})$ la borne supérieure des diamètres des ouverts U de $\mathcal{U}$.

Corollaire 10.7 *Soit $\mathcal{U}_n$ une suite de recouvrements ouverts de X . Si $\lim_{n \rightarrow \infty} \text{diam}(\mathcal{U}_n) = 0$, alors*

$$h_{\text{top}}(f) = \lim_{n \rightarrow \infty} h(f, \mathcal{U}_n) .$$

Preuve. En effet, pour tout $\epsilon > 0$, le recouvrement $\mathcal{U}_n$ est plus fin que $\mathcal{O}_\epsilon$ dès que $\text{diam}(\mathcal{U}_n) < \epsilon$. $\square$

10.3 Entropie topologique : propriétés

Soient X un espace topologique compact, $f : X \rightarrow X$ une application continue et $\mathcal{U}$ un recouvrement ouvert de X .

La proposition suivante est l'analogie topologique des propositions 9.14 et 9.15.

Proposition 10.8 1. Pour tout $k \geq 1$, on a $h(f, \mathcal{U}) = h(f, \bigvee_{i=0}^k f^{-i}\mathcal{U})$, et, si f est inversible, alors $h(f, \mathcal{U}) = h(f, \bigvee_{i=-k}^k f^{-i}\mathcal{U})$.

2. Pour tout $k \geq 1$, on a $h_{\text{top}}(f^k) = k h_{\text{top}}(f)$.

Preuve. La preuve est analogue à celle des propositions 9.14 et 9.15.

(1) Notons $a_n(\mathcal{U}) = \log(N(\bigvee_{i=0}^n f^{-i}\mathcal{U}))$. On a $a_n(\bigvee_{i=0}^k f^{-i}\mathcal{U}) = a_{n+k}(\mathcal{U})$, et lorsque f est inversible, $a_n(\bigvee_{i=-k}^k f^{-i}\mathcal{U}) = a_{n+2k-1}(\mathcal{U})$. Le résultat s'en déduit, car les suites $\frac{a_n(\mathcal{U})}{n}$ et $\frac{a_{n+k}(\mathcal{U})}{n}$ ont la même limite quand n tend vers $+\infty$.

(2) On a

$$h(f^k, \mathcal{U}) \leq h(f^k, \bigvee_{i=0}^k \mathcal{U}) = \lim_{n \rightarrow \infty} \frac{1}{n} a_{nk}(\mathcal{U}) = k h(f, \mathcal{U}).$$

Ceci prouve que

$$h(f^k, \mathcal{U}) \leq k h_{\text{top}}(f),$$

mais aussi que $k h_{\text{top}}(f) \leq h_{\text{top}}(f^k)$. Donc $h_{\text{top}}(f^k) = k h_{\text{top}}(f)$. $\square$

Le résultat suivant est analogue au théorème de Kolmogorov-Sinai, il permet d'éviter de prendre la borne supérieure dans la définition de l'entropie topologique.

Un recouvrement ouvert $\mathcal{U}$ de X est dit *générateur* pour f si $\lim_{n \rightarrow \infty} \text{diam}(\bigvee_{i=0}^n f^{-i}\mathcal{U}) = 0$, ou si f est inversible et si $\lim_{n \rightarrow \infty} \text{diam}(\bigvee_{i=-n}^n f^{-i}\mathcal{U}) = 0$.

Théorème 10.9 Si $\mathcal{U}$ est générateur pour f , alors $h_{\text{top}}(f) = h(f, \mathcal{U})$.

Preuve. D'après la proposition 10.8, la suite $h(f, \bigvee_{i=0}^n f^{-i}\mathcal{U})$ est constante égale à $h(f, \mathcal{U})$. D'après le corollaire 10.7, cette suite converge vers $h_{\text{top}}(f)$. On a donc $h_{\text{top}}(f) = h(f, \mathcal{U})$. On procède de même lorsque f est inversible, avec la suite $h(f, \bigvee_{i=-n}^n f^{-i}\mathcal{U})$. $\square$

Soit maintenant (X, d) un espace métrique compact. Une application continue $f : X \rightarrow X$ est dite *expansive* s'il existe $\epsilon > 0$ tel que, si $x \neq y$, alors

$$\sup_{n \in \mathbb{N}} d(f^n(x), f^n(y)) \geq \epsilon,$$

ou si f est inversible et s'il existe $\epsilon > 0$ tel que, si $x \neq y$, alors

$$\sup_{n \in \mathbb{Z}} d(f^n(x), f^n(y)) \geq \epsilon.$$

Exemple. Le décalage σ d'un système symbolique est expansif.

Remarque. Une application continue $f : X \rightarrow X$ est expansive si et seulement s'il existe $\epsilon_0 > 0$ tel que le recouvrement ouvert $\mathcal{O}_{\epsilon_0}$ de X est générateur pour f .

Tout ϵ_0 comme ci-dessus est appelé une *constante d'expansivité* de f .

Corollaire 10.10 *Pour toute application continue expansive $f : X \rightarrow X$, on a $h_{\text{top}}(f) < \infty$. De plus, si ϵ_0 est une constante d'expansivité de f , alors, pour tout $0 < \epsilon < \epsilon_0$, on a*

$$h_{\text{top}}(f) = \lim_{n \rightarrow \infty} \frac{1}{n} \log N_{d_n^f}(\epsilon) = \limsup_{n \rightarrow \infty} \frac{1}{n} \log N'_{d_n^f}(\epsilon) .$$

Preuve. En effet, $h_{\text{top}}(f) = h(f, \mathcal{O}_{\epsilon_0})$ (et on utilise l'égalité précédant la proposition 10.5). $\square$

10.4 Le principe variationnel : première inégalité

Théorème 10.11 (Principe variationnel) *Soient X un espace métrique compact et $f : X \rightarrow X$ une application continue. Alors on a l'égalité*

$$h_{\text{top}}(f) = \sup_{\mu \in \mathcal{M}(X)^f} h_{\mu}(f) .$$

Rappelons que $\mathcal{M}(X)^f$ est l'espace des mesures de probabilités boréliennes f -invariantes sur X . Nous montrerons l'inégalité $\geq$ dans cette partie, et l'inégalité $\leq$ dans la suivante.

Nous aurons besoin du lemme suivant, qui est un des points clefs de la preuve du théorème de Riesz.

Lemme 10.12 *Soient X un espace métrisable localement compact séparable, μ une mesure de probabilité borélienne sur X et A un borélien de X . Alors, pour tout $\epsilon > 0$, il existe un compact K de X et un ouvert U de X tels que $K \subset A \subset U$ et $\mu(U - K) \leq \epsilon$.*

Preuve. On vérifie, comme au lemme 9.8, que l'ensemble $\mathcal{B}'$ des boréliens A de X qui satisfont cette propriété est une σ -algèbre, qui contient les ouverts de X . $\square$

Preuve de la majoration $h_{\mu}(f) \leq h_{\text{top}}(f)$ pour tout μ dans $\mathcal{M}(X)^f$.

Soit $\alpha = \{A_1, \dots, A_p\}$ une partition mesurable de X . D'après le lemme, il existe des compacts disjoints $K_1, \dots, K_p$ tels que le nombre réel $\epsilon = \sup_i \mu(A_i \Delta K_i)$ est aussi petit que l'on veut. On pose $U_0 = X - (K_1 \cup \dots \cup K_p)$, et on introduit la partition $\beta = \{K_1, \dots, K_p, U_0\}$. D'après le lemme 9.9, on peut s'arranger pour que $H(\alpha|\beta) \leq 1$. L'inégalité de Rokhlin (proposition 9.14 (2)) assure que $h_{\mu}(f, \alpha) \leq h_{\mu}(f, \beta) + 1$. Introduisons maintenant le recouvrement ouvert $\mathcal{U} = \{U_0 \cup K_1, \dots, U_0 \cup K_p\}$ de X . D'après l'assertion (2) de la dernière remarque du paragraphe 9.1, on a la majoration

$$H_{\mu} \left(\bigvee_{i=0}^{n-1} f^{-i} \beta \right) \leq \log \left(\text{Card} \left(\bigvee_{i=0}^{n-1} f^{-i} \beta \right) \right) \leq \log \left(2^n N \left(\bigvee_{i=0}^{n-1} f^{-i} \mathcal{U} \right) \right) .$$

(En effet, notons $K_0 = U_0$ et $K_{i_0, \dots, i_{n-1}} = K_{i_0} \cap f^{-1}(K_{i_1}) \cap \dots \cap f^{-1}(K_{i_{n-1}})$. Alors les $K_{i_0, \dots, i_{n-1}}$ sont deux à deux disjoints et

$$\text{Card} \left(\bigvee_{i=0}^{n-1} f^{-i} \beta \right) = \text{Card} \{ (i_0, \dots, i_{n-1}) \in \{0, \dots, p\}^n / K_{i_0, \dots, i_{n-1}} \neq \emptyset \} .$$

La dernière inégalité ci-dessus découle alors du fait que $K_{i_0, \dots, i_{n-1}}$ est contenu dans

$$(U_0 \cup K_{\text{sup}\{i_0, 1\}}) \cap f^{-1}(U_0 \cup K_{\text{sup}\{i_1, 1\}}) \cap \dots \cap f^{-1}(U_0 \cup K_{\text{sup}\{i_{n-1}, 1\}}) ,$$

ce qui donne 2^n choix possibles d'un élément de $\bigvee_{i=0}^{n-1} f^{-i}\mathcal{U}$ pour tout élément de $\bigvee_{i=0}^{n-1} f^{-i}\beta$.)

En divisant par n et en passant à la limite, on obtient

$$h_\mu(f, \alpha) \leq h_\mu(f, \beta) + 1 \leq h(f, \mathcal{U}) + \log 2 + 1 \leq h_{\text{top}}(f) + \log 2 + 1.$$

On en déduit la majoration $h_\mu(f) \leq h_{\text{top}}(f) + \log 2 + 1$. En remplaçant dans cette égalité f par une puissance f^n , en utilisant les propositions 9.15 et 10.8, et en divisant par n , on obtient $h_\mu(f) \leq h_{\text{top}}(f) + \frac{\log 2 + 1}{n}$. Par conséquent, on a $h_\mu(f) \leq h_{\text{top}}(f)$. $\square$

10.5 Le principe variationnel : seconde inégalité

Commençons la démonstration de l'autre inégalité par quelques lemmes préliminaires.

Lemme 10.13 *Soient X un espace métrisable compact, $(\mu_n)_{n \in \mathbb{N}}$ une suite de mesures de probabilité boréliennes sur X qui convergent faiblement vers une mesure de probabilité μ et A un borélien de X , tel que $\mu(\partial A) = 0$. Alors*

$$\lim_{n \rightarrow \infty} \mu_n(A) = \mu(A).$$

Preuve. Comme l'adhérence $\overline{A}$ de A est fermée, la fonction caractéristique $\mathbb{I}_{\overline{A}}$ est une limite d'une suite décroissante de fonctions continues f_k : on peut par exemple prendre $f_k(x) = \max\{1 - k d(x, \overline{A}), 0\}$. Les égalités $\lim_{n \rightarrow \infty} \mu_n(f_k) = \mu(f_k)$ et $\lim_{k \rightarrow \infty} \mu(f_k) = \mu(\overline{A})$ (et l'inégalité $\mu_n(\overline{A}) \leq \mu_n(f_k)$) impliquent que

$$\limsup_{n \rightarrow \infty} \mu_n(\overline{A}) \leq \mu(\overline{A}).$$

De même, l'intérieur $\overset{\circ}{A}$ de A vérifie

$$\liminf_{n \rightarrow \infty} \mu_n(\overset{\circ}{A}) \geq \mu(\overset{\circ}{A}).$$

L'hypothèse assure que $\mu(\overline{A}) = \mu(\overset{\circ}{A}) = \mu(A)$. Donc $\lim_{n \rightarrow \infty} \mu_n(A) = \mu(A)$. $\square$

Lemme 10.14 *Soient X un espace métrisable compact, μ une mesure de probabilité borélienne sur X et $\epsilon > 0$. Alors il existe une partition mesurable α de X telle que, pour tout A dans α , on a $\text{diam}(A) \leq \epsilon$ et $\mu(\partial A) = 0$.*

Preuve. Pour tout point x de X , il existe un réel ϵ_x dans $]0, \frac{\epsilon}{2}[$ tel que la sphère de centre x et de rayon ϵ_x soit de mesure nulle pour μ . On peut extraire du recouvrement de X par les boules $B(x, \epsilon_x)$ un sous-recouvrement fini $\{B_1, \dots, B_p\}$. Par construction, on a $\mu(\partial B_i) = 0$. On construit $\alpha = \{A_1, \dots, A_p\}$ avec $A_i = B_i - (B_1 \cup \dots \cup B_{i-1})$. On a alors $\partial A_i \subset \partial B_1 \cup \dots \cup \partial B_p$, donc $\mu(\partial A_i) = 0$. $\square$

Le lemme crucial est le suivant. Il permet de construire des mesures de probabilités boréliennes f -invariantes de grande entropie. On rappelle que δ_x désigne la masse de Dirac en x .

Lemme 10.15 *Soient X un espace métrisable compact et $f : X \rightarrow X$ une application continue. Pour tout $n \geq 1$, on choisit une partie E_n de X qui est ϵ -séparée pour la distance*

d_n^f , et on note $\nu_n = \frac{1}{\text{Card } E_n} \sum_{x \in E_n} \delta_x$ et $\mu_n = \frac{1}{n} \sum_{i=0}^{n-1} (f^i)_*(\nu_n)$. Alors la suite μ_n admet une sous-suite faiblement convergente vers une limite μ . De plus, une telle mesure de probabilité μ est f -invariante et on a

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log(\text{Card } E_n) \leq h_\mu(f) .$$

Preuve. L'existence de μ résulte de la compacité faible de $\mathcal{M}(X)$ (voir la proposition 4.10). La f -invariance de μ résulte d'un passage à la limite dans l'égalité

$$f_*\mu_n - \mu_n = \frac{1}{n}((f^n)_*\nu_n - \nu_n) .$$

D'après le lemme 10.14, il existe une partition mesurable α de X telle que, pour tout A dans α , on a $\text{diam}(A) \leq \epsilon$ et $\mu(\partial A) = 0$. Posons $\alpha_n = \bigvee_{i=0}^{n-1} f^{-i}\alpha$. Comme chaque partie de α_n contient au plus un point de E_n , on a l'égalité $H_{\nu_n}(\alpha_n) = \log(\text{Card } E_n)$.

Soient $0 \leq k < q < n$, alors

$$\alpha_n = \left(\bigvee_{j=0}^{E[\frac{n}{q}]-1} f^{-jq-k}\alpha_q \right) \vee \left(\bigvee_{j=0}^{k-1} f^{-j}\alpha \right) \vee \left(\bigvee_{j=qE[\frac{n}{q}]-q+k+1}^{n-1} f^{-j}\alpha \right) .$$

Par sous-additivité de l'entropie, on a alors

$$\log(\text{Card } E_n) = H_{\nu_n}(\alpha_n) \leq \sum_{j=0}^{E[\frac{n}{q}]-1} H_{\nu_n}(f^{-k-qj}\alpha_q) + 2q \log(\text{Card } \alpha) .$$

En sommant sur k , puis en appliquant la concavité de l'entropie d'une partition en fonction de la mesure (lemme 9.17), on obtient

$$q \log(\text{Card } E_n) \leq \sum_{i=0}^{n-1} H_{\nu_n}(f^{-i}\alpha_q) + 2q^2 \log(\text{Card } \alpha) \leq n H_{\mu_n}(\alpha_q) + 2q^2 \log(\text{Card } \alpha) .$$

En divisant par nq et en faisant tendre n vers l'infini, on obtient

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log(\text{Card } E_n) \leq \frac{1}{q} \lim_{n \rightarrow \infty} H_{\mu_n}(\alpha_q) = \frac{1}{q} H_\mu(\alpha_q)$$

par le lemme 10.13. On fait alors tendre q vers l'infini, ce qui donne

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log(\text{Card } E_n) \leq h_\mu(f, \alpha) \leq h_\mu(f) .$$

C'est ce que l'on voulait. □

Fin de la preuve du théorème 10.11 : Appliquons le lemme précédent avec E_n de cardinal maximal. Pour tout $\epsilon > 0$, on peut trouver un élément μ dans $\mathcal{M}(X)^f$ telle que $h_\mu(f) \geq \limsup_{n \rightarrow \infty} \frac{1}{n} \log N'_{d_n^f}(\epsilon)$. D'après la proposition 10.5, le membre de droite tend vers $h_{\text{top}}(f)$ quand ϵ tend vers 0. Donc $\sup_{\mu \in \mathcal{M}(X)^f} h_\mu(f) \geq h_{\text{top}}(f)$. □

Le corollaire suivant affirme que, dans certain cas, cette borne supérieure est atteinte.

Corollaire 10.16 *Soit X un espace métrique compact et $f : X \rightarrow X$ une application continue expansive. Alors il existe une mesure de probabilité borélienne f -invariante μ sur X telle que $h_\mu(f) = h_{\text{top}}(f)$.*

Une telle mesure s'appelle une *mesure (de probabilité) d'entropie maximale*.

Preuve. D'après le paragraphe 10.3, il existe $\epsilon_0 > 0$ tel que $h_{\text{top}}(f) = h(f, \mathcal{O}_{\epsilon_0})$. Par le lemme 10.2, on a $h(f, \mathcal{O}_{\epsilon_0}) = \limsup_{n \rightarrow \infty} \frac{1}{n} \log N'_{d_n}(f)(\epsilon_0)$. Par le lemme 10.15, il existe donc un élément μ dans $\mathcal{M}(X)^f$ tel que $h_\mu(f) \geq h_{\text{top}}(f)$. $\square$

Par exemple la mesure de Lebesgue sur le cercle est l'unique mesure d'entropie maximale pour l'application de doublement de l'angle (voir l'exercice E.10.10). Voir le paragraphe 10.3) pour une preuve qu'un système dynamique symbolique sur un alphabet fini admet une unique mesure d'entropie maximale. Voir la proposition 12.7 pour une preuve que la mesure de Haar sur le tore est l'unique mesure d'entropie maximale pour un automorphisme linéaire hyperbolique du tore.

10.6 Exercices

Exercice E.10.1 Soient (X, d) un espace métrique compact et $f : X \rightarrow X$ une application non dilatante (i.e. telle que, pour tous x, y dans X , on a $d(f(x), f(y)) \leq d(x, y)$). Montrer que $h_{\text{top}}(f) = 0$.

Exercice E.10.2 Soit (X, d) un espace métrique compact. On note

$$D(X) = \limsup_{\epsilon \rightarrow 0} \frac{\log N_d(\epsilon)}{\log \epsilon}.$$

- a) Montrer que, si X est un compact de $\mathbb{R}^m$ muni de la distance euclidienne, alors on a $D(X) \leq m$.
- b) On suppose que $D(X) < +\infty$. Soient $C > 0$ et $f : X \rightarrow X$ une application C -lipschitzienne. Montrer que $h_{\text{top}}(f) \leq D(X) \log(C)$.

Exercice E.10.3 Soit f un homéomorphisme d'un espace métrique compact X . Montrer que $h_{\text{top}}(f^{-1}) = h_{\text{top}}(f)$.

Exercice E.10.4 Soient f un homéomorphisme d'un espace métrique compact X , et $\mathcal{U}$ un recouvrement ouvert de X tel que

$$\lim_{n \rightarrow \infty} \text{diam} \left(\bigvee_{i=0}^n f^{-i} \mathcal{U} \right) = 0.$$

Montrer que X est un ensemble fini.

Exercice E.10.5 Soient X un espace métrique compact, $f : X \rightarrow X$ une application continue et Y un fermé de X tel que $f(Y) \subset Y$. Montrer que $h_{\text{top}}(f|_Y) = h_{\text{top}}(f)$.

Exercice E.10.6 Soient X, Y des espaces métriques compacts, $f : X \rightarrow X, g : Y \rightarrow Y, F : X \rightarrow Y$ des applications continues telles que $F \circ f = g \circ F$. Montrer que $h_{\text{top}}(g) \leq h_{\text{top}}(f)$.

Exercice E.10.7 Soient X un espace métrique compact et $\mathcal{U}, \mathcal{U}', \mathcal{V}, \mathcal{V}'$ des recouvrements ouverts de X . Par analogie avec l'entropie relative des partitions mesurables, on pose

$$N(\mathcal{U}|\mathcal{V}) = \frac{N(\mathcal{U} \vee \mathcal{V})}{N(\mathcal{V})}.$$

1. Montrer que si $\mathcal{U} \preceq \mathcal{U}'$, alors $N(\mathcal{U}|\mathcal{V}) \leq N(\mathcal{U}'|\mathcal{V})$.

2. Est-il vrai que si $V \preceq \mathcal{V}'$, alors $N(\mathcal{U}|\mathcal{V}) \geq N(\mathcal{U}|\mathcal{V}')$?

Exercice E.10.8 Soient X un espace métrique compact et $f : X \rightarrow X$ une application expansive. Montrer que f admet une mesure d'entropie maximale.

Exercice E.10.9 Soient X un espace métrique compact et $f : X \rightarrow X$ une application continue. Montrer que si f admet une unique mesure d'entropie maximale μ , alors μ est un point extrémal du convexe $\mathcal{M}(X)^f$, et donc que f est ergodique.

Exercice E.10.10 Soient X le cercle $\mathbb{R}/\mathbb{Z}$ (identifié en tant qu'ensemble avec $[0, 1[$) et $f : X \rightarrow X$ l'application définie par $x \mapsto Nx$, pour N dans $\mathbb{N} - \{0, 1\}$. Le but de cet exercice est, entre autre, de montrer que la mesure de Lebesgue est l'unique mesure d'entropie maximale pour f .

1. Soit μ une mesure de probabilité borélienne f -invariante sur X . Montrer que $h_\mu(f) \leq \log N$.
2. Montrer qu'il y a égalité si et seulement si μ est la mesure de Lebesgue λ sur X .
3. Calculer l'entropie topologique de f .
4. Dédire de ce qui précède que la mesure de Lebesgue λ est ergodique.

Exercice E.10.11 Soient X un espace métrique compact et $f : X \rightarrow X$ une application continue. Montrer que si f admet une unique mesure d'entropie maximale μ , et si $g : X \rightarrow X$ est une application mesurable telle que $f \circ g = f$, alors μ est g -invariante.

Dédire de l'exercice précédent que la mesure de Lebesgue sur le cercle est invariante par toutes les rotations du cercle.

10.7 Indications pour la résolution des exercices

Exercice E.10.1 Les distances d_n^f et d coïncident. On applique alors la proposition 10.5.

Exercice E.10.2 On majorera l'entier $N_{d_n^f}(\epsilon)$.

Exercice E.10.3 On vérifiera que, pour tout recouvrement ouvert $\mathcal{U}$ de X , on a

$$h(f^{-1}, \mathcal{U}) = h(f, \mathcal{U}) .$$

Exercice E.10.4

- i) Montrer que l'on peut supposer que $\mathcal{U}$ est un recouvrement fini.
- ii) Montrer qu'il existe $k \geq 1$ tel que le recouvrement fini $\mathcal{V} = \bigvee_{i=1}^k f^{-i}\mathcal{U}$ est plus fin que $\mathcal{U}$.
- iii) Montrer que $\mathcal{V} \preceq f^{-1}\mathcal{V}$, puis que, pour tout $n \geq 1$, on a $\mathcal{V} \preceq f^{-n}\mathcal{V}$.
- iv) En déduire que $\lim_{n \rightarrow \infty} \text{diam} \left(\bigvee_{i=0}^n f^{-i}\mathcal{V} \right) = 0$. Conclure.

Exercice E.10.7 On construira un contre-exemple sur un ensemble X à 5 éléments avec $N(\mathcal{U}|\mathcal{V}) = \frac{3}{2}$ et $N(\mathcal{U}|\mathcal{V}') = \frac{5}{3}$.

Exercice E.10.8 Soit ϵ_0 une constante d'expansivité de f . Pour tout n , soit E_n une partie ϵ_0 -séparée dont le cardinal est égal à $N_{d_n^f}(\epsilon_0)$. Par le corollaire 10.10, on a $h_{\text{top}}(f) = \limsup_{n \rightarrow \infty} \frac{1}{n} \log N'_{d_n^f}(\epsilon_0)$. Soit μ une mesure construite dans le lemme 10.15 pour E_n . Alors μ est une mesure d'entropie maximale.

Exercice E.10.9 Si $\mu = t\mu_1 + (1-t)\mu_2$ avec $0 < t < 1$, alors comme $h_{t\mu_1 + (1-t)\mu_2}(f) = th_{\mu_1}(f) + (1-t)h_{\mu_2}(f)$ (voir la proposition 9.18), et comme $h_{\mu_i}(f) \leq h_\mu(f)$ car μ maximise l'entropie, on a forcément $h_{\mu_1}(f) = h_{\mu_2}(f) = h_\mu(f)$, et donc par unicité, $\mu_1 = \mu_2 = \mu$, donc μ est bien un point extrémal.

On a vu (proposition 4.12) que les points extrémaux de $\mathcal{M}(X)^f$ sont précisément les mesures de $\mathcal{M}(X)^f$ qui sont ergodiques pour f .

Exercice E.10.10 (1) Soit ξ la partition mesurable $\{[\frac{k}{N}, \frac{k+1}{N}[/ k = 0, \dots, N-1\}$. Alors, par récurrence sur n dans $\mathbb{N} - \{0\}$, on a

$$\xi_n = \bigvee_{i=0}^{n-1} f^{-i}\xi = \left\{ \left[\frac{k}{N^n}, \frac{k+1}{N^n} \right[/ k = 0, \dots, N^n - 1 \right\} ,$$

donc ξ est génératrice pour f . Par le théorème de Kolmogorov-Sinai et la sous-additivité de la suite $(H_\mu(\xi_p))_{p \geq 1}$, pour tout n dans $\mathbb{N} - \{0\}$, on a

$$h_\mu(f) = h_\mu(f, \xi) = \inf_{p \geq 1} \frac{1}{p} H_\mu(\xi_p) \leq \frac{1}{n} H_\mu(\xi_n) \leq \frac{1}{n} \log(\text{Card } \xi_n) = \log N ,$$

la dernière inégalité ayant lieu, par un argument de concavité déjà vu (voir la remarque juste avant le paragraphe 9.2), avec égalité si et seulement si $\mu([\frac{k}{N^n}, \frac{k+1}{N^n}[) = \frac{1}{N^n}$ pour $k = 0, \dots, N^n - 1$.

(2) Il découle de ce qu'il précède que si $h_\mu(f) = \log N$, alors les mesures μ et λ coïncident sur tous les intervalles d'extrémités N -adiques (de la forme $[\frac{k}{N^n}, \frac{k+1}{N^n}[$ pour $k = 0, \dots, N^n - 1$ et n dans $\mathbb{N} - \{0\}$). Par densité, ces deux mesures sont donc égales.

(3) Par le principe variationnel, l'entropie topologique de f vaut donc

$$h_{\text{top}}(f) = \sup_{\mu \in \mathcal{M}(X)^f} h_\mu(f) = \log N .$$

(4) Ceci découle de l'exercice E.10.9 précédent.

Exercice E.10.11 Soit ξ une partition mesurable et $\xi_n = \bigvee_{i=0}^{n-1} f^{-i}\xi$. Alors

$$\begin{aligned} H_{g_*\mu}(\xi_n) &= H_\mu(g^{-1}\xi_n) = H_\mu\left(\left(\bigvee_{i=1}^{n-1} g^{-1}f^{-i}\xi\right) \vee g^{-1}\xi\right) = H_\mu\left(\left(\bigvee_{i=1}^{n-1} f^{-i}\xi\right) \vee g^{-1}\xi\right) \\ &\leq H_\mu\left(\bigvee_{i=1}^{n-1} f^{-i}\xi\right) + H_\mu(g^{-1}\xi) = H_\mu\left(\bigvee_{i=0}^{n-2} f^{-i}\xi\right) + H_\mu(g^{-1}\xi) . \end{aligned}$$

(La dernière inégalité découle de la sous-additivité de l'entropie des partitions, et la dernière égalité découle de l'invariance de μ par f .)

En divisant par n et en faisant tendre n vers $+\infty$, on a donc $h_{g_*\mu}(f, \xi) = h_\mu(f, \xi)$. Ceci étant vrai pour tout ξ , on a donc $h_{g_*\mu}(f) = h_\mu(f)$. Par unicité, on a donc $g_*\mu = \mu$. C'est ce qu'il fallait démontrer.

Par l'exercice précédent, la mesure de Lebesgue est invariante par toute rotation de la forme $x \mapsto x + \frac{k}{N}$ pour $k = 0, \dots, N-1$, et ceci pour tout N . Par densité, elle est donc invariante par toutes les rotations.

11 Sous-décalages de type fini

Une description combinatoire ou symbolique d'un système dynamique est parfois souhaitable. Un "codage" d'un système dynamique topologique consiste à repérer l'orbite d'un point x de X à partir d'une décomposition finie $X = \bigcup_{i \in \Lambda} R_i$ en fermés d'intérieurs disjoints. On dira qu'une suite ω de $\Lambda^{\mathbb{N}}$ (ou de $\Lambda^{\mathbb{Z}}$ si f est inversible) *code* le point x si, à chaque instant n , le point $f^n(x)$ appartient à R_{ω_n} , autrement dit si x est dans $\bigcap_n f^{-n}(R_{\omega_n})$.

Choisir un bon codage, c'est choisir une décomposition de sorte que les intersections $\bigcap_n f^{-n}(R_{\omega_n})$ contiennent au plus un point et telles que l'on puisse facilement décrire les suites ω pour lesquelles cette intersection est non vide (plus précisément telles que l'ensemble de ces suites forme un "système de Markov" ou "sous-décalage de type fini", au sens ci-dessous). Le système dynamique (X, f) apparaît alors comme l'"image" d'un sous-décalage de type fini (X_A, σ_A) (on parle aussi de "semi-conjugaison", voir le paragraphe 2.1).

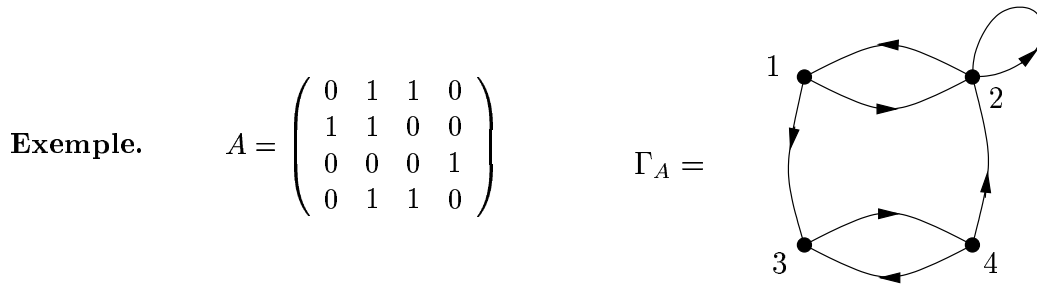
Le prototype d'un codage est l'écriture décimale d'un réel x de $[0, 1[$. Ceci code le système dynamique $(X = \mathbb{R}/\mathbb{Z}, f : x \mapsto 10x)$, en prenant pour alphabet fini $\Lambda = \{0, 1, \dots, 9\}$ et pour parties les $R_i = [\frac{i}{10}, \frac{i+1}{10}]$ pour i dans Λ . L'application de codage est l'application de $\Lambda^{\mathbb{N}}$ dans $[0, 1[\simeq \mathbb{R}/\mathbb{Z}$ définie par

$$(a_n)_{n \in \mathbb{N}} \mapsto \sum_{n \in \mathbb{N}} \frac{a_n}{10^{n+1}}.$$

Cette discussion justifie l'importance des sous-décalages de type fini. Ce chapitre leur est entièrement consacré. On calculera leur entropie topologique. On montrera que, pour les sous-décalages de type fini, il existe une unique mesure d'entropie maximale, et que cette mesure reflète la répartition statistique des orbites périodiques.

11.1 Systèmes de Markov

Soit $A = (A_{\alpha\beta})_{\alpha, \beta \in \Lambda}$ une matrice carrée $q \times q$ de coefficients égaux à 0 ou 1. Se donner une telle matrice revient à se donner un graphe orienté $\Gamma = \Gamma_A$ d'ensemble de sommets Λ de cardinal q , tel qu'il existe une arête allant du sommet α au sommet β si et seulement si $A_{\alpha, \beta} = 1$. La matrice A est alors appelée la *matrice d'incidence* du graphe orienté Γ_A .



On note

$$X_A = \{\omega \in \Lambda^{\mathbb{Z}} / \forall n \in \mathbb{Z}, A_{\omega_n, \omega_{n+1}} = 1\}$$

et σ_A la restriction du décalage ("shift" en anglais) à X_A . On munit X_A de la distance et de la topologie induite par la distance d sur $\Lambda^{\mathbb{Z}}$ (voir paragraphe 2.3). Le système dynamique topologique (X_A, σ_A) s'appelle le *sous-décalage de type fini* ou *système de Markov* (ou encore chaîne de Markov topologique, et "subshift of finite type" en anglais) de matrice A

(ou de graphe orienté Γ). Lorsque Γ est un graphe complet (i.e. lorsque A est une matrice carrée de coefficients tous égaux à 1), alors $(X_A, \Sigma_A) = (\Lambda^{\mathbb{Z}}, \sigma)$.

On dit qu'une matrice carrée B à coefficients réels positifs ou nuls est *transitive* (certains ouvrages, comme [PY], disent *apériodique*) s'il existe $n > 0$ tel que les coefficients de B^n sont tous non nuls. Un sous-décalage de type fini (X_A, σ_A) est dit *transitif* si A est transitive. C'est par exemple le cas lorsque A est une matrice carrée de coefficients tous égaux à 1.

Calculons l'entropie topologique d'un sous-décalage de type fini. Nous aurons besoin du lemme classique suivant surtout dans la prochaine partie.

Lemme 11.1 (Théorème de Perron-Frobenius) *Soient B une matrice carrée à coefficients réels positifs ou nuls. On suppose que B est transitive. Alors*

- a) *la matrice B admet un vecteur propre $v_{\max}$ à coefficients strictement positifs. Tout vecteur propre à coefficients positifs ou nuls est proportionnel à $v_{\max}$;*
- b) *si $\lambda_{\max} = \lambda_{\max}(B)$ est la valeur propre correspondant à $v_{\max}$, alors $\lambda_{\max}$ est simple, et toute autre valeur propre λ vérifie*

$$|\lambda| < \lambda_{\max} .$$

Preuve. Quitte à remplacer B par B^n , on peut supposer que $n = 1$. (En effet, soit $v_{\max}$ un vecteur propre de B^n à coefficients strictement positifs tel que tout vecteur propre de B^n à coefficients positifs ou nuls lui est proportionnel, et dont la valeur propre $\lambda_{\max}$ est simple et strictement plus grande que les valeurs absolues des autres valeurs propres de B^n . Alors $\sqrt[n]{\lambda_{\max}}$ est une valeur propre simple de B , strictement plus grande que les valeurs absolues des autres valeurs propres de B . De plus $Bv_{\max}$, qui est un vecteur propre de B^n pour la valeur propre $\lambda_{\max}$, est proportionnel à $v_{\max}$, donc $v_{\max}$ est vecteur propre de B , et sa valeur propre est nécessairement $\sqrt[n]{\lambda_{\max}}$. Si v est un vecteur propre de B à coefficients positifs ou nuls, alors il l'est aussi pour B^n , donc est proportionnel à $v_{\max}$.)

- a) Considérons le convexe compact

$$C = \{v = (x_1, \dots, x_q) \in \mathbb{R}^q / x_i \geq 0 \text{ et } x_1 + x_2 + \dots + x_q = 1\} .$$

L'application B induit une application $f_B : C \rightarrow C$ uniquement définie par $\mathbb{R}f_B(v) = \mathbb{R}Bv$ pour tout v dans C .

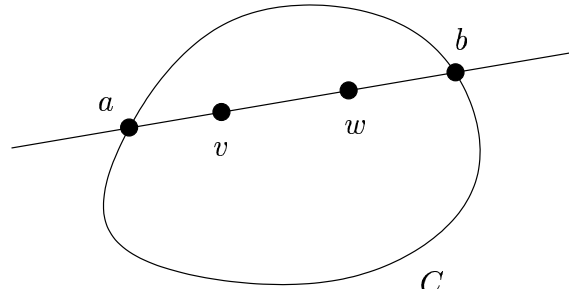
Munissons C de la distance (dite *distance de Hilbert*)

$$d(v, w) = \log[a, b, v, w]$$

avec a, b les points du bord de C sur la droite $\mathcal{D}$ passant par v et w (supposés distincts), avec a du côté de v , et où $[a, b, v, w]$ désigne le birapport de ces quatre points (rappelons que

$$[a, b, v, w] = \frac{(w - a)(b - v)}{v - a)(b - w)} ,$$

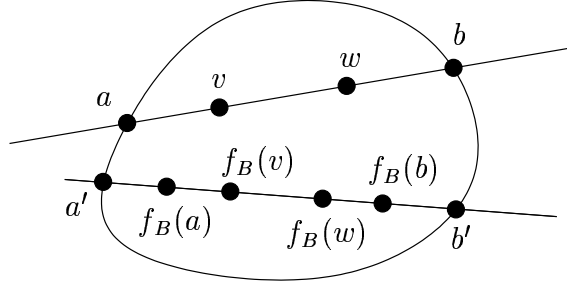
en utilisant une (et n'importe laquelle) coordonnée affine sur la droite $\mathcal{D}$.)



(Le fait que d vérifie l'inégalité triangulaire peut être démontré aisément. Nous omettons cette vérification car ce fait n'est pas utilisé dans l'argument ci-dessous. Voir par exemple [Har].)

Par l'hypothèse sur B^n , l'image $f_B(C)$ est contenue dans l'intérieur relatif du convexe C . Comme f_B conserve le birapport, on a donc, pour tous les v, w distincts dans C ,

$$d(f_B(v), f_B(w)) < d(v, w) .$$



Ceci prouve que f_B contient un unique point fixe dans C , qui est dans l'intérieur relatif de C : c'est le point v_0 qui minimise la fonction continue $v \mapsto d(v, f_B(v))$ sur le compact $f_B(C)$.

b) Sinon, en utilisant le théorème de Jordan, on pourrait trouver un sous-espace vectoriel B -invariant V de dimension 2 ou 3 de $\mathbb{R}^q$, contenant $v_{\max}$ et tel que dans une base de V ayant $v_{\max}$ comme premier vecteur, la restriction de B soit donnée par une matrice de la forme

$$\begin{pmatrix} \lambda_{\max} & 0 \\ 0 & \lambda \end{pmatrix}, \quad \begin{pmatrix} \lambda_{\max} & 1 \\ 0 & \lambda \end{pmatrix} \quad \text{ou} \quad \begin{pmatrix} \lambda_{\max} & 0 & 0 \\ 0 & \lambda \cos \theta & -\lambda \sin \theta \\ 0 & \lambda \sin \theta & \lambda \cos \theta \end{pmatrix}$$

avec $\lambda_{\max} \leq |\lambda|$. Dans ces trois cas, il n'existe pas dans V de cône C' convexe fermé *saillant* (i.e. ne contenant pas de droite vectorielle), de sommet l'origine de V , tel que $v_{\max}$ appartienne à l'intérieur de C' et dont l'image par B est contenue dans l'intérieur de C' (auquel on a rajouté 0). Ceci contredit ce que l'on a vu dans la preuve de l'assertion a). $\square$

L'entropie topologique d'un sous-décalage de type fini est donné par le résultat suivant.

Proposition 11.2 *Soit $\rho(A) = \lim_{n \rightarrow \infty} \|A^n\|^{\frac{1}{n}}$ le rayon spectral de la matrice A (pour $\|\cdot\|$ une norme matricielle), alors*

$$h_{\text{top}}(\sigma_A) = \log \rho(A) .$$

Remarque. Lorsque A est transitive, on a bien sûr $\rho(A) = \lambda_{\max}(A)$. En particulier, le rayon spectral d'une matrice carrée de coefficients tous égaux à 1 étant égal à son nombre de lignes, l'entropie topologique du décalage σ sur $\Lambda^{\mathbb{Z}}$, pour Λ un alphabet de cardinal p , est

$$h_{\text{top}}(\sigma) = \log p .$$

Pour montrer cette proposition, nous utiliserons le lemme suivant.

Lemme 11.3 *Pour α dans Λ , on note $X_\alpha = \{\omega \in X_A / \omega_0 = \alpha\}$. Alors l'ensemble*

$$\{(\alpha_0, \dots, \alpha_n) \in \Lambda^{n+1} / X_{\alpha_0} \cap \sigma^{-1}(X_{\alpha_1}) \cap \dots \cap \sigma^{-1}(X_{\alpha_n}) \neq \emptyset, \alpha_0 = \alpha, \alpha_n = \beta\}$$

a pour cardinal le coefficient α - β de la matrice A^n .

Preuve. Cet ensemble est aussi l'ensemble

$$\{(\alpha_0, \dots, \alpha_n) \in \Lambda^{n+1} / \alpha_0 = \alpha, \alpha_n = \beta, \text{ et } \forall k = 0, \dots, n-1, A_{\alpha_k, \alpha_{k+1}} = 1\}$$

Son cardinal est donc le nombre de chemins joignant α à β de longueur n dans le graphe Γ_A . Ce nombre est exactement le coefficient $A_{\alpha\beta}^n$ de la matrice A^n . $\square$

Comme $\text{tr } A^N = \sum_{\alpha \in \Lambda} A_{\alpha\alpha}^N$, le résultat suivant découle alors immédiatement de ce lemme.

Corollaire 11.4 *Soit $P_N = \{\omega \in X_A / \sigma_A^N \omega = \omega\}$. Alors on a*

$$\text{Card } P_N = \text{tr } A^N . \quad \square$$

Preuve de la proposition 11.2. Il existe des constantes c et C telles que, pour toute matrice B de taille $q \times q$, on a

$$c \|B\| \leq \sum_{\alpha, \beta} |B_{\alpha, \beta}| \leq C \|B\| .$$

Rappelons que la distance d sur X_A est définie par

$$d(\omega, \omega') = \sup_{i \in \mathbb{Z}} \frac{1}{2^{|i|}} d(\omega_i, \omega'_i) .$$

Le recouvrement ouvert $\mathcal{O}_1$ (par les boules ouvertes de rayon 1) de X_A est formé des q ouverts disjoints $\{X_\alpha\}_{\alpha \in \Lambda}$. C'est un recouvrement ouvert générateur pour σ_A , car

$$\text{diam}(\sigma_A^{-n} \mathcal{O}_1 \vee \dots \vee \sigma_A^{+n} \mathcal{O}_1) = 2^{-n} .$$

Par le théorème 10.9 et le lemme 11.3, on a donc

$$h_{\text{top}}(\sigma_A) = \lim_{n \rightarrow \infty} \frac{1}{n} \log \sum_{\alpha, \beta \in \Lambda} A_{\alpha\beta}^n = \lim_{n \rightarrow \infty} \frac{1}{n} \log \|A^n\| = \log \rho(A) . \quad \square$$

11.2 Chaînes de Markov

Il existe de nombreuses mesures invariantes sur le sous-décalage de type fini (X_A, σ_A) . Le but de cette partie est d'en construire à l'aide de matrices stochastiques. Nous allons appliquer le théorème de Carathéodory (le théorème 2.1 du chapitre 2.1) avec $\Omega = \Lambda^{\mathbb{Z}}$ et $\mathcal{A}$ l'algèbre des réunions finies disjointes de cylindres de Ω .

Rappelons (voir chapitre 2.1) que, par définition, un *cylindre* sur Ω est une partie de Ω de la forme

$$C_{\alpha_m, \alpha_{m+1}, \dots, \alpha_n}^{m, m+1, \dots, n} = \{\omega \in \Omega / \forall i = m, m+1, \dots, n, \omega_i = \alpha_i\} ,$$

avec $m \leq n$ dans $\mathbb{Z}$ et $\alpha_m, \alpha_{m+1}, \dots, \alpha_n$ dans Λ .

Pour construire une mesure de probabilité borélienne invariante par le décalage sur Ω , il suffit donc de se donner des réels positifs ou nuls $\mu(C_{\alpha_m, \alpha_{m+1}, \dots, \alpha_n}^{m, m+1, \dots, n})$ pour $m \leq n$ et $\alpha_m, \dots, \alpha_n$ dans Λ , de sorte que $\sum_{\alpha \in \Lambda} \mu(C_\alpha^m) = 1$ et

$$\mu(C_{\alpha_m, \alpha_{m+1}, \dots, \alpha_n}^{m, m+1, \dots, n}) = \sum_{\alpha \in \Lambda} \mu(C_{\alpha, \alpha_m, \dots, \alpha_n}^{m-1, m, \dots, n}) = \sum_{\alpha \in \Lambda} \mu(C_{\alpha_m, \dots, \alpha_n, \alpha}^{m, \dots, n, n+1}) .$$

En effet, la condition iii) du théorème 2.1 de Carathéodory est automatiquement satisfaite, car les cylindres $C_{\alpha_m, \dots, \alpha_n}^{m, \dots, n}$ sont des compacts de X , donc si $(C_n)_{n \in \mathbb{N}}$ est une suite décroissante dans $\mathcal{A}$ telle que $\bigcap_{n \in \mathbb{N}} C_n = \emptyset$, alors il existe n_0 tel que, pour tout $n \geq n_0$, on a $C_n = \emptyset$.

On appelle *matrice stochastique* $\pi = (\pi_{\alpha\beta})_{\alpha, \beta \in \Lambda}$ une matrice à coefficients positifs ou nuls tels que la somme des coefficients de chaque colonne vaut 1 : pour tout β dans Λ , on a

$$\sum_{\alpha \in \Lambda} \pi_{\alpha\beta} = 1 .$$

Le nombre réel $\pi_{\alpha\beta}$ s'interprète comme la “probabilité de passage du sommet α au sommet β ” (ou celle de provenir du sommet α sachant que l'on est sur le sommet β). Si l'on suppose que $A_{\alpha\beta}$ vaut 1 si et seulement si $\pi_{\alpha\beta}$ est non nul, alors “le processus se déroulera sur le graphe Γ_A ”.

Lemme 11.5 *Soit $\pi = (\pi_{\alpha\beta})_{\alpha, \beta \in \Lambda}$ une matrice stochastique transitive. Alors il existe un unique vecteur $v^\pi = (v_\alpha^\pi)_{\alpha \in \Lambda}$ tel que $\pi.v^\pi = v^\pi$ et $\sum_{\alpha \in \Lambda} v_\alpha^\pi = 1$. Les coefficients de v^π sont strictement positifs. Le réel 1 est une valeur propre simple de π et les autres valeurs propres vérifient $|\lambda| < 1$.*

Preuve. On applique le lemme de Perron-Frobenius. Il reste à démontrer que $\lambda_{\max}(\pi) = 1$. Or, en reprenant les notations de la démonstration de ce lemme, on remarque que l'hypothèse que π est stochastique assure que π préserve les hyperplans affines $H_t = \{v \in \mathbb{R}^\Lambda / \sum_{\alpha \in \Lambda} v_\alpha = t\}$, et donc que, pour tout v dans C , on a $f_\pi(v) = \pi v$. D'où $\lambda_{\max}(\pi) = 1$.

(Une autre manière de montrer que $\lambda_{\max}(\pi) = 1$ est de remarquer que le produit de deux matrices stochastiques est encore stochastique, donc que $\|\pi^n\|$ reste borné quand n tend vers $+\infty$, donc que le rayon spectral $\rho(\pi)$ de la matrice π vaut 1.) $\square$

À partir de toute matrice stochastique π , on peut construire une mesure de probabilité borélienne μ_π , appelée *mesure de Markov* sur $\Lambda^{\mathbb{Z}}$ par la formule

$$\mu_\pi(C_{\alpha_m, \alpha_{m+1}, \dots, \alpha_n}^{m, m+1, \dots, n}) = \left(\prod_{i=m}^{n-1} \pi_{\alpha_i \alpha_{i+1}} \right) v_{\alpha_n}^\pi ,$$

où v^π est le vecteur normalisé π -invariant (donné par le lemme précédent). Les conditions de compatibilité sont vérifiées, car π est stochastique et v^π est normalisé et π -invariant. Cette mesure de probabilité μ_π est invariante par le décalage σ . Le lemme suivant calcule son entropie (métrique), et généralise l'exemple de la fin du paragraphe 9.4.

Remarque. Il n'y a pas vraiment besoin de l'hypothèse de transitivité sur π pour construire une mesure de Markov (voir par exemple [PY]), mais par contre cette propriété sera utilisée pour avoir le mélange (voir le lemme 11.7 ci-dessous).

Lemme 11.6 *Soit π une matrice stochastique transitive. Alors*

$$h_{\mu_\pi}(\sigma) = - \sum_{\alpha, \beta \in \Lambda} \pi_{\alpha\beta} \log(\pi_{\alpha\beta}) v_\beta^\pi .$$

Preuve. On note ξ_0 la partition mesurable $\xi_0 = \{C_\alpha^0 / \alpha \in \Lambda\}$ de $\Lambda^{\mathbb{Z}}$. Cette partition est génératrice pour σ . On a donc, d'après le théorème de Kolmogorov-Sinai et le lemme 9.13,

$$h_{\mu_\pi}(\sigma) = h_{\mu_\pi}(\sigma, \xi_0) = \lim_{n \rightarrow \infty} H \left(\xi_0 \left| \bigvee_{i=1}^n \sigma^{-i} \xi_0 \right. \right).$$

Or

$$\begin{aligned} H \left(\xi_0 \left| \bigvee_{i=1}^n \sigma^{-i} \xi_0 \right. \right) &= - \sum_{\alpha_0, \dots, \alpha_n \in \Lambda} \log \left(\frac{\mu_\pi(C_{\alpha_0, \dots, \alpha_n}^{0, \dots, n})}{\mu_\pi(C_{\alpha_1, \dots, \alpha_n}^{1, \dots, n})} \right) \mu_\pi(C_{\alpha_0, \dots, \alpha_n}^{0, \dots, n}) = \\ &= - \sum_{\alpha_0, \alpha_1 \in \Lambda} \log(\pi_{\alpha_0 \alpha_1}) \mu_\pi(C_{\alpha_0, \alpha_1}^{0, 1}) = - \sum_{\alpha, \beta \in \Lambda} \pi_{\alpha \beta} \log(\pi_{\alpha \beta}) v_\beta^\pi. \end{aligned}$$

Ceci montre le résultat. $\square$

Lemme 11.7 *Soit π une matrice stochastique transitive. Alors le décalage σ est mélangeant pour μ_π .*

Preuve. D'après le lemme 9.8, les fonctions caractéristiques $\mathbb{I}_C$ des cylindres C engendrent un sous-espace dense de $\mathbb{L}^2(X, \mu_\pi)$. D'après la proposition 6.1, il suffit donc de montrer que, pour tous les cylindres C_1, C_2 de X , on a

$$\lim_{N \rightarrow \infty} \mu_\pi(C_1 \cap \sigma^{-N} C_2) = \mu_\pi(C_1) \mu_\pi(C_2).$$

On peut supposer que $C_1 = C_{\alpha_m, \dots, \alpha_n}^{m, \dots, n}$ et $C_2 = C_{\beta_p, \dots, \beta_q}^{p, \dots, q}$. On a alors, pour $N \geq n - p$,

$$\mu_\pi(C_1 \cap \sigma^{-N} C_2) = \left(\prod_{i=m}^{n-1} \pi_{\alpha_i \alpha_{i+1}} \right) (\pi^{N-n+p})_{\alpha_n \beta_p} \left(\prod_{i=p}^{m-1} \pi_{\beta_i \beta_{i+1}} \right) v_{\beta_q}^\pi.$$

Pour conclure, il suffit donc de montrer que, pour tout α, β dans Λ , on a $\lim_{N \rightarrow \infty} (\pi^N)_{\alpha \beta} = v_\alpha^\pi$. Or, par le lemme 11.5, la limite $\tau = \lim_{N \rightarrow \infty} \pi^N$ est le projecteur sur la droite $\mathbb{R} v_\alpha^\pi$ parallèlement à l'hyperplan $H_0 = \{v \in \mathbb{R}^\Lambda / \sum_{\alpha \in \Lambda} v_\alpha = 0\}$. Les coefficients matriciels de τ sont $\tau_{\alpha \beta} = v_\alpha^\pi$. C'est ce que l'on voulait. $\square$

Remarque. Avec l'exercice E.11.4, ceci montre en particulier qu'un système dynamique symbolique $(\Lambda^{\mathbb{Z}}, \sigma)$ admet une famille à plusieurs paramètres (donc non dénombrable) de mesures de probabilité invariantes, mélangeantes et deux à deux étrangères.

11.3 Mesure de Parry

La mesure de Parry est une mesure invariante sur le sous-décalage de type fini (X_A, σ_A) . Nous verrons que cette mesure est l'unique mesure qui maximise l'entropie. Nous verrons aussi que cette mesure reflète la statistique des orbites périodiques. Mais commençons par la construction de cette mesure.

Reprenons notre matrice $A = (A_{\alpha \beta})_{\alpha, \beta \in \Lambda}$ de coefficients égaux à 0 ou 1, et supposons que A est transitive. Notons $\lambda = \lambda_{\max}(A)$, $v = (v_\alpha)_{\alpha \in \Lambda}$ un vecteur propre de A pour la valeur propre λ , et $f = (f_\alpha)_{\alpha \in \Lambda}$ un vecteur propre de ${}^t A$ pour la valeur propre λ , donnés par le théorème de Perron-Frobenius. On a donc, pour tout α dans Λ ,

$$v_\alpha > 0, \quad f_\alpha > 0, \quad \sum_{\beta \in \Lambda} A_{\alpha \beta} v_\beta = \lambda v_\alpha, \quad \text{et} \quad \sum_{\beta \in \Lambda} A_{\beta \alpha} f_\beta = \lambda f_\alpha.$$

On normalise les vecteurs v et f de sorte que $\sum_{\alpha \in \Lambda} v_\alpha f_\alpha = 1$. On note alors $\pi^\bullet$ la matrice stochastique de coefficients

$$\pi_{\alpha\beta}^\bullet = \frac{A_{\alpha\beta} f_\alpha}{\lambda f_\beta}.$$

La mesure (de Markov associée) $\mu = \mu_{\pi^\bullet}$ s'appelle la *mesure de Parry*.

Par exemple, lorsque A est une matrice carrée de coefficients tous égaux à 1, alors λ vaut $\text{Card } \Lambda$, que l'on note q ci-dessous, $v_\alpha = f_\alpha = \frac{1}{\sqrt{q}}$ pour tout α dans Λ , et $\pi^\bullet$ est la matrice carrée de coefficients tous égaux à $\frac{1}{q}$. La mesure de Parry pour le décalage $\Sigma_A = \sigma$ est alors la mesure produit $\nu_0^{\mathbb{Z}}$, pour ν_0 la mesure d'équiprobabilité sur Λ .

Proposition 11.8 *Soit (X_A, σ_A) un sous-décalage de type fini transitif, μ la mesure de Parry sur X_A , $P_N = \{\omega \in X_A / \sigma^N \omega = \omega\}$ l'ensemble des point périodiques de période divisant N , et $\mu_N = \frac{1}{\text{Card } P_N} \sum_{\omega \in P_N} \delta_\omega$. Alors*

- a) on a $h_{\text{top}}(\sigma_A) = h_\mu(\sigma_A)$;
- b) la suite des mesures de probabilité μ_N converge faiblement vers μ ;
- c) la seule mesure de probabilité σ -invariante ν sur X_A telle que $h_\nu(\sigma_A) = h_{\text{top}}(\sigma_A)$ est $\nu = \mu$.

En particulier, l'unique mesure d'entropie maximale du décalage σ sur $\Lambda^{\mathbb{Z}}$, pour Λ un alphabet fini, est la mesure produit $\nu_0^{\mathbb{Z}}$, pour ν_0 l'équiprobabilité sur Λ .

Rappelons que d'après la proposition 11.2, on a

$$h_{\text{top}}(\sigma_A) = \log \lambda_{\max}(A).$$

Preuve. a) Remarquons que le vecteur $\pi^\bullet$ -invariant normalisé est

$$v^{\pi^\bullet} = (v_\alpha f_\alpha)_{\alpha \in \Lambda}.$$

Le lemme 11.6 donne alors

$$h_\mu(\sigma_A) = \sum_{\alpha, \beta \in \Lambda} (-\log(A_{\alpha\beta} f_\alpha) + \log f_\beta + \log \lambda) \frac{A_{\alpha\beta} f_\alpha v_\beta}{\lambda}.$$

Or

$$\begin{aligned} \sum_{\alpha, \beta} -\log(A_{\alpha\beta} f_\alpha) \frac{A_{\alpha\beta} f_\alpha v_\beta}{\lambda} &= -\frac{1}{\lambda} \sum_{\alpha, \beta} \log(f_\alpha) A_{\alpha\beta} f_\alpha v_\beta = -\sum_{\alpha} \log(f_\alpha) f_\alpha v_\alpha, \\ \sum_{\alpha, \beta} \log(f_\beta) \frac{A_{\alpha\beta} f_\alpha v_\beta}{\lambda} &= \sum_{\beta} \log(f_\beta) f_\beta v_\beta, \end{aligned}$$

et

$$\sum_{\alpha, \beta} \log \lambda \frac{A_{\alpha\beta} f_\alpha v_\beta}{\lambda} = \log \lambda \sum_{\alpha} f_\alpha v_\alpha = \log \lambda = h_{\text{top}}(\sigma_A).$$

Le résultat en découle.

b) Comme les fonctions localement constantes sont denses dans les fonctions continues sur l'espace de Cantor $\Lambda^{\mathbb{Z}}$, il suffit de montrer que, pour tout cylindre $C = C_{\alpha_m, \dots, \alpha_n}^{m, \dots, n}$ de

$\Lambda^{\mathbb{Z}}$, on a $\lim_{N \rightarrow \infty} \mu_N(C) = \mu(C)$. D'après le lemme 11.3 et son corollaire 11.4, on a les égalités $\text{Card } P_N = \text{tr } A^N$ et, pour $N > n - m$,

$$\mu_N(C) = \frac{\text{Card } (P_N \cap C)}{\text{Card } P_N} = \frac{1}{\text{tr } A^N} \left(\prod_{i=m}^{n-1} A_{\alpha_i \alpha_{i+1}} \right) (A^{N-n+m})_{\alpha_n \alpha_m}.$$

D'après le théorème de Perron-Frobenius, on a $\lim_{N \rightarrow \infty} \frac{1}{\lambda^N} A^N = \tau$, où τ est le projecteur sur la droite $\mathbb{R}v$ parallèlement à l'hyperplan noyau de la forme linéaire f , donc $\tau = (v_\alpha f_\beta)_{\alpha, \beta \in \Lambda}$. En particulier, on a $\lim_{N \rightarrow \infty} \frac{1}{\lambda^N} \text{tr } A^N = 1$ et $\lim_{N \rightarrow \infty} \frac{1}{\lambda^N} (A^N)_{\alpha, \beta} = v_\alpha f_\beta$. D'où

$$\lim_{N \rightarrow \infty} \mu_N(C) = \left(\prod_{i=m}^{n-1} A_{\alpha_i \alpha_{i+1}} \right) \frac{f_{\alpha_m} v_{\alpha_n}}{\lambda^{n-m}} = \left(\prod_{i=m}^{n-1} \pi_{\alpha_i \alpha_{i+1}}^\bullet \right) f_{\alpha_n} v_{\alpha_n} = \mu(C).$$

C'est ce que l'on voulait.

c) Soit ν une mesure de probabilité borélienne σ_A -invariante sur X_A telle que $h_\nu(\sigma_A) = h_{\text{top}}(\sigma_A) = \log \lambda$.

Si ν est absolument continue par rapport à la mesure de Parry μ , alors il existe, d'après le théorème de Radon-Nokodym, une fonction ψ dans $\mathbb{L}^1(X_A, \mu)$ telle que $\nu = \psi\mu$. Cette fonction est σ_A -invariante μ -presque partout. Comme μ est ergodique (voir le lemme 11.7), on a $\psi = 1$ et $\nu = \mu$, ce qui est exactement ce que l'on cherche à démontrer.

Supposons donc par l'absurde que ν n'est pas absolument continue par rapport à μ . Il existe donc un borélien B de X_A tel que $\nu(B) > 0$ et $\mu(B) = 0$. Le lemme 9.8 appliqué à la mesure $\frac{1}{2}(\nu + \mu)$, permet de trouver une suite B_p d'unions finies disjointes de cylindres de X_A tels que $\lim_{p \rightarrow \infty} \nu(B_p) = \nu(B) \neq 0$ et $\lim_{p \rightarrow \infty} \mu(B_p) = 0$.

Notons ξ_0 la partition mesurable de X_A définie par $\xi_0 = \{X_A \cap C_\alpha^0 / \alpha \in \Lambda\}$ et ξ_n la partition $\xi_n = \bigvee_{i=-n}^n \sigma_A^{-i} \xi_0$. Comme ξ_0 est une partition génératrice pour σ_A (celle-ci étant inversible), on a, d'après le théorème de Kolmogorov-Sinai,

$$h_\nu(\sigma_A) = h_\nu(\sigma_A, \xi_0) = \inf_{n \geq 1} \frac{1}{n} H_\nu \left(\bigvee_{i=0}^{n-1} \sigma^{-i} \xi_0 \right) = \inf_{n \geq 1} \frac{1}{2n} H_\nu(\xi_n).$$

On a donc, pour tout $n \geq 1$, l'inégalité $H_\nu(\xi_n) - 2n \log \lambda \geq 0$.

Lemme 11.9 Soient $x_1, \dots, x_k$ des réels strictement positifs avec $\sum_{i=1}^k x_i = a$. Alors $-\sum_{i=1}^k x_i \log x_i \leq a \log \frac{k}{a}$.

Preuve. Par concavité de la fonction $\varphi(t) = -t \log t$, on a

$$\frac{1}{k} \sum_{i=1}^k -x_i \log x_i \leq - \left(\frac{\sum x_i}{k} \right) \log \left(\frac{\sum x_i}{k} \right) = \frac{a}{k} \log \frac{k}{a}.$$

□

Lemme 11.10 Il existe $c > 0$ tel que, pour tout cylindre $C_{\alpha_m, \dots, \alpha_n}^{m, \dots, n}$ non vide de X_A , on a $\mu(C_{\alpha_m, \dots, \alpha_n}^{m, \dots, n}) \geq c \lambda^{-(n-m)}$.

Preuve. On a

$$\mu(C_{\alpha_m, \dots, \alpha_n}^m, \dots, n) = \left(\prod_{i=m}^{n-1} A_{\alpha_i \alpha_{i+1}} \right) f_{\alpha_m} v_{\alpha_n} \lambda^{-(n-m)}.$$

Il suffit de prendre $c = \inf_{\alpha, \beta \in \Lambda} f_{\alpha} v_{\beta} > 0$. $\square$

Fin de la preuve de la proposition 11.8. Fixons p et choisissons n de sorte que B_p soit une réunion finie de parties de la partition ξ_n . On a alors, en utilisant les deux lemmes précédents,

$$\begin{aligned} 0 \leq H_{\nu}(\xi_n) - 2n \log \lambda &\leq - \sum_{C \in \xi_n, C \subset B_p} \nu(C) \log \nu(C) - \sum_{C \in \xi_n, C \subset {}^c B_p} \nu(C) \log \nu(C) - 2n \log \lambda \\ &\leq \nu(B_p) \log \frac{\text{Card}\{C \in \xi_n / C \subset B_p\}}{\nu(B_p)} + \nu({}^c B_p) \log \frac{\text{Card}\{C \in \xi_n / C \subset {}^c B_p\}}{\nu({}^c B_p)} - 2n \log \lambda \\ &\leq \nu(B_p) \log \frac{\mu(B_p) \lambda^{2n}}{c \nu(B_p)} + \nu({}^c B_p) \log \frac{\mu({}^c B_p) \lambda^{2n}}{c \nu({}^c B_p)} - 2n \log \lambda \\ &= \nu(B_p) \log \frac{\mu(B_p)}{\nu(B_p)} + \nu({}^c B_p) \log \frac{\mu({}^c B_p)}{\nu({}^c B_p)} - \log c. \end{aligned}$$

Ces inégalités sont vraies pour tout p , mais le dernier terme tend vers $-\infty$ quand p tend vers l'infini, contradiction. $\square$

Remarque. La proposition 11.8 est le prototype d'un résultat plus général du à Bowen et valable pour les "ensembles hyperboliques" (localement maximaux et topologiquement transitifs) des systèmes dynamiques (voir par exemple [Bow, KH]).

11.4 Exercices

Exercice E.11.1 Soient (X_A, σ_A) un sous-décalage de type fini de matrice d'incidence A transitive et $P_n = \{\omega \in X_A / \sigma_A^n(\omega) = \omega\}$ l'ensemble des points périodiques de période divisant n . Montrer que, quand n tend vers $+\infty$, on a

$$\text{Card } P_n \sim e^{n h_{\text{top}}(\sigma_A)}.$$

Exercice E.11.2 Soit $A = (A_{\alpha, \beta})_{\alpha, \beta \in \Lambda}$ une matrice carrée de coefficients égaux à 0 ou 1. On dit que A est *irréductible* si, pour tout α, β dans Λ , il existe $n > 0$ tel que $(A^n)_{\alpha\beta} \neq 0$. On appelle *période* d de A le plus grand commun diviseur de l'ensemble $\{n > 0 / \exists \alpha \in \Lambda, (A^n)_{\alpha\alpha} \neq 0\}$. Montrer que A est transitive si et seulement si A est irréductible de période 1.

Exercice E.11.3 Soit $A = (A_{\alpha, \beta})_{\alpha, \beta \in \Lambda}$ une matrice carrée de coefficients égaux à 0 ou 1. On suppose que A n'a pas de ligne ni de colonne nulle. Montrer que A est irréductible (voir l'exercice précédent) si et seulement si le décalage σ_A admet une orbite dense dans X_A .

Exercice E.11.4 Soient π, π' deux matrices stochastiques transitives distinctes, indexées par un même ensemble fini Λ . Montrer que les mesures de probabilité μ_{π} et $\mu_{\pi'}$ sont étrangères.

Exercice E.11.5 Soit π une matrice stochastique transitive. Montrer, pour presque tout ω de $\Lambda^{\mathbb{Z}}$, et tout α, β dans Λ , que l'on a

- a) $\lim_{n \rightarrow \infty} \frac{1}{n} \text{Card} \{i \in \{1, \dots, n\} / \omega_i = \alpha\} = v_{\alpha}^{\pi}.$
b) $\lim_{n \rightarrow \infty} \frac{1}{n} \text{Card} \{i \in \{1, \dots, n\} / \omega_i = \alpha \text{ et } \omega_{i+1} = \beta\} = \pi_{\alpha\beta} v_{\beta}^{\pi}.$

Exercice E.11.6 (Examen de janvier 2002) Les parties I et II sont indépendantes, et sont utilisées dans les parties III et IV. Pour tout réel x , on note $[x]$ sa partie entière et $\{x\}$ sa partie fractionnaire. On fixe un réel $\beta > 1$. Le but principal de ce problème est d'étudier la dynamique de la tranformation de l'intervalle $I = [0, 1[$ donnée par $f(x) = \{\beta x\}$.

Partie I : Une transformation de l'intervalle.

Soient $t_0 = 0 < t_1 < \dots < t_n < \dots$ une suite strictement croissante telle que $\lim_{t \rightarrow \infty} t_n = 1$ et $g : I \rightarrow I$ l'application donnée par, pour tout x dans l'intervalle $I_N = [t_n, t_{n+1}[$,

$$g(x) = \frac{x - t_n}{t_{n+1} - t_n}.$$

- (1) Soit $\lambda = dx$ la mesure de Lebesgue sur I . Montrer que g préserve λ .
(2) Soit $a = (a_1, \dots, a_m)$ dans $\mathbb{N}^m$. On note $I_a = I_{a_1} \cap g^{-1}(I_{a_2}) \cap \dots \cap g^{-(m-1)}(I_{a_m})$. Montrer que, pour tout borélien A de I , on a $\lambda(g^{-m}(A) \cap I_a) = \lambda(A)\lambda(I_a)$.
(3) Montrer que g est mélangeante pour λ .

Partie II : Un système dynamique symbolique.

Soient $q = (q_n)_{n \in \mathbb{N} - \{0\}}$ et $(x_n)_{n \in \mathbb{N}}$ les suites définies par la condition initiale $x_0 = 0$ et la relation de récurrence suivante: pour $n \geq 1$, q_n est le plus grand des entiers q tel que le réel $x_n = x_{n+1} + q\beta^{-n}$ vérifie $x_n \leq 1$.

- (1) a) Vérifier que $q_1 = [\beta]$ et que, pour tout $n \geq 1$, on a $q_n \leq [\beta]$.
b) Vérifier que $\sum_{i \in \mathbb{N} - \{0\}} q_i \beta^{-i} = 1$.
(2) Soient $\Lambda = \{0, 1, \dots, [\beta]\}$, $X = \Lambda^{\mathbb{N} - \{0\}}$ muni de sa structure d'espace métrique compact, et $\sigma : X \rightarrow X$ le décalage (on rappelle qu'il est donné par $\sigma(\omega)_i = \omega_{i+1}$). On munit X de l'ordre lexicographique, défini par $\omega \leq \omega'$ si et seulement si ou bien $\omega = \omega'$ ou bien il existe $i_0 \in \mathbb{N} - \{0\}$ tel que $\omega_{i_0} < \omega'_{i_0}$ et $\omega_i = \omega'_i$ pour tout $i < i_0$.
a) Soit $X_{\beta} = \{\omega \in X / \forall n \in \mathbb{N}, \sigma^n(\omega) \leq q\}$. Montrer que X_{β} est compact.
b) Vérifier que $\sigma(X_{\beta}) \subset X_{\beta}$. On notera alors σ_{β} la restriction de σ à X_{β} .
(3) Soit $p_n = \text{Card}\{\omega_1, \dots, \omega_n\} \in \Lambda^n / \exists \omega' \in X_{\beta}, \forall i \in \{1, \dots, n\}, \omega'_i = \omega_i$.
a) Montrer que $p_1 = [\beta] + 1$ et, pour tout $n \geq 2$, $p_n = q_1 p_{n-1} + q_2 p_{n-2} + \dots + q_{n-1} p_1 + q_n + 1$.
b) Montrer que, pour tout $n \geq 1$, on a $p_n \leq p_1 n \beta^n$.
c) Montrer que $\lim_{n \rightarrow \infty} \sqrt[n]{p_n} = \beta$. (On pourra chercher, pour tout $\gamma < \beta$, une minoration de la forme $p_n \geq c \gamma^n$.)

d) Montrer que l'entropie topologique de σ_β est égale à $\log \beta$.

Partie III : Ergodicité de f .

Soit $\varphi : I \rightarrow]0, +\infty[$ la fonction définie par

$$\varphi = C \sum_{n \in \mathbb{N}} \beta^{-n} \mathbb{I}_{\{x / 0 \leq x \leq y_n\}}$$

où $y_n = (1 - x_n)\beta^n$ et où la constante $C > 0$ est choisie pour que la mesure $\mu = \varphi\lambda$ soit une mesure de probabilité. On veut montrer que f préserve μ . Soit A un borélien de I . On posera $f^0(A) = A$.

- (1) a) Montrer que f^n induit une bijection affine croissante de $[x_n, 1[$ sur $[0, y_n[$.
b) Montrer que $\mu(A) = C \sum_{n \in \mathbb{N}} \lambda(f^{-n}(A) \cap [x_n, 1[)$.
c) Montrer que, pour tout $n \in \mathbb{N} - \{0\}$, on a $\lambda(f^{-n}(A) \cap [x_{n-1}, x_n]) = q_n \beta^{-n} \lambda(A)$.
d) Montrer que f préserve μ .
- (2) Soit $F : I \rightarrow I$ l'application donnée par, pour tout x dans $[x_{n-1}, x_n[$, $F(x) = f^n(x)$.
a) Montrer que F préserve λ et que F est ergodique pour λ .
b) Montrer que f est ergodique pour μ . (On pourra montrer l'implication $(f^{-1}(A) = A) \Rightarrow (F^{-1}(A) = A)$.)
c) Montrer que μ est l'unique mesure de probabilité f -invariante sur I qui est absolument continue par rapport à la mesure de Lebesgue.

Partie IV : Développement β -adique et entropie de f .

Soient $\Omega_1 : I \rightarrow \Lambda$ l'application donnée par $\Omega_1(x) = [\beta x]$ et, pour $n \in \mathbb{N} - \{0, 1\}$, $\omega_n = \Omega_1 \circ f^n$. La suite $\Omega(x) = (\Omega_i(x))_{i \in \mathbb{N} - \{0\}}$ est appelée le *développement β -adique* de x .

- (1) a) Montrer que $\Omega(x)$ appartient à X_β et que l'application $\Omega : I \rightarrow X_\beta$ est injective.
b) Montrer que, pour λ -presque tout x , la limite

$$\ell_\beta = \lim_{n \rightarrow \infty} \frac{1}{n} \text{Card } \{i \leq n / \Omega_i(x) = 0\}$$

existe et ne dépend pas de x .

- (2) On note α la partition mesurable

$$\alpha = \left\{ \left[0, \frac{1}{\beta} \right[, \left[\frac{1}{\beta}, \frac{2}{\beta} \right[, \dots , \left[\frac{q_1 - 1}{\beta}, \frac{q_1}{\beta} \right[, \left[\frac{q_1}{\beta}, 1 \right[\right\}.$$

- a) Montrer que tout élément B de la partition mesurable $\bigvee_{0 \leq k \leq n-1} f^{-k} \alpha$ vérifie $\lambda(B) \leq \beta^{-n}$.
b) Montrer la minoration suivante de l'entropie métrique de f : $h_\mu(f) \geq \log \beta$.
c) Soit $\nu = \Omega_* \mu$ la mesure de probabilité sur X_β image de μ . Montrer que σ_β préserve ν .
d) Montrer que $h_\nu(\sigma_\beta) = h_\nu(f)$.
e) Calculer l'entropie métrique $h_\mu(f)$ de f .
- (3) On suppose que β est le nombre d'or $\beta = \frac{1+\sqrt{5}}{2}$. On note Y le support de ν .
a) Calculer la suite $(q_n)_{n \in \mathbb{N} - \{0\}}$.
b) Calculer ℓ_β .
c) Montrer que $(Y, \sigma|_Y)$ est un sous-décalage de type fini dont ν est la mesure de Parry.

Exercice E.11.7 (Examen de janvier 2003) Les parties II et III (sauf III (9)) sont indépendantes.

Dans tout ce problème, (X, d) est un espace métrique compact, $f : X \rightarrow X$ un homéomorphisme, et $\varphi : X \rightarrow \mathbb{R}$ une application continue. On rappelle que

$$d_n^f(x, y) = \sup_{i=0, \dots, n-1} d(f^i(x), f^i(y)) ,$$

qu'une partie E de X est ϵ -dense pour une distance d' si $\forall x \in X, \exists y \in E, d'(x, y) \leq \epsilon$, et ϵ -séparée pour une distance d' si $\forall x, y \in E, x \neq y \Rightarrow d'(x, y) \geq \epsilon$.

Partie I: Pression métrique et topologique.

Pour $n \in \mathbb{N} - \{0\}$, $x \in X$, $\epsilon > 0$, on note

- $S_n^f \varphi(x) = \sum_{i=0}^{n-1} \varphi(f^i(x))$,
- $N_{f, \varphi}(n, \epsilon)$ la borne inférieure, sur toutes les parties ϵ -denses E de X pour la distance d_n^f , des sommes $\sum_{x \in E} e^{S_n^f \varphi(x)}$,
- $N'_{f, \varphi}(n, \epsilon)$ la borne supérieure, sur toutes les parties ϵ -séparées E de X pour la distance d_n^f , des sommes $\sum_{x \in E} e^{S_n^f \varphi(x)}$.

(1) Montrer que

$$\lim_{\epsilon \rightarrow 0} \left(\limsup_{n \rightarrow \infty} \frac{1}{n} \log N_{f, \varphi}(n, \epsilon) \right) = \lim_{\epsilon \rightarrow 0} \left(\limsup_{n \rightarrow \infty} \frac{1}{n} \log N'_{f, \varphi}(n, \epsilon) \right) .$$

Ce nombre, noté $P(f, \varphi)$, est appelé la *pression topologique* de f (par rapport à φ).

Si μ est une mesure de probabilité f -invariante sur X , le nombre

$$P_\mu(f, \varphi) = h_\mu(f) + \int \varphi d\mu$$

est appelé la *pression (métrique)* de f pour μ (par rapport à φ).

(2) Montrer que si $\varphi = 0$, alors $P(f, \varphi) = h_{\text{top}}(f)$, $P_\mu(f, \varphi) = h_\mu(f)$. Montrer que $P(f, \varphi) \leq h_{\text{top}}(f) + \sup_X |\varphi|$, $P_\mu(f, \varphi) \leq h_\mu(f) + \sup_X |\varphi|$.

(3) Pour tout $k \geq 1$, montrer que si $g = f^k$ et $\psi = S_k^f \varphi$, alors $P_\mu(g, \psi) = k P_\mu(f, \varphi)$ et $P(g, \psi) = k P(f, \varphi)$.

(4) Montrer que si $\varphi = \varphi' + \psi \circ f - \psi$, avec $\varphi', \psi : X \rightarrow \mathbb{R}$ continues, alors $P(f, \varphi) = P(f, \varphi')$ et $P_\mu(f, \varphi) = P_\mu(f, \varphi')$.

Partie II: Le principe variationnel pour la pression.

Le but de cette partie est de montrer le *principe variationnel pour la pression*:

$$P(f, \varphi) = \sup_{\mu \in \mathcal{M}(X)^f} P_\mu(f, \varphi) .$$

(1) Soient $a_1, \dots, a_k$ dans $\mathbb{R}$, $A = \sum_{i=1}^k e^{a_i}$, et $p_1, \dots, p_k \geq 0$ tels que $\sum_{i=1}^k p_i = 1$. Montrer que $\sum_{i=1}^k p_i(-\log p_i + a_i) \leq \log A$, avec égalité si $p_i = \frac{e^{a_i}}{A}$.

Pour $n \in \mathbb{N} - \{0\}$ et $\epsilon > 0$, soit E_n une partie ϵ -séparée pour la distance d_n^f . On note

$$\nu_n = \frac{1}{\sum_{x \in E_n} e^{S_n^f \varphi(x)}} \sum_{x \in E_n} e^{S_n^f \varphi(x)} \delta_x \quad \text{et} \quad \mu_n = \frac{1}{n} \sum_{i=0}^{n-1} (f^i)_* \nu_n.$$

(2) Montrer que la suite $(\mu_n)_{n \in \mathbb{N} - \{0\}}$ admet une sous-suite faiblement convergente vers une limite μ . Montrer qu'une telle mesure de probabilité μ est f -invariante.

(3) Soit α une partition mesurable de X telle que $\text{diam } A < \epsilon$ et $\mu(\partial A) = 0$ pour tout élément A de α . Posons $\alpha_k = \bigvee_{i=0}^{k-1} f^{-i} \alpha$. Montrer, en utilisant (1), que

$$H_{\nu_n}(\alpha_n) + \int S_n^f \varphi d\nu_n = \log \sum_{x \in E_n} e^{S_n^f \varphi(x)}.$$

(4) En déduire que, pour $0 < q < n$, on a

$$\frac{q}{n} \log \sum_{x \in E_n} e^{S_n^f \varphi(x)} \leq H_{\mu_n}(\alpha_q) + \frac{2q^2}{n} \log \text{Card}(\alpha) + q \int \varphi d\mu_n.$$

(5) En déduire que $\limsup_{n \rightarrow \infty} \frac{1}{n} \log \sum_{x \in E_n} e^{S_n^f \varphi(x)} \leq P_\mu(f, \varphi)$, puis que

$$P(f, \varphi) \leq \sup_{\nu \in \mathcal{M}(X)^f} P_\nu(f, \varphi).$$

(6) Maintenant, on considère μ un élément quelconque de $\mathcal{M}(X)^f$. Soit $\alpha = \{A_1, \dots, A_p\}$ une partition mesurable de X , B_i un compact contenu dans A_i , et $U_0 = X - (B_1 \cup \dots \cup B_p)$, de sorte que, si β est la partition mesurable $\{U_0, B_1, \dots, B_p\}$, alors $h_\mu(f, \alpha) \leq h_\mu(f, \beta) + 1$. Pour tout n dans $\mathbb{N}$, soit $\epsilon > 0$ tel que $d(B_i, B_j) > 2\epsilon$ pour $1 \leq i \neq j \leq p$, et $|\varphi(x) - \varphi(y)| \leq 1$ si $d(x, y) \leq \epsilon$. Soit E une partie ϵ -dense pour la distance d_n^f , et $\beta_n = \bigvee_{i=0}^{n-1} f^{-i} \beta$. Montrer, en utilisant (1), que

$$H_\mu(\beta_n) + \int S_n \varphi d\mu \leq \log \left(2^n \sum_{x \in E} e^{S_n \varphi(x) + n} \right).$$

(On pourra, pour tout élément C de β_n , fixer un point x_C dans l'adhérence de C tel que $\sup_{x \in C} S_n \varphi(x) = S_n \varphi(x_C)$ et y_C dans E tel que $d_n^f(x_C, y_C) \leq \epsilon$.)

(7) En déduire que $P_\mu(f, \varphi) \leq 2 + \log 2 + P(f, \varphi)$.

(8) En déduire le résultat.

Partie III : Mesures d'équilibre de la pression pour les sous-décalages de type fini.

Soit Λ un alphabet fini, $A = (A_{\alpha\beta})_{\alpha, \beta \in \Lambda}$ une matrice carrée de coefficients 0 ou 1 transitive, et (X_A, σ_A) le sous-décalage de type fini associé. Le but de cette partie est de

montrer que, pour certaines applications continues $\varphi : X_A \rightarrow \mathbb{R}$, il existe une et une seule mesure μ dans $\mathcal{M}(X_A)^{\sigma_A}$ telle que $P_\mu(\sigma_A, \varphi) = P(\sigma_A, \varphi)$. Une telle mesure est appelée *mesure d'équilibre* pour la pression.

Pour tout x dans $\Lambda^{\mathbb{Z}}$, on note x_i la i -ème composante de x , de sorte que $x = (x_i)_{i \in \mathbb{Z}}$. On rappelle que $\Lambda^{\mathbb{Z}}$ est muni de la distance $d(x, y) = \sup_{i \in \mathbb{Z}} \frac{1}{2^{|i|}} \delta_{x_i}^{y_i}$, avec $\delta_s^t = 0$ si $s = t$ et $\delta_s^t = 1$ sinon.

Soit $\psi : X_A \rightarrow \mathbb{R}$ une application, et k dans $\mathbb{N} - \{0\}$. On dit que ψ *ne dépend que des k premières coordonnées* si, pour tous les x, y dans X_A , si $x_i = y_i$ pour $i = 0, \dots, k-1$, alors $\psi(x) = \psi(y)$. On note alors $\psi(x) = \psi_{x_0, \dots, x_{k-1}}$.

Dans la suite, on fixe une application $\varphi : X_A \rightarrow \mathbb{R}$ qui ne dépend que des deux premières coordonnées.

(1) Montrer que $S_n^{\sigma_A} \varphi$ ne dépend que des $n+1$ premières coordonnées, et que, pour tous les x, y dans X_A , si $d_n^{\sigma_A}(x, y) < \frac{1}{2}$, alors $S_n^{\sigma_A} \varphi(x) = S_n^{\sigma_A} \varphi(y)$.

(2) Montrer que pour tout $\epsilon \leq \frac{1}{2}$, il existe une constante $C_\epsilon > 0$ telle que pour tout n dans $\mathbb{N} - \{0\}$,

$$N'_{\sigma_A, \varphi}(\epsilon, n) \leq C_\epsilon N'_{\sigma_A, \varphi}\left(\frac{1}{2}, n\right).$$

(3) En déduire que $P(\sigma_A, \varphi) = \limsup_{n \rightarrow \infty} \frac{1}{n} \log N'_{\sigma_A, \varphi}\left(\frac{1}{2}, n\right)$.

On note A^φ la matrice $(A^\varphi_{\alpha\beta})_{\alpha, \beta \in \Lambda}$ avec

$$A^\varphi_{\alpha\beta} = A_{\alpha\beta} e^{\varphi_{\alpha\beta}}.$$

(4) Montrer que A^φ admet un vecteur propre $v^\varphi = (v_\alpha)_{\alpha \in \Lambda}$ de coefficients tous strictement positifs, unique à scalaire strictement positif près, dont la valeur propre λ^φ est simple, et vérifie $\lambda^\varphi > |\lambda|$ pour tout autre valeur propre λ de A^φ .

(5) On rappelle que, pour $B = (B_{\alpha\beta})_{\alpha, \beta \in \Lambda}$, on note $B_{\alpha\beta}^n$ le coefficient α - β de la puissance n -ème de la matrice B . Montrer que pour tous les α, β dans Λ et tout n dans $\mathbb{N} - \{0\}$,

$$\sum_{(\alpha_0, \dots, \alpha_n) \in \Lambda^{n+1} / \alpha_0 = \alpha, \alpha_n = \beta, A_{\alpha_i \alpha_{i+1}} = 1} e^{(S_n^{\sigma_A} \varphi)_{\alpha_0, \dots, \alpha_n}} = (A^\varphi)_{\alpha\beta}^n.$$

En déduire, en utilisant la question (3) que $P(\sigma_A, \varphi) = \log \lambda^\varphi$.

Soit $f^\varphi = (f_\alpha)_{\alpha \in \Lambda}$ un vecteur propre de la matrice ${}^t A^\varphi$ pour la valeur propre λ^φ tel que $\sum_{\alpha \in \Lambda} f_\alpha v_\alpha = 1$. On note π la matrice $(\pi_{\alpha\beta})_{\alpha, \beta \in \Lambda}$ où

$$\pi_{\alpha\beta} = \frac{A^\varphi_{\alpha\beta} f_\alpha}{\lambda^\varphi f_\beta}.$$

(6) Montrer que π est une matrice stochastique. On note μ^φ la mesure de Markov sur X_A associée à π . Montrer que $P_{\mu^\varphi}(\sigma_A, \varphi) = P(\sigma_A, \varphi)$.

(7) Soit $P_N = \{x \in X_A / \sigma_A^N(x) = x\}$ l'ensemble des points périodiques pour σ_A , de période divisant N . Soit

$$\mu_N = \frac{1}{\sum_{x \in P_N} e^{S_N \varphi(x)}} \sum_{x \in P_N} e^{S_N \varphi(x)} \delta_x,$$

où δ_x est la masse de Dirac au point x . Montrer que la suite des mesures de probabilité $(\mu_N)_{N \in \mathbb{N}}$ converge faiblement vers μ^φ .

(8) Montrer qu'il existe des constantes $c_1, c_2 > 0$ telles que, pour tout cylindre $C = C_{\alpha_m, \dots, \alpha_n}^{m, \dots, n}$ non vide de X_A , on a

$$c_1(\lambda^\varphi)^{-(n-m)} e^{(S_{n-m}\varphi)_{\alpha_m, \dots, \alpha_n}} \leq \mu^\varphi(C) \leq c_2(\lambda^\varphi)^{-(n-m)} e^{(S_{n-m}\varphi)_{\alpha_m, \dots, \alpha_n}}.$$

(9) En déduire que la mesure μ^φ est l'unique mesure ν dans $\mathcal{M}(X_A)^{\sigma_A}$ telle que $P_\nu(\sigma_A, \varphi) = P(\sigma_A, \varphi)$.

11.5 Indications pour la résolution des exercices

Exercice E.11.1 Utiliser la proposition 11.2 et le corollaire 11.4.

Exercice E.11.2 Montrer d'abord qu'il existe $n > 0$ tel que, pour tout α , on a $(A^n)_{\alpha\alpha} \neq 0$.

Exercice E.11.5 Utiliser le théorème ergodique de Birkhoff.

Exercice E.11.6 Partie I (1) Comme g est une bijection affine croissante de chaque intervalle I_n sur $[0, 1]$, on a, pour tout borélien A de I , $\lambda(g^{-1}(A) \cap I_n) = \lambda(A)\lambda(I_n)$. Donc

$$\lambda(g^{-1}(A)) = \sum_{n \in \mathbb{N}} \lambda(g^{-1}(A) \cap I_n) = \lambda(A) \sum_{n \in \mathbb{N}} \lambda(I_n) = \lambda(A).$$

Donc la mesure de probabilité λ est g -invariante.

(2) Comme g^m est une bijection affine croissante de chaque intervalle I_a sur $[0, 1]$, on a, de la même façon, $\lambda(g^{-m}(A) \cap I_a) = \lambda(A)\lambda(I_a)$.

(3) Pour $n \geq m$, on a encore $\lambda(g^{-n}(A) \cap I_a) = \lambda(g^{-(n-m)}(A)\lambda(I_a) = \lambda(A)\lambda(I_a)$. Donc $\lim_{n \rightarrow \infty} \lambda(g^{-n}(A) \cap I_a) = \lambda(A)\lambda(I_a)$. Comme les parties I_a engendrent la σ -algèbre des boréliens de I , les fonctions caractéristiques $\mathbb{I}_{I_a}$ engendrent un sous-espace vectoriel dense de $\mathbb{L}^1(I, \lambda)$ (voir le lemme 9.8). On en déduit que g est mélangeante (voir la proposition 6.1).

Partie II (1) a) Il est clair que $q_1 = [\beta]$. Pour tout n dans $\mathbb{N}$, on a $1 - x_n \leq \beta^{-n}$. Donc, pour tout $n \geq 1$, on a $q_n \beta^{-n} = x_n - x_{n-1} \leq 1 - x_{n-1} \leq \beta^{-(n-1)}$. D'où $q_n \leq \beta$ et $q_n \leq [\beta]$.

b) On a $x_n = \sum_{i=0}^n q_i \beta^{-i}$. Donc $\sum_{i=1}^{\infty} q_i \beta^{-i} = \lim_{n \rightarrow \infty} x_n = 1$.

(2) a) Il suffit de montrer que la partie $Z = \{\omega \in X / w \leq q\}$ est fermée, car $X_\beta = \bigcap_{n \in \mathbb{N}} \sigma^{-n}(Z)$. Or si ω est dans le complémentaire ${}^c Z$ de Z , on peut trouver $i_0 \geq 1$ tel que $\omega_{i_0} > q_{i_0}$ et, pour tout $i < i_0$, $\omega_i = q_i$. Les mots ω' de X tels que $\omega'_i = \omega_i$ pour tout $i \leq i_0$ sont aussi dans ${}^c Z$. Donc ${}^c Z$ est ouvert et Z est fermé.

b) C'est clair.

(3) a) Notons $E_n = \{(\omega_1, \dots, \omega_n) \in \Lambda^n / \exists \omega' \in X_\beta, \forall i = 1, \dots, n, \omega'_i = \omega_i\}$, de sorte que $p_n = \text{Card } E_n$. Si $\omega_1 < q_1$, alors on a $\omega \in E_n \Leftrightarrow (\omega_2, \dots, \omega_n) \in E_{n-1}$. Si $\omega_1 = q_1$ et $\omega_2 < q_2$, alors on a $\omega \in E_n \Leftrightarrow (\omega_3, \dots, \omega_n) \in E_{n-2}$. Par récurrence, si $\omega_1 = q_1, \dots, \omega_{i-1} = q_{i-1}$ et $\omega_i < q_i$, alors on a $\omega \in E_n \Leftrightarrow (\omega_{i+1}, \dots, \omega_n) \in E_{n-i}$. Donc $p_n = \sum_{i=1}^{n-1} q_i p_{n-i} + 1$, pour $n \geq 2$. Il est clair que $p_1 = [\beta] + 1$.

b) Par récurrence, on a

$$p_n \leq p_1(n-1)\beta^n \left(\sum_{i=1}^n q_i \beta^{-i} \right) + 1 \leq p_1(n-1)\beta^n + 1 \leq p_1 n \beta^n .$$

c) Fixons $\gamma < \beta$, et $n_0 \geq 1$ tels que $\sum_{i=1}^{n_0} q_i \gamma^{-i} \geq 1$ et posons $c = \max_{i=1, \dots, n_0} p_i \gamma^i$. Montrons, par récurrence, que pour tout $n \geq 0$, on a $p_n \geq c \gamma^n$. En effet, c'est vrai pour $n \leq n_0$, et pour tout $n > n_0$, on a

$$p_n = \left(\sum_{i=1}^{n-1} q_i p_{n-i} \right) + 1 \geq c \gamma^n \left(\sum_{i=1}^{n_0} q_i \gamma^{-i} \right) \geq c \gamma^n .$$

On en déduit que, pour tout $\gamma < \beta$, on a $\gamma \leq \liminf_{n \rightarrow \infty} \sqrt[n]{p_n} \leq \limsup_{n \rightarrow \infty} \sqrt[n]{p_n} \leq \beta$. Donc $\lim_{n \rightarrow \infty} \sqrt[n]{p_n} = \beta$.

d) Pour tout u dans Λ , notons $X_\beta^u = \{\omega \in X_\beta / \omega_0 = u\}$. Notons $\mathcal{O}_1$ le recouvrement ouvert $\mathcal{O}_1 = \{X_\beta^0, \dots, X_\beta^{[\beta]}\}$ et $\mathcal{O}_n = \bigvee_{i=0}^{n-1} \sigma_\beta^{-i} \mathcal{O}_1$. Notons que $\mathcal{O}_1$ est aussi une partition mesurable et que $N(\mathcal{O}_n) = p_n$. Le recouvrement ouvert $\mathcal{O}_1$ est générateur pour σ_β . Par le théorème 10.9,

$$h_{\text{top}}(\sigma_\beta) = h(\sigma_\beta, \mathcal{O}_1) = \lim_{n \rightarrow \infty} \frac{1}{n} \log N(\mathcal{O}_n) = \lim_{n \rightarrow \infty} \log p_n = \log \beta .$$

[Partie III] (1) a) Cela résulte de la construction, en raisonnant par récurrence. Si f^{n-1} est une bijection affine croissante de $[x_{n-1}, 1[$ sur $[0, y_{n-1}[$, alors $f^{n-1}(x_n) = \beta^{n-1}(x_n - x_{n-1}) = q_n \beta^{-1}$. Comme q_n est le plus grand entier tel que $q_n \beta^{-1} \leq y_{n-1}$, l'application f est une bijection affine croissante de $[q_n \beta^{-1}, y_{n-1}[$ sur $[0, \beta y_{n-1} - q_n[= [0, y_n$ et f^n est une bijection affine croissante de $[x_n, 1[$ sur $[0, y_n[$.

b) Par a),

$$\mu(A) = C \sum_{n \in \mathbb{N}} \beta^{-n} \lambda(A \cap [0, y_n]) = C \sum_{n \in \mathbb{N}} \lambda(f^{-n}(A) \cap [x_n, 1]) .$$

c) Découpons l'intervalle $[x_{n-1}, x_n[$ en q_n intervalles $I_k = [x_{n-1} + \frac{k}{\beta^n}, x_{n-1} + \frac{k+1}{\beta^n}[$, avec $0 \leq k \leq q_n$. D'après a), l'application $f^n : I_k \rightarrow [0, 1[$ est une bijection affine de pente β^n . On a donc $\lambda(f^{-1}(A) \cap I_k) = \beta^{-n} \lambda(A)$ et

$$\lambda(f^{-n}(A) \cap [x_{n-1}, x_n]) = \sum_{k=0}^{q_n-1} \lambda(f^{-1}(A) \cap I_k) = \beta^{-n} q_n \lambda(A) .$$

d) On a

$$\begin{aligned} \mu(f^{-1}(A)) &= C \sum_{n \in \mathbb{N} - \{0\}} \lambda(f^{-n}(A) \cap [x_{n-1}, 1]) = \\ C \sum_{n \in \mathbb{N} - \{0\}} \lambda(f^{-n}(A) \cap [x_{n-1}, x_n]) &+ C \sum_{n \in \mathbb{N} - \{0\}} \lambda(f^{-n}(A) \cap [x_n, 1]) = \\ C \left(\sum_{n \in \mathbb{N} - \{0\}} \beta^{-n} q_n \right) \lambda(A) &+ (\mu(A) - C \lambda(A)) = \mu(A) . \end{aligned}$$

(2) a) L'application F est la transformation étudiée dans la première partie, avec q_1 intervalles de taille β^{-1} suivis de q_2 intervalles de taille β^{-2} , etc. Elle préserve donc λ et est mélangeante pour λ . Par conséquent, F est ergodique pour λ .

b) Soit A un borélien de I tel que $f^{-1}(A) = A$. On a

$$F^{-1}(A) = \bigcup_{n \in \mathbb{N} - \{0\}} f^{-n}(A) \cap [x_{n-1}, x_n[= \bigcup_{n \in \mathbb{N} - \{0\}} A \cap [x_{n-1}, x_n[= A.$$

Comme F est ergodique pour λ , on a $\lambda(A) = 0$ ou $\lambda(A) = 1$. Donc $\mu(A) = 0$ ou $\mu(A) = 1$. Donc f est ergodique pour μ .

c) Soit μ' une mesure de probabilité f -invariante sur I absolument continue par rapport à λ . Par le théorème de Radon-Nykodym, on peut écrire $\mu' = \psi\mu$ avec ψ dans $\mathbb{L}^1(I, \mu)$. Comme μ' est f -invariante, on a l'égalité $\psi \circ f = \psi$ dans $\mathbb{L}^1(I, \mu)$. Comme f est ergodique pour μ , l'application ψ est constante μ -presque partout, et $\mu' = \mu$ car ces mesures sont de probabilité.

Partie IV (1) a) Par construction, on a

$$\sum_{i=1}^{\infty} \Omega_i(x) \beta^{-i} = x \quad (*)$$

et, pour tout n dans $\mathbb{N}$, on a $\sum_{i=1}^{\infty} \Omega_{n+i}(x) \beta^{-i} < 1$. On en déduit que $\sigma^n(\Omega(x)) \leq q$ et donc que $\Omega(x)$ appartient à X_β . La formule (*) prouve aussi l'injectivité de Ω .

b) Cela résulte de l'ergodicité de f et du théorème ergodique de Birkhoff. En outre, on a $\ell_\beta = \mu([0, \beta^{-1}[)$.

(2) a) Les éléments B de la partition $\alpha_n = \bigvee_{k=0}^{n-1} f^{-k}(\alpha)$ sont des intervalles sur lesquels l'application f^n est une application affine de pente β^n . Donc $\lambda(B) \leq \beta^{-n}$.

b) On a donc, pour tout β dans α_n , l'inégalité $\mu(B) \leq D\beta^{-n}$ avec $D = \frac{C}{1-\beta^{-1}}$. Donc

$$H_\mu(\alpha_n) = - \sum_{B \in \alpha_n} \mu(B) \log \mu(B) \geq \left(\sum_{B \in \alpha_n} \mu(B) \right) (n \log \beta - \log D) = n \log \beta - \log D,$$

et

$$h_\mu(f) \geq h_\mu(f, \alpha) = \lim_{n \rightarrow \infty} \frac{1}{n} H_\mu(\alpha_n) \geq \log \beta.$$

c) Cela résulte de l'égalité $\Omega \circ f = \sigma_\beta \circ \Omega$.

d) La partition mesurable $\mathcal{O}_1$ de X_β introduite en II (3) d) vérifie $\Omega^{-1}\mathcal{O}_1 = \alpha$. On a donc

$$\frac{1}{n} H_\nu(\mathcal{O}_n) = \frac{1}{n} H_\mu(\Omega^{-1}\mathcal{O}_n) = \frac{1}{n} H_\mu(\alpha_n).$$

Prenons la limite de ces égalités quand n tend vers l'infini. Comme la partition mesurable $\mathcal{O}_1$ est génératrice pour σ_β , et que α est génératrice pour f , on obtient, par le théorème de Kolmogorov-Sinai,

$$h_\nu(\sigma_\beta) = h_\nu(\sigma_\beta, \mathcal{O}_1) = \lim_{n \rightarrow \infty} \frac{1}{n} H_\nu(\mathcal{O}_n) = \lim_{n \rightarrow \infty} \frac{1}{n} H_\mu(\alpha_n) = h_\mu(f, \alpha) = h_\mu(f).$$

e) Par le principe variationnel et les questions précédentes, on a

$$\log \beta \leq h_\mu(f) = h_\nu(\sigma_\beta) \leq h_{\text{top}}(\sigma_A) = \log \beta.$$

Donc $h_\mu(f) = \log \beta$.

(3) a) Comme on a $1 = \beta^{-1} + \beta^{-2}$, on obtient $q_1 = q_2 = 1$, et $q_n = 0$ pour tout $n \geq 3$.

b) On a $\varphi = C(\mathbb{I}_{[0,1[} + \mathbb{I}_{[0,\beta^{-1}[})$ avec $C = \frac{1}{1-\beta^{-2}}$. Donc $\ell_\beta = \mu([0, \beta^{-1}[) = \frac{1}{1-\beta^{-2}} = \frac{5+\sqrt{5}}{10} \simeq 0,723$.

c) On a

$$X_\beta = \{\omega \in X / \forall n \geq 0, (\omega_n = \omega_{n+1} = 1) \Rightarrow (\forall i \geq 2, \omega_{n+i} = 0)\}.$$

Posons $Y' = \{\omega \in X / \forall n \geq 0, (\omega_n = 1) \Rightarrow (\omega_{n+1} = 0)\}$. Le système dynamique $(Y', \sigma|_{Y'})$ est un sous-décalage de type fini, associé à la matrice $A = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$. Comme Y' est un fermé de X_β qui contient $\Omega(I)$, le support Y de ν est contenu dans Y' . D'après les questions précédentes, la mesure ν est une mesure de probabilité σ -invariante sur Y' d'entropie maximale. C'est donc la mesure de parry et on a $Y' = Y$.

Exercice E.11.7 Sauf dans les questions I (3) et I (4), nous noterons pour simplifier $S_n \varphi = S_n^f \varphi$, $N(n, \epsilon) = N_{f, \varphi}(n, \epsilon)$, $N'(n, \epsilon) = N'_{f, \varphi}(n, \epsilon)$, $P(f) = P(f, \varphi)$ et $P_\mu(f) = P_\mu(f, \varphi)$.

[Partie I] (1) Les nombres $N(n, \epsilon)$, $N'(n, \epsilon)$ sont finis par compacité de X . Notons que si $\epsilon \leq \epsilon'$, alors une partie ϵ' -séparée est ϵ -séparée, et une partie ϵ -dense est ϵ' -dense, donc $N(n, \epsilon) \geq N(n, \epsilon')$ et $N'(n, \epsilon) \geq N'(n, \epsilon')$.

Les points d'une partie 2ϵ -séparée pour d_n^f sont dans des boules distinctes de rayon ϵ pour d_n^f . Toute partie ϵ -séparée E pour d_n^f rendant la somme $\sum_{x \in E} e^{S_n \varphi(x)}$ maximale, qui existe par compacité, est ϵ -dense pour d_n^f . Donc $N'(n, 2\epsilon) \leq N(n, \epsilon) \leq N'(n, \epsilon)$. En prenant la limite supérieure sur les n , puis la limite sur les ϵ (qui existe par monotonie), le résultat s'en déduit.

(2) Si $\varphi = 0$, alors $N(n, \epsilon)$ est le cardinal minimal d'une partie ϵ -dense de X pour d_n^f et $N'(n, \epsilon)$ est le cardinal maximal d'une partie ϵ -séparée pour d_n^f , et le premier résultat découle de la proposition 10.5.

Posons $\|\varphi\|_\infty = \sup_X |\varphi|$, et remarquons que $\|S_n \varphi\|_\infty \leq n \|\varphi\|_\infty$. Donc $N_{f, \varphi}(n, \epsilon) \leq e^{n \|\varphi\|_\infty} N_{f, 0}(n, \epsilon)$ et en prenant le logarithme et en divisant par n , puis en prenant les limites, on a $P(f) \leq h_{\text{top}}(f) + \|\varphi\|_\infty$. Comme μ est une mesure de probabilité, on a $P_\mu(f) \leq h_\mu(f) + \|\varphi\|_\infty$.

(3) Comme μ est f -invariante, on a $\int_X S_k^f \varphi d\mu = k \int_X \varphi d\mu$. On sait que $h_\mu(f^k) = k h_\mu(f)$, donc $P_\mu(f^k, S_k^f \varphi) = k P_\mu(f, \varphi)$.

Remarquons que $S_n^{f^k} S_k^f \varphi = S_{nk}^f \varphi$ et que $d_n^{f^k} \leq d_{nk}^f$.

Soit E une partie ϵ -dense pour d_{nk}^f . Alors E est aussi ϵ -dense pour $d_n^{f^k}$ et

$$\sum_{x \in E} e^{S_n^{f^k} S_k^f \varphi(x)} = \sum_{x \in E} e^{S_{nk}^f \varphi(x)}. \quad (*)$$

Donc $N_{f^k, S_k^f \varphi}(n, \epsilon) \leq N_{f, \varphi}(nk, \epsilon)$. Par conséquent, $P(f^k, S_k^f \varphi) \leq k P(f, \varphi)$.

Soit E une partie ϵ -séparée pour $d_n^{f^k}$. Alors E est aussi ϵ -séparée pour d_{nk}^f . Par la formule (*), on a donc $N'_{f^k, S_k^f \varphi}(nk, \epsilon) \leq N'_{f, \varphi}(nk, \epsilon)$. Par conséquent, $k P(f, \varphi) \leq P(f^k, S_k^f \varphi)$.

Ceci montre que $P(f^k, S_k^f \varphi) = kP(f, \varphi)$.

(4) Comme $\|S_n^f(\psi \circ f - \psi)\|_\infty = \|\psi \circ f^n - \psi\|_\infty \leq 2\|\psi\|_\infty$, on a

$$\frac{1}{n} \log N_{f, \varphi}(n, \epsilon) - \frac{2\|\psi\|_\infty}{n} \leq \frac{1}{n} \log N_{f, \varphi'}(n, \epsilon) \leq \frac{1}{n} \log N_{f, \varphi}(n, \epsilon) + \frac{2\|\psi\|_\infty}{n},$$

ce qui montre que $P(f, \varphi) = P(f, \varphi')$.

Par f -invariance de μ , on a $\int \psi \circ f d\mu = \int \psi d\mu$, donc $\int \varphi d\mu = \int \varphi' d\mu$ et $P_\mu(f, \varphi) = P_\mu(f, \varphi')$.

Partie II (1) On peut supposer que les p_i sont strictement positifs. On pose $t_i = \frac{e^{a_i}}{p_i} > 0$. Par la concavité de la fonction $t \mapsto \log t$, comme $\sum_{i=1}^k p_i = 1$, on a

$$\sum_{i=1}^k p_i \log t_i \leq \log \left(\sum_{i=1}^k p_i t_i \right),$$

avec égalité si $t_i = A$ pour tout i . Le résultat en découle.

(2) Par compacité faible de $\mathcal{M}(X)$, la suite des mesures de probabilité boréliennes μ_n converge faiblement, quitte à extraire, vers une limite μ . Comme $f_*\mu_n - \mu_n = \frac{1}{n}((f^n)_*\nu_n - \nu_n)$, la mesure μ est bien f -invariante.

(3) Posons $E_n = \{x_1, \dots, x_k\}$, $a_i = S_n \varphi(x_i)$ et $p_i = \nu_n(\{x_i\})$, de sorte que $p_i = \frac{e^{a_i}}{\sum e^{a_i}}$. Chaque élément de α_n contient au plus un élément de E_n . Donc, par le cas d'égalité de la question II (1), on a

$$H_{\nu_n}(\alpha_n) + \int S_n \varphi d\nu_n = \sum_{i=1}^k -p_i \log p_i + p_i a_i = \log \sum_{x \in E_n} e^{S_n \varphi(x)}.$$

(4) Notons que si $0 \leq i < q < n$, alors

$$\alpha_n = \left(\bigvee_{j=0}^{E[\frac{n}{q}]-1} f^{-jq-i} \alpha_q \right) \vee \left(\bigvee_{k=0}^{i-1} f^{-k} \alpha \right) \vee \left(\bigvee_{k=qE[\frac{n}{q}]-q+i+1}^{n-1} f^{-k} \alpha \right).$$

Par sous-additivité de l'entropie et majoration triviale de celle-ci, on a

$$H_{\nu_n}(\alpha_n) \leq \sum_{j=0}^{E[\frac{n}{q}]-1} H_{\nu_n}(f^{-jq-i} \alpha_q) + 2q \log(\text{Card } \alpha).$$

En sommant sur $i = 0, \dots, q-1$, on obtient

$$qH_{\nu_n}(\alpha_n) \leq \sum_{k=0}^{n-1} H_{\nu_n}(f^{-k} \alpha_q) + 2q^2 \log(\text{Card } \alpha).$$

Par la concavité de l'entropie métrique (comme fonction de la mesure), on a

$$\frac{q}{n} H_{\nu_n}(\alpha_n) \leq H_{\mu_n}(\alpha_q) + \frac{2q^2}{n} \log(\text{Card } \alpha).$$

Donc, comme $\int S_n \varphi d\nu_n = n \int \varphi d\nu_n$, on a

$$\frac{q}{n} \log \sum_{x \in E_n} e^{S_n \varphi(x)} = \frac{q}{n} H_{\nu_n}(\alpha_n) + q \int \varphi d\mu_n \leq H_{\mu_n}(\alpha_q) + \frac{2q^2}{n} \log(\text{Card } \alpha) + q \int \varphi d\mu_n .$$

(5) On a donc, pour tout $q > 0$, comme μ_n converge faiblement vers μ , et comme le bord des éléments de α est négligeable pour la mesure limite μ des μ_n ,

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \sum_{x \in E_n} e^{S_n \varphi(x)} \leq \lim_{n \rightarrow \infty} \left(\frac{1}{q} H_{\mu_n}(\alpha_q) + \int \varphi d\mu_n \right) = \frac{1}{q} H_{\mu}(\alpha_q) + \int \varphi d\mu .$$

Ceci étant vrai pour tout q , par définition de $h_{\mu}(f, \alpha)$, on a

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \sum_{x \in E_n} e^{S_n \varphi(x)} \leq h_{\mu}(f, \alpha) + \int \varphi d\mu \leq h_{\mu}(f) + \int \varphi d\mu = P_{\mu}(f, \varphi) .$$

Pour tout $\epsilon > 0$ et $n \geq 1$, soit E_n une partie ϵ -séparée pour la distance d_n^f telle que $\sum_{x \in E_n} e^{S_n^f \varphi(x)} = N'(n, \epsilon)$ (qui existe d'après la solution de la question I (1), mais on peut aussi remplacer " $= N'(n, \epsilon)$ " par " $\geq N'(n, \epsilon) - \epsilon$ "). On a vu dans le cours qu'il existe une partition mesurable α telle que $\text{diam } A < \epsilon$ et $\mu(\partial A) = 0$ pour tout élément A de α . Alors par ce qui précède, il existe μ dans $\mathcal{M}(X)^f$ telle que

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log N'(n, \epsilon) \leq P_{\mu}(f, \varphi) \leq \sup_{\nu \in \mathcal{M}(X)^f} P_{\nu}(f, \varphi) .$$

En prenant la limite quand ϵ tend vers 0, on a bien $P(f, \varphi) \leq \sup_{\mu \in \mathcal{M}(X)^f} P_{\mu}(f, \varphi)$.

(6) Pour $C \in \beta_n$, soit x_C un point adhérent à C tel que $\sup_{x \in C} S_n \varphi(x) = S_n \varphi(x_C)$. Soit y_C un point de E tel que $d_n^f(x_C, y_C) \leq \epsilon$, donc tel que $S_n \varphi(x_C) \leq S_n \varphi(y_C) + n$. La première hypothèse sur ϵ entraîne que, pour tout y dans E , le nombre d'éléments C de β_n tels que $y_C = y$ est au plus 2^n . Donc, en utilisant la question (1) pour la seconde inégalité, on a

$$\begin{aligned} H_{\mu}(\beta_n) + \int S_n \varphi &\leq \sum_{C \in \beta_n} \mu(C) (-\log \mu(C) + S_n \varphi(x_C)) \leq \log \left(\sum_{C \in \beta_n} e^{S_n \varphi(x_C)} \right) \\ &\leq \log \left(\sum_{C \in \beta_n} e^{S_n \varphi(y_C) + n} \right) \leq \log \left(2^n \sum_{y \in E} e^{S_n \varphi(y) + n} \right) . \end{aligned}$$

(7) Donc

$$\frac{1}{n} H_{\mu}(\beta_n) + \int \varphi d\mu \leq 1 + \log 2 + \frac{1}{n} \log \sum_{x \in E} e^{S_n \varphi(x)} \leq 1 + \log 2 + \frac{1}{n} \log \sum_{x \in E} e^{S_n \varphi(x)} .$$

En prenant la borne inférieure sur toutes les parties ϵ -denses pour d_n^f , on a

$$\frac{1}{n} H_{\mu}(\beta_n) + \int \varphi d\mu \leq 1 + \log 2 + \frac{1}{n} \log N(n, \epsilon) .$$

En prenant la limite supérieure quand n tend vers l'infini, puisque $\limsup_{n \rightarrow \infty} \frac{1}{n} N(n, \epsilon) \leq P(f, \varphi)$ par décroissance en ϵ de $N(n, \epsilon)$, on a

$$h_\mu(f, \alpha) + \int \varphi d\mu \leq h_\mu(f, \beta) + 1 + \int \varphi d\mu \leq 2 + \log 2 + P(f, \varphi) .$$

En prenant la borne supérieure sur les partitions mesurables α , comme une partition β satisfaisant les conditions de la question (6) existe par le cours, on a donc $P_\mu(f, \varphi) \leq 2 + \log 2 + P(f, \varphi)$.

(8) En appliquant cette formule en remplaçant f, φ par $f^k, S_k \varphi$, en utilisant la question I (3), en divisant par k et en faisant tendre k vers $+\infty$, on obtient $P_\mu(f, \varphi) \leq P(f, \varphi)$. Ceci étant vrai pour tout μ dans $\mathcal{M}(X)^f$, on a $\sup_{\mu \in \mathcal{M}(X)^f} P_\mu(f, \varphi) \leq P(f, \varphi)$. Avec l'inégalité inverse traitée dans la question (5), ceci montre le résultat.

Partie III Pour simplifier, on note $X = X_A$ et $f = \sigma_A$, ce qui permet d'utiliser les notations du début du corrigé.

(1) Soient x, y dans X tels que $x_i = y_i$ pour $i = 0, \dots, n$. Comme $f^k(x)_j = x_{j+k}$, on a

$$S_n \varphi(x) = \sum_{k=0}^{n-1} \varphi(f^k(x)) = \sum_{k=0}^{n-1} \varphi_{f^k(x)_0, f^k(x)_1} = \sum_{k=0}^{n-1} \varphi_{x_k, x_{k+1}} = S_n \varphi(y) .$$

Soient x, y dans X tels que $d_n^f(x, y) < \frac{1}{2}$. Par définition de la distance d , pour tous x', y' dans X_A , on a $d(x', y') < \frac{1}{2^p}$ si et seulement si $x'_i = y'_i$ pour $|i| \leq p$. Donc, par définition de d_n^f , on a $x_i = y_i$ pour $i = 0, \dots, n$. Par ce qui précède, on a donc $S_n \varphi(x) = S_n \varphi(y)$.

(2) Soit E une partie ϵ -séparée pour d_n^f telle que $\sum_{x \in E} S_n \varphi(x)$ est maximal (voir la question I (1) pour l'existence). Soit F une partie $\frac{1}{2}$ -séparée pour d_n^f telle que $\sum_{x \in F} S_n \varphi(x)$ est maximal. Par maximalité, pour tout x dans E , il existe y_x dans F tel que $d_n^f(x, y_x) < \frac{1}{2}$. Par la question précédente, on a $S_n \varphi(x) = S_n \varphi(y_x)$. Donc

$$\sum_{x \in E} S_n \varphi(x) \leq \sum_{y \in F} S_n \varphi(y) \text{ Card } E_y$$

avec $E_y = \{x \in E / y_x = y\}$. Il suffit donc de montrer que le cardinal de E_y est borné par une constante qui ne dépend que de ϵ .

Soit N dans $\mathbb{N}$ tel que si $\sup_{|i| \leq N} d(f^i(x), f^i(y)) < 1$, alors $d(x, y) < \epsilon$ (l'entier $N = E(-\log_2 \epsilon) + 1$ convient). Soit $\alpha > 0$ tel que si $d(x, y) < \alpha$, alors $\sup_{|i| \leq N} d(f^i(x), f^i(y)) < \epsilon$.

Soit $n > 2N$. Pour tous les x, x' dans E_y avec $x \neq x'$, on a $d_n^f(x, x') < 1$ par inégalité triangulaire, donc $d(f^i(x), f^i(x')) < \epsilon$ pour $i = N, N+1, \dots, n-N$. On a $d(x, x') \geq \alpha$ ou $d(f^n(x), f^n(x')) \geq \alpha$, sinon par définition de N et α , on aurait $d(f^i(x), f^i(x')) < \epsilon$ pour $i = 0, \dots, N-1$ et $d(f^i(x), f^i(x')) < \epsilon$ pour $i = n-N, \dots, n-1$, et on obtiendrait une contradiction au fait que E est ϵ -séparé pour d_n^f .

Donc E_y s'envoie injectivement par $x \mapsto (x, f^n(x))$ sur une partie E' de $X \times X$ formée de points deux à deux à distance (pour la distance produit $d((a, b), (a', b)) = \sup\{d(a, a'), d(b, b')\}$) au moins α . Par compacité, le cardinal d'une telle partie E' est borné par une constante ne dépendant que de α , donc de ϵ .

(3) Comme $\limsup_{n \rightarrow \infty} \frac{\log C_\epsilon}{n} = 0$ et $N'(n, \epsilon) \geq N'(n, \frac{1}{2})$ si $\epsilon \leq \frac{1}{2}$, le résultat découle de la définition de $P(f, \varphi)$.

(4) Comme A est transitive, il en est de même de A^φ , et le résultat découle du théorème de Perron-Frobenius.

(5) Pour tous les $\alpha_0, \dots, \alpha_n$ dans Λ tels que $A_{\alpha_i, \alpha_{i+1}} = 1$ pour $i = 0, \dots, n-1$, comme A est transitive, il existe x dans X_A tel que $x_i = \alpha_i$ pour $i = 0, \dots, n$. Donc les quantités sont bien définies. On a $e^{(S_n \varphi)_{\alpha_0, \dots, \alpha_n}} = \prod_{i=0}^{n-1} e^{\varphi_{\alpha_i, \alpha_{i+1}}} = \prod_{i=0}^{n-1} A_{\alpha_i, \alpha_{i+1}}^\varphi$. Le premier résultat s'en déduit par sommation.

Si E est une partie $\frac{1}{2}$ -séparée pour d_n^f , alors pour tous les x, y distincts dans E , on a $(x_{-1}, x_0, \dots, x_n) \neq (y_{-1}, y_0, \dots, y_n)$, donc

$$\sum_{x \in E} e^{S_n \varphi(x)} \leq (\text{Card } \Lambda) \sum_{(\alpha_0, \dots, \alpha_n) \in \Lambda^{n+1} / A_{\alpha_i, \alpha_{i+1}} = 1} e^{(S_n \varphi)_{\alpha_0, \dots, \alpha_n}} = (\text{Card } \Lambda) \sum_{\alpha, \beta \in \Lambda} (A^\varphi)_{\alpha\beta}^n .$$

Il existe aussi une partie E qui est $\frac{1}{2}$ -séparée pour d_n^f et pour laquelle l'inégalité dans l'autre sens est vraie en omettant la constante $\text{Card } \Lambda$. Donc, comme deux normes matricielles sont équivalentes, et puisque A^φ est une matrice de Perron-Frobenius, on a

$$\begin{aligned} P(\sigma_A, \varphi) &= \lim_{n \rightarrow \infty} \frac{1}{n} \log N_{\sigma_A, \varphi}(\frac{1}{2}, n) = \lim_{n \rightarrow \infty} \frac{1}{n} \log \sum_{\alpha, \beta \in \Lambda} (A^\varphi)_{\alpha\beta}^n \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} \log \|(A^\varphi)^n\| = \log(\lambda^\varphi) . \end{aligned}$$

(6) Comme f^φ est un vecteur propre de ${}^t A^\varphi$, on a $\sum_{\alpha \in \Lambda} A_{\alpha\beta}^\varphi f_\alpha = \lambda^\varphi f_\beta$, donc la matrice à coefficients positifs ou nuls π est bien stochastique.

Si $w = (w_\alpha)_{\alpha \in \Lambda}$ est le vecteur propre de π , associé à la valeur propre 1, vérifiant $\sum_{\alpha \in \Lambda} w_\alpha = 1$, alors on sait que (voir paragraphe 11.2)

$$h_{\mu^\varphi}(\sigma_A) = - \sum_{\alpha, \beta \in \Lambda} \pi_{\alpha\beta} \log(\pi_{\alpha\beta}) w_\beta ,$$

et que si C est le cylindre $C_{\alpha_m, \dots, \alpha_n}^{m, \dots, n}$, alors

$$\mu(C) = \left(\prod_{i=m}^{n-1} \pi_{\alpha_i \alpha_{i+1}} \right) w_{\alpha_n} .$$

Par ce qui précède, on a $w_\alpha = v_\alpha f_\alpha$. Comme φ ne dépend que des deux premières coordonnées, on a donc

$$\int \varphi d\mu^\varphi = \sum_{\alpha, \beta \in \Lambda} \varphi_{\alpha\beta} \mu^\varphi(C_{\alpha, \beta}^{0,1}) = \sum_{\alpha, \beta \in \Lambda} \varphi_{\alpha\beta} \frac{A_{\alpha\beta}^\varphi f_\alpha v_\beta}{\lambda^\varphi} .$$

Donc

$$\begin{aligned} P_{\mu^\varphi}(\sigma_A, \varphi) &= h_{\mu^\varphi}(\sigma_A) + \int \varphi d\mu^\varphi = \sum_{\alpha, \beta \in \Lambda} (-\log(A_{\alpha\beta}^\varphi f_\alpha) + \log f_\beta + \log(\lambda^\varphi) + \varphi_{\alpha\beta}) \frac{A_{\alpha\beta}^\varphi f_\alpha v_\beta}{\lambda^\varphi} \\ &= \sum_{\alpha, \beta \in \Lambda} (-\log(A_{\alpha\beta}^\varphi f_\alpha) + \log f_\beta + \log(\lambda^\varphi)) \frac{A_{\alpha\beta}^\varphi f_\alpha v_\beta}{\lambda^\varphi} . \end{aligned}$$

Or

$$\sum_{\alpha, \beta} -\log(A_{\alpha\beta} f_{\alpha}) \frac{A_{\alpha\beta}^{\varphi} f_{\alpha} v_{\beta}}{\lambda^{\varphi}} = -\frac{1}{\lambda^{\varphi}} \sum_{\alpha, \beta} \log(f_{\alpha}) A_{\alpha\beta}^{\varphi} f_{\alpha} v_{\beta} = -\sum_{\alpha} \log(f_{\alpha}) f_{\alpha} v_{\alpha} ,$$

$$\sum_{\alpha, \beta} \log(f_{\beta}) \frac{A_{\alpha\beta}^{\varphi} f_{\alpha} v_{\beta}}{\lambda^{\varphi}} = \sum_{\beta} \log(f_{\beta}) f_{\beta} v_{\beta} ,$$

et

$$\sum_{\alpha, \beta} \log(\lambda^{\varphi}) \frac{A_{\alpha\beta}^{\varphi} f_{\alpha} v_{\beta}}{\lambda^{\varphi}} = \log(\lambda^{\varphi}) \sum_{\alpha} f_{\alpha} v_{\alpha} = \log \lambda^{\varphi} = P(\sigma_A, \varphi) ,$$

par la question (5). Le résultat en découle.

(7) Comme l'ensemble des combinaisons linéaires finies de fonctions caractéristiques de cylindres est contenu et dense dans l'espace des fonctions continues sur l'espace de Cantor $\Lambda^{\mathbb{Z}}$, il suffit de montrer que, pour tout cylindre $C = C_{\alpha_m, \dots, \alpha_n}^{m, \dots, n}$ de $\Lambda^{\mathbb{Z}}$, on a $\lim_{N \rightarrow \infty} \mu_N(C) = \mu^{\varphi}(C)$. Comme P_N est invariant par σ , on peut supposer que $m = 0$. D'après la question (5), on a $\sum_{x \in P_N} e^{S_N \varphi(x)} = \text{tr} (A^{\varphi})^N$ et, pour $N > n$, si $C \cap X_A$ est non vide, alors

$$\begin{aligned} \mu_N(C) &= \frac{\sum_{x \in P_N \cap C} e^{S_N \varphi(x)}}{\text{tr} (A^{\varphi})^N} \\ &= \frac{1}{\text{tr} (A^{\varphi})^N} \sum_{(\beta_{n+1}, \dots, \beta_{N-1}) \in \Lambda^{N-n-1} / A_{\beta_i, \beta_{i+1}} = 1} e^{(S_N \varphi)_{\alpha_0, \dots, \alpha_n, \beta_{n+1}, \dots, \beta_{N-1}, \alpha_0}} \\ &= \left(\prod_{i=m}^{n-1} A_{\alpha_i \alpha_{i+1}}^{\varphi} \right) (A^{\varphi})_{\alpha_n \alpha_m}^{N-n+m} . \end{aligned}$$

D'après le théorème de Perron-Frobenius, on a $\lim_{N \rightarrow \infty} \frac{1}{\lambda^N} (A^{\varphi})^N = \tau$, où τ est la matrice du projecteur sur la droite $\mathbb{R} v^{\varphi}$ parallèlement à l'hyperplan noyau de la forme linéaire f^{φ} , donc $\tau = (v_{\alpha} f_{\beta})_{\alpha, \beta \in \Lambda}$. En particulier, on a $\lim_{N \rightarrow \infty} \frac{1}{(\lambda^{\varphi})^N} \text{tr} (A^{\varphi})^N = 1$ et $\lim_{N \rightarrow \infty} \frac{1}{(\lambda^{\varphi})^N} ((A^{\varphi})^N)_{\alpha, \beta} = v_{\alpha} f_{\beta}$. D'où

$$\lim_{N \rightarrow \infty} \mu_N(C) = \left(\prod_{i=0}^{n-1} A_{\alpha_i \alpha_{i+1}}^{\varphi} \right) \frac{f_{\alpha_0} v_{\alpha_n}}{(\lambda^{\varphi})^n} = \left(\prod_{i=0}^{n-1} \pi_{\alpha_i \alpha_{i+1}} \right) f_{\alpha_n} v_{\alpha_n} = \mu^{\varphi}(C) .$$

C'est ce que l'on voulait.

(8) Comme vu ci-dessus, et puisque $C \cap X_A$ est non vide, on a

$$\begin{aligned} \mu^{\varphi}(C) &= \left(\prod_{i=m}^{n-1} A_{\alpha_i \alpha_{i+1}}^{\varphi} \right) \frac{f_{\alpha_m} v_{\alpha_n}}{(\lambda^{\varphi})^{n-m}} = \left(\prod_{i=m}^{n-1} A_{\alpha_i \alpha_{i+1}} \right) e^{\varphi_{\alpha_i, \alpha_{i+1}}} f_{\alpha_0} v_{\alpha_n} (\lambda^{\varphi})^{-n+m} \\ &= f_{\alpha_0} v_{\alpha_n} (\lambda^{\varphi})^{-n+m} e^{(S_{n-m} \varphi)_{\alpha_m, \dots, \alpha_n}} . \end{aligned}$$

Il suffit de prendre $c_1 = \min_{\alpha, \beta \in \Lambda} f_{\alpha} v_{\beta} > 0$ et $c_2 = \max_{\alpha, \beta \in \Lambda} f_{\alpha} v_{\beta} > 0$.

(9) Soit ν un élément de $\mathcal{M}(X_A)^{\sigma_A}$ tel que $P_{\nu}(\sigma_A, \varphi) \geq P_{\mu^{\varphi}}(\sigma_A, \varphi) = \log \lambda^{\varphi}$. Montrons que $\nu = \mu^{\varphi}$, ce qui par le principe variationnel pour la pression (partie II) assurera le résultat.

Si ν est absolument continue par rapport à μ^φ , alors il existe, d'après le théorème de Radon-Nokodym, une fonction ψ dans $\mathbb{L}^1(X_A, \mu^\varphi)$ telle que $\nu = \psi \mu^\varphi$. Cette fonction est σ_A -invariante μ^φ -presque partout. Comme μ^φ est ergodique (car toute mesure de Markov pour (X_A, σ_A) est mélangeante), et que ces mesures sont des mesures de probabilité, on a $\psi = 1$ et $\nu = \mu^\varphi$, ce qui est exactement ce que l'on cherche à démontrer.

Supposons donc par l'absurde que ν n'est pas absolument continue par rapport à μ^φ . Il existe donc un borélien B de X_A tel que $\nu(B) > 0$ et $\mu^\varphi(B) = 0$. Par un argument de densité déjà vu, on peut trouver une suite B_p d'unions finies disjointes de cylindres de X_A tels que $\lim_{p \rightarrow \infty} \nu(B_p) = \nu(B) \neq 0$ et $\lim_{p \rightarrow \infty} \mu^\varphi(B_p) = 0$.

Notons ξ_0 la partition mesurable de X_A définie par $\xi_0 = \{X_A \cap C_\alpha^0 / \alpha \in \Lambda\}$ et ξ_n la partition $\xi_n = \bigvee_{i=-n}^n \sigma_A^{-i} \xi_0$. Comme ξ_0 est une partition génératrice pour σ_A (cette transformation étant inversible), on a, d'après le théorème de Kolmogorov-Sinai,

$$h_\nu(\sigma_A) = h_\nu(\sigma_A, \xi_0) = \inf_{n \geq 1} \frac{1}{n} H_\nu \left(\bigvee_{i=0}^{n-1} \sigma^{-i} \xi_0 \right) = \inf_{n \geq 1} \frac{1}{2n} H_\nu(\xi_n) .$$

Comme par hypothèse $P_\nu(\sigma_A, \varphi) = \log \lambda^\varphi$, et puisque $2n \int \varphi d\nu = \int S_{2n} \varphi d\nu$, on a donc, pour tout $n \geq 1$, l'inégalité $H_\nu(\xi_n) - 2n \log \lambda^\varphi + \int S_{2n} \varphi d\nu \geq 0$.

Notons que ξ_n est la partition par les cylindres de la forme $C = C_{\alpha_{-n}, \dots, \alpha_n}^{-n, \dots, n}$, et que $S_{2n} \varphi$ est constante sur C , valant $S_{2n} \varphi|_C = (S_{2n} \varphi)_{\alpha_{-n}, \dots, \alpha_n}$. Fixons p et choisissons n assez grand pour que B_p soit une réunion finie disjointe d'éléments de ξ_n .

On a alors

$$\begin{aligned} 0 &\leq H_\nu(\xi_n) + \int S_{2n} \varphi d\nu - 2n \log \lambda^\varphi = \sum_{C \in \xi_n} -\nu(C) \log \nu(C) + \sum_{C \in \xi_n} S_{2n} \varphi|_C \nu(C) - 2n \log \lambda^\varphi \\ &\leq \sum_{C \in \xi_n, C \subset B_p} \nu(C) (-\log \nu(C) + S_{2n} \varphi|_C) + \sum_{C \in \xi_n, C \subset {}^c B_p} \nu(C) (-\log \nu(C) + S_{2n} \varphi|_C) - 2n \log \lambda^\varphi . \end{aligned}$$

En utilisant la question II (1), on a donc

$$0 \leq \nu(B_p) \log \frac{\sum_{C \in \xi_n, C \subset B_p} e^{S_{2n} \varphi|_C}}{\nu(B_p)} + \nu({}^c B_p) \log \frac{\sum_{C \in \xi_n, C \subset {}^c B_p} e^{S_{2n} \varphi|_C}}{\nu({}^c B_p)} - 2n \log \lambda^\varphi .$$

En utilisant la question (8) précédente, on a alors

$$\begin{aligned} 0 &\leq \nu(B_p) \log \frac{\mu(B_p)(\lambda^\varphi)^{2n}}{c_1 \nu(B_p)} + \nu({}^c B_p) \log \frac{\mu({}^c B_p)(\lambda^\varphi)^{2n}}{c_1 \nu({}^c B_p)} - 2n \log \lambda^\varphi \\ &= \nu(B_p) \log \frac{\mu(B_p)}{\nu(B_p)} + \nu({}^c B_p) \log \frac{\mu({}^c B_p)}{\nu({}^c B_p)} - \log c_1 . \end{aligned}$$

Cette inégalité est vraie pour tout p , mais le dernier terme tend vers $-\infty$ quand p tend vers l'infini, contradiction. Le résultat en découle.

12 Codage

Ce chapitre est entièrement consacré à notre exemple favori de “système dynamique hyperbolique”, le tore $\mathbb{T}^N$ muni d’un automorphisme linéaire hyperbolique f_M . Nous renvoyons par exemple à [KH] pour des (définitions et) résultats généraux sur les systèmes dynamiques hyperboliques, pour lesquels il est de toutes façons primordial de se familiariser avec cet exemple, en particulier en ce qui concerne les “variétés stables et instables”, le “lemme de pistage”, le “lemme de fermeture”, trois outils dont les avatars sont nombreux en dynamique hyperbolique.

Grâce à ces objets, nous définirons des “partitions de Markov” de $(\mathbb{T}^N, f_M)$ et nous montrerons qu’elles permettent de coder (au sens de l’introduction du chapitre précédent) notre système dynamique par un sous-décalage de type fini (X_A, σ_A) . Bien sûr, pour pouvoir coder, il nous faudra démontrer l’existence de partitions de Markov.

Comme application, nous montrerons que la mesure de Haar sur $\mathbb{T}^N$ est l’unique mesure d’entropie maximale pour f_M , et nous calculerons cette entropie (qui est donc l’entropie topologique de $(\mathbb{T}^N, f_M)$).

12.1 Variétés stables, lemme de pistage et lemme de fermeture

Dans tout ce chapitre, on reprend les notations du chapitre 7 sur la stabilité structurelle. En particulier, on note

- M une matrice $N \times N$ à coefficients entiers, hyperbolique (i.e. sans valeur propre de module 1), que l’on suppose de plus de déterminant égal à ± 1 (pour que f_M soit inversible), par exemple $M = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$;
- $E = E_s \oplus E_u$ la décomposition de $E = \mathbb{R}^N$ en sous-espaces stables et instables pour M ;
- $\|\cdot\|$ une norme sur E adaptée à M (i.e. telle que, si l’on note encore $\|\cdot\|$ la norme d’opérateur associée, alors $\|M|_{E_s}\| < 1$ et $\|M|_{E_u}\| < 1$ et, pour tous x_s dans E_s et x_u dans E_u , $\|x_s + x_u\| = \max\{\|x_s\|, \|x_u\|\}$) ; attention, il ne s’agit pas d’une norme euclidienne, mais d’une norme “carrée”, voir le paragraphe 12.3 pour un exemple ;
- $ch(M)$ la constante d’hyperbolicité de M

$$ch(M) = \max\{\|M|_{E_s}\|, \|M^{-1}|_{E_u}\|\} < 1 ;$$

- $p : \mathbb{R}^N \rightarrow \mathbb{T}^N$ la projection canonique ;
- $f = f_M$ l’automorphisme linéaire hyperbolique du tore $\mathbb{T}^N$ défini par $f_M \circ p = p \circ M$;
et
- d la distance quotient sur $\mathbb{T}^N$

$$d(x, y) = \inf\{\|v - w\| / p(v) = x, p(w) = y\} .$$

On fixe $\epsilon_0 > 0$ tel que p est injectif sur la boule $B(0, \epsilon_1)$ dans $\mathbb{R}^N$ de rayon

$$\epsilon_1 = 2\epsilon_0 \frac{\max\{\|M\|, \|M^{-1}\|\}}{1 - ch(M)} \geq \epsilon_0 .$$

Voir le paragraphe 12.3 pour un exemple explicite de telles constantes $ch(M)$, ϵ_1 , ϵ_0 .

Pour $\epsilon < \epsilon_0$ et x dans $\mathbb{T}^N$, on note

$$W_\epsilon^s(x) = \{y \in \mathbb{T}^N / \forall n \geq 0, d(f^n(x), f^n(y)) \leq \epsilon\}$$

la *variété stable locale* en x et

$$W_\epsilon^u(x) = \{y \in \mathbb{T}^N / \forall n \geq 0, d(f^{-n}(x), f^{-n}(y)) \leq \epsilon\}$$

la *variété instable locale* en x .

Rassemblons quelques propriétés élémentaires de ces parties.

Lemme 12.1 Soient $\epsilon < \epsilon_0$ et x, y dans $\mathbb{T}^N$.

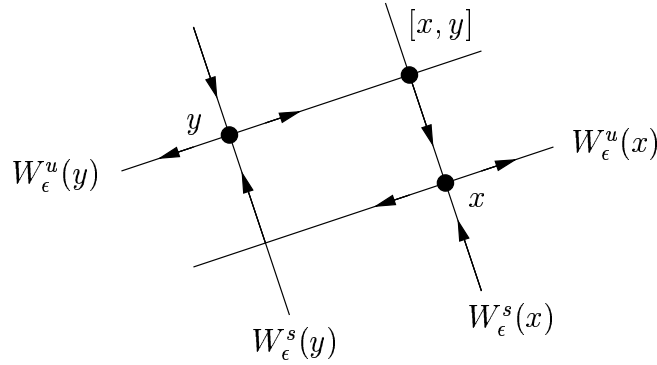
1. $f(W_\epsilon^s(x)) \subset W_\epsilon^s(f(x))$ et $W_\epsilon^u(f(x)) \subset f(W_\epsilon^u(x))$;
2. le point y appartient à $W_\epsilon^s(x)$ si et seulement si pour tout $n \geq 0$, on a

$$d(f^n(x), f^n(y)) \leq \epsilon ch(M)^n ;$$

3. le point y appartient à $W_\epsilon^u(x)$ si et seulement si pour tout $n \geq 0$, on a

$$d(f^{-n}(x), f^{-n}(y)) \leq \epsilon ch(M)^n ;$$

4. si $d(x, y) \leq \epsilon$, alors l'intersection $W_\epsilon^s(x) \cap W_\epsilon^u(y)$ est réduite à un point, noté $[x, y]$.
L'application $(x, y) \mapsto [x, y]$ est continue, et appelée *produit local*.



Preuve. On relève les points x, y de $\mathbb{T}^N$ en des points v, w de $\mathbb{R}^N$ tels que $\|v - w\| \leq \epsilon$. On a les égalités $W_\epsilon^s(x) = p(B(v, \epsilon) \cap (v + E_s))$ et $W_\epsilon^u(x) = p(B(v, \epsilon) \cap (v + E_u))$. Les quatre assertions s'en déduisent aussitôt. $\square$

Soit (X, f) un système dynamique topologique inversible, avec X un espace métrique. On appelle ϵ -pseudo-orbite de f toute suite $(x_n)_{n \in \mathbb{Z}}$ de X telle que

$$\sup_{n \in \mathbb{Z}} d(x_{n+1}, f(x_n)) \leq \epsilon .$$

On dit qu'une suite $(x_n)_{n \in \mathbb{Z}}$ de X est ϵ -*pistée* par l'orbite d'un point x de X si on a

$$\sup_{n \in \mathbb{Z}} d(x_n, f^n(x)) \leq \epsilon .$$

Énonçons maintenant le lemme de pistage (“shadowing lemma” en anglais) dans notre cas particulier.

Lemme 12.2 (Lemme de pistage) *Si $\epsilon \leq \epsilon_0$, alors toute ϵ -pseudo-orbite de f_M sur $\mathbb{T}^N$ est $\frac{\epsilon}{1-ch(M)}$ -pistée par une unique orbite de f_M sur $\mathbb{T}^N$.*

Remarque. L'unicité signifie que f_M est expansive, de constante d'expansivité $\delta_0 = \frac{\epsilon_0}{1-ch(M)}$, au sens défini au paragraphe 10.3.

Preuve. Pour montrer l'existence, on relève (par récurrence en fixant un relevé v_0 de x_0) la ϵ -pseudo-orbite $(x_n)_{n \in \mathbb{Z}}$ de f_M sur $\mathbb{T}^N$ en une ϵ -pseudo-orbite $(v_n)_{n \in \mathbb{Z}}$ de M sur $\mathbb{R}^N$ et on applique l'exercice E.7.3.

Pour montrer l'unicité, si y et y' sont deux points dont l'orbite $\frac{\epsilon}{1-ch(M)}$ -piste $(x_n)_{n \in \mathbb{Z}}$, alors

$$\forall n \in \mathbb{Z}, \quad d(f^n(y), f^n(y')) \leq \frac{2\epsilon}{1-ch(M)} .$$

Soit v un relevé de $y - y'$ par p dans la boule $B(0, \frac{2\epsilon}{1-ch(M)})$. Comme P est injectif sur $B(0, \epsilon_1)$, on montre par récurrence que les points $T^n v$ restent dans $B(0, \frac{2\epsilon}{1-ch(M)})$. Par l'hyperbolicité de M , ceci implique que $v = 0$, donc que $y = y'$. $\square$

Énonçons maintenant le lemme de fermeture (“closing lemma” en anglais) dans notre cas particulier. Pour des versions plus générales (dont le lemme de fermeture d'Anosov pour le flot géodésique des surfaces hyperboliques), voir par exemple [KH].

Ce lemme est un corollaire du lemme de pistage, qui permet de construire des orbites périodiques près de tout point récurrent.

Lemme 12.3 (Lemme de fermeture) *Pour tout $\epsilon < \epsilon_0$, pour tout x dans $\mathbb{T}^N$ et tout $n \geq 1$ tel que $d(x, f^n(x)) \leq \epsilon$, il existe un unique point y de $\mathbb{T}^N$ tel que $f^n(y) = y$ et*

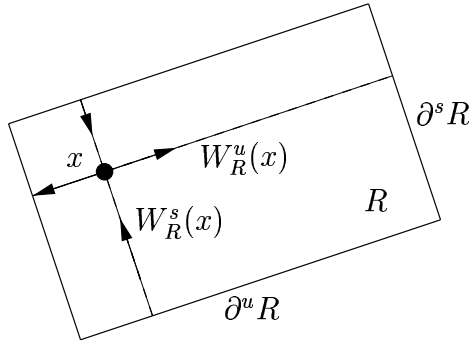
$$\forall i = 0, \dots, n-1, \quad d(f^i(x), f^i(y)) \leq \frac{\epsilon}{1-ch(M)} .$$

Preuve. On note $(x_k)_{k \in \mathbb{Z}}$ la ϵ -pseudo-orbite $x_k = f^{r_k}(x)$ où r_k est le reste de la division euclidienne de k par n . Le point y est forcément l'unique point dont l'orbite $\frac{\epsilon}{1-ch(M)}$ -piste la suite $(x_k)_{k \in \mathbb{Z}}$. Comme $f^n(y)$ vérifie la même propriété, on a $f^n(y) = y$. $\square$

12.2 Partition de Markov

On appelle *rectangle* de $\mathbb{T}^N$ pour $f = f_M$ toute partie R de diamètre au plus ϵ_0 telle que si $x, y \in R$, alors $[x, y] \in R$.

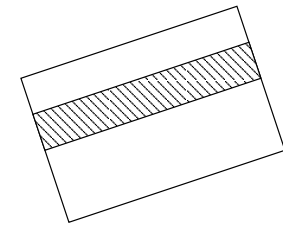
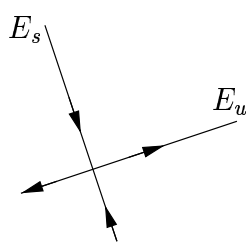
Autrement dit, dans une identification locale de $\mathbb{T}^N$ avec $E = E_s \times E_u$, on a $R = R^s \times R^u$ avec $\text{diam}(R^s) \leq \epsilon_0$ et $\text{diam}(R^u) \leq \epsilon_0$. Le bord de R s'écrit alors $\partial R = \partial^s R \cup \partial^u R$ où $\partial^s R = R^s \times \partial R^u$ et $\partial^u R = \partial R^s \times R^u$.



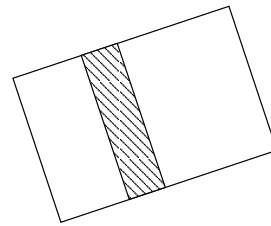
Le rectangle R est dit *propre* si $R = \overline{\overset{\circ}{R}}$, i.e. si R est l'adhérence de son intérieur. Pour x dans R , on note

$$W_R^s(x) = W_{\epsilon_0}^s(x) \cap R \quad \text{et} \quad W_R^u(x) = W_{\epsilon_0}^u(x) \cap R,$$

que l'on appelle les feuilles stables et instables dans R de x . Une partie S de R est appelée *sous-rectangle stable* (resp. *instable*) si pour tout x dans S , on a $W_R^s(x)$ (resp. $W_R^u(x)$) est contenu dans S .



sous-rectangle instable



sous-rectangle stable

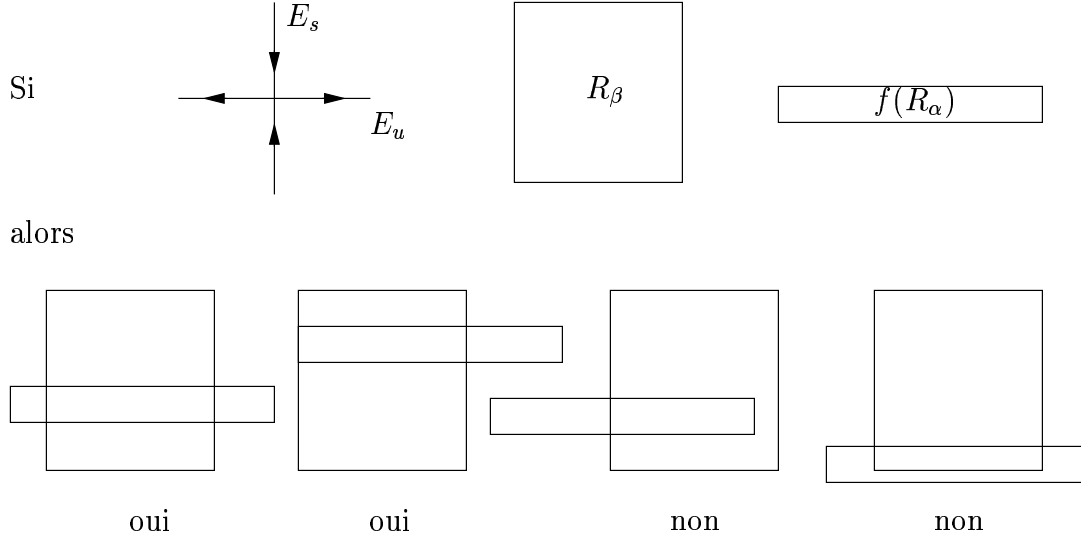
Une *partition de Markov* $\mathcal{R}$ est un recouvrement fini $\mathcal{R} = (R_\alpha)_{\alpha \in \Lambda}$ de $\mathbb{T}^N$ par des rectangles propres, tel que, pour tous les α, β dans Λ , on a

- a) $\overset{\circ}{R}_\alpha \cap \overset{\circ}{R}_\beta = \emptyset$ si $\alpha \neq \beta$,
- b) si $x \in \overset{\circ}{R}_\alpha$ et $f(x) \in \overset{\circ}{R}_\beta$, alors $f(W_{R_\alpha}^s(x)) \subset W_{R_\beta}^s(f(x))$ et $W_{R_\beta}^u(f(x)) \subset f(W_{R_\alpha}^u(x))$.

Remarque. La terminologie “partition de Markov” est classique, bien que $\mathcal{R}$ ne soit jamais une partition au sens strict.

Voici quelques dessins de configuration possibles et impossibles, pour illustrer la défi-

dition des partitions de Markov.



A chaque partition de Markov, on associe sa *matrice d'incidence* $A = (A_{\alpha\beta})_{\alpha,\beta \in \Lambda}$ où

$$A_{\alpha\beta} = \begin{cases} 1 & \text{si } f(\overset{\circ}{R}_\alpha) \cap \overset{\circ}{R}_\beta \neq \emptyset \\ 0 & \text{sinon} \end{cases}.$$

Théorème 12.4 Soient $\mathcal{R} = (R_\alpha)_{\alpha \in \Lambda}$ une partition de Markov pour $f = f_M$, A sa matrice d'incidence et (X_A, σ_A) le sous-décalage de type fini associé.

- a) Pour tout ω dans X_A , l'intersection $\bigcap_{n \in \mathbb{Z}} f^{-n}(R_{\omega_n})$ est réduite à un point $\pi(\omega)$.
- b) L'application $\pi : X_A \rightarrow \mathbb{T}^N$ est continue, surjective et vérifie $f \circ \pi = \pi \circ \sigma_A$.
- c) Pour toute mesure de probabilité σ_A -invariante et ergodique μ , de support X_A , on a

$$\mu(\{\omega \in X_A / \text{Card } \pi^{-1}(\pi(\omega)) > 1\}) = 0.$$

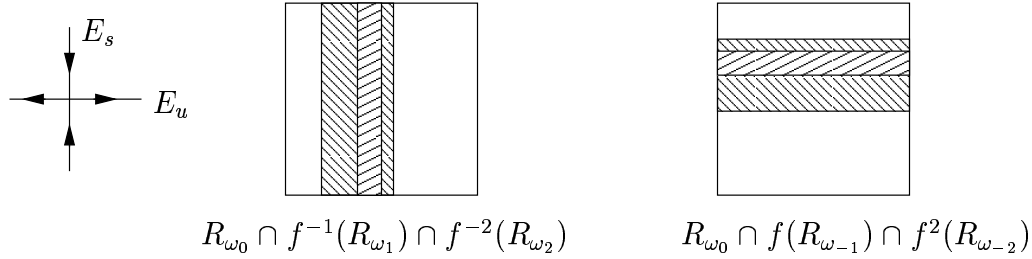
Ce théorème (avec l'existence de partitions de Markov qui sera vue au paragraphe 12.4) dit en particulier que pour tout automorphisme linéaire hyperbolique du tore, il existe un sous-décalage de type fini qui lui est (topologiquement) semi-conjugué. De plus, quitte à enlever un ensemble négligeable (pour toute mesure raisonnable), cette semi-conjugaison est en fait une conjugaison.

Preuve. On notera

$$\overset{\circ}{\mathcal{R}} = \bigcup_{\alpha \in \Lambda} \overset{\circ}{R}_\alpha, \quad \partial \mathcal{R} = \bigcup_{\alpha \in \Lambda} \partial R_\alpha, \quad \partial^s \mathcal{R} = \bigcup_{\alpha \in \Lambda} \partial^s R_\alpha, \quad \partial^u \mathcal{R} = \bigcup_{\alpha \in \Lambda} \partial^u R_\alpha.$$

a) Soit ω un élément de X_A . Par la définition de X_A et la propriété b) des partitions de Markov, la partie $R_{\omega_0} \cap f^{-1}(R_{\omega_1})$ est un sous-rectangle stable non vide de R_{ω_0} . Par récur-

rence, l'intersection $\bigcap_{n=0}^{n_0} f^{-n}(R_{\omega_n})$ est un sous-rectangle stable non vide. Par compacité, l'intersection $\bigcap_{n \geq 0} f^{-n}(R_{\omega_n})$ est aussi un sous-rectangle stable non vide.



On montre de même que $\bigcap_{n \leq 0} f^{-n}(R_{\omega_n})$ est aussi un sous-rectangle instable non vide. Donc l'intersection $\bigcap_{n \in \mathbb{Z}} f^{-n}(R_{\omega_n})$ est non vide. Deux points de cette intersection ont des orbites qui se ϵ_0 -pistent, et sont donc égaux (par le lemme de pistage 12.2).

b) Par compacité, il résulte de a) que l'on a

$$\lim_{n_0 \rightarrow +\infty} \text{diam} \left(\bigcap_{n=-n_0}^{n_0} f^{-n}(R_{\omega_n}) \right) = 0 .$$

Ceci prouve la continuité de π . Par construction, on a $f \circ \pi = \pi \circ \sigma_A$. Par construction, tout point de l'intersection $\bigcap_{n \in \mathbb{Z}} f^{-n}(\mathcal{R})$ est dans l'image $\pi(X_A)$ de π . Par le théorème de Baire, cette intersection est dense dans $\mathbb{T}^N$. Comme $\pi(X_A)$ est compact, l'application π est surjective.

c) La mesure de probabilité $\nu = \pi_* \mu$ est donc f -invariante, ergodique et de support $\mathbb{T}^N$. Soit $Z = \{x \in \mathbb{T}^N / \text{Card } \pi^{-1}(x) > 1\}$. On veut montrer que $\nu(Z) = 0$. Par construction, l'ensemble Z est contenu dans $\bigcup_{n \in \mathbb{Z}} f^n(\partial \mathcal{R})$. Écrivons $\partial \mathcal{R} = \partial^u \mathcal{R} \cup \partial^s \mathcal{R}$. La propriété b) des partitions de Markov implique que $f(\partial^s \mathcal{R}) \subset \partial^s \mathcal{R}$. Comme ν est f -invariante, on a $\nu(\partial^s \mathcal{R}) = \nu(f^n(\partial^s \mathcal{R})) = \nu(\bigcap_{n \geq 0} f^n(\partial^s \mathcal{R}))$. Or le fermé $F = \bigcap_{n \geq 0} f^n(\partial^s \mathcal{R})$ vérifie $f^{-1}(F) = F$ et n'est pas de mesure 1, car le support de ν vaut $\mathbb{T}^n$. Par ergodicité, il est donc de mesure nulle, et donc $\nu(\partial^s \mathcal{R}) = 0$. De même, on a $\nu(\partial^u \mathcal{R}) = 0$, donc $\nu(\partial \mathcal{R}) = 0$ et $\nu(Z) = 0$. $\square$

Remarque. (1) Sous les hypothèses du théorème, on peut montrer que, pour tout x dans $\mathbb{T}^N$, on a

$$\text{Card } \pi^{-1}(x) \leq (\text{Card } \Lambda)^2 .$$

(2) Comme f est topologiquement mélangeante (proposition 6.3 et exercice E.6.3, le sous-décalage σ_A est aussi topologiquement mélangeant et la matrice A est irréductible (voir l'exercice E.11.3). Ceci permet, pour toute matrice stochastique $\Pi = (\pi_{\alpha\beta})_{\alpha, \beta \in \Lambda}$ telle que $\pi_{\alpha\beta} = 0$ si et seulement si $A_{\alpha\beta} = 0$, de construire une mesure de probabilité σ_A -invariante et mélangeante μ_Π de support X_A : la construction donnée au paragraphe 11.2 pour les matrices transitives peut se généraliser aux matrices irréductibles (voir [PY]). En prenant l'image dans $\mathbb{T}^N$ de ces mesures de probabilité, on construit *une famille à plusieurs paramètres réels (donc une famille non dénombrable) de mesures de probabilité ν_Π sur $\mathbb{T}^N$ qui sont f_M -invariantes, mélangeantes et deux à deux étrangères!*

12.3 Un codage sur le tore $\mathbb{T}^2$

Même pour la matrice $M = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$, une partition de Markov est toujours formée d'un grand nombre de rectangles, à cause de la condition sur la petitesse du diamètre de ces rectangles. Si on ôte cette hypothèse, et que l'on remplace la condition *b*) par

b') si $x \in \overset{\circ}{R}_\alpha$ et $f(x) \in \overset{\circ}{R}_\beta$, alors

$$f(W_{R_\alpha}^s(x)) = W_{R_\beta}^s(f(x)) \cap f(R_\alpha) \text{ et } W_{R_\beta}^u(f(x)) = f(W_{R_\alpha}^u(x)) \cap R_\beta ,$$

alors la “partition” $\mathcal{R}$ permet encore de construire un codage $\pi : X_A \rightarrow \mathbb{T}^n$, qui vérifie aussi les conclusions du théorème, voir par exemple l'exercice [KH] (mais π n'est plus forcément à fibres finies, voir la remarque précédente).

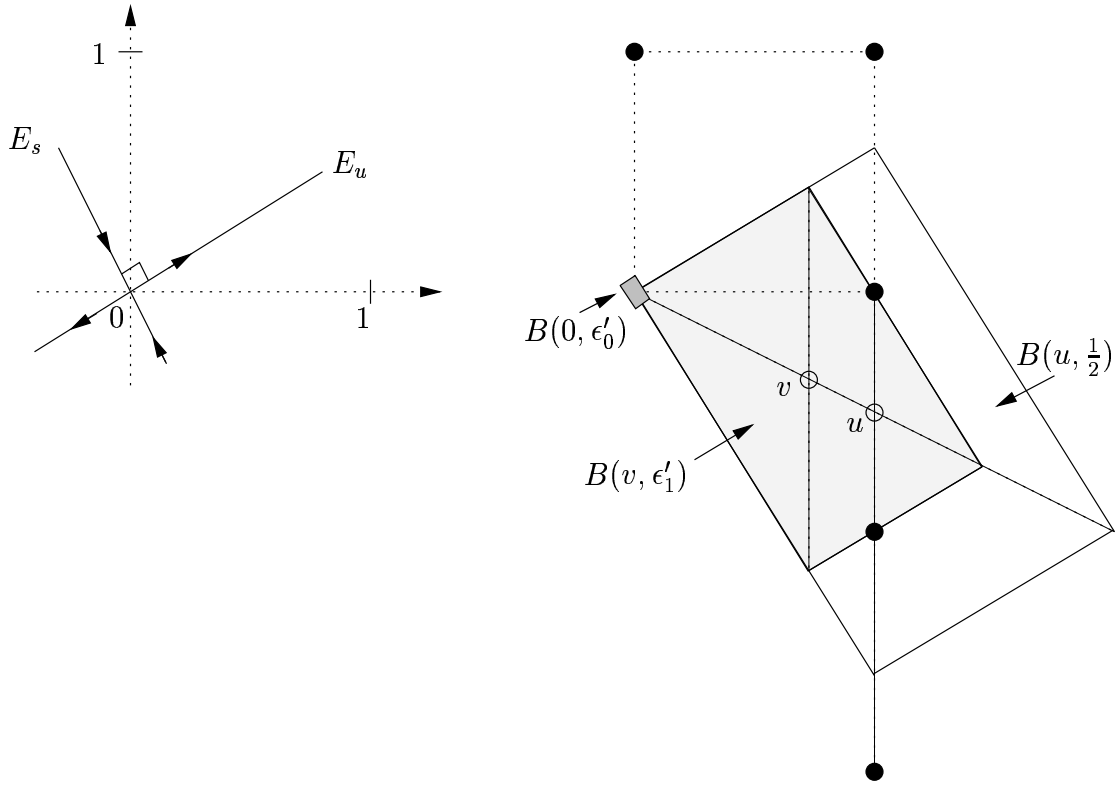
Donnons un exemple explicite de “partition” $\mathcal{R}$ de $\mathbb{T}^2$ qui permet un tel codage avec $M = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$. Les valeurs propres de M sont k^2 et k^{-2} , où k est le nombre d'or

$$k = \frac{1 + \sqrt{5}}{2} \simeq 1,618 ,$$

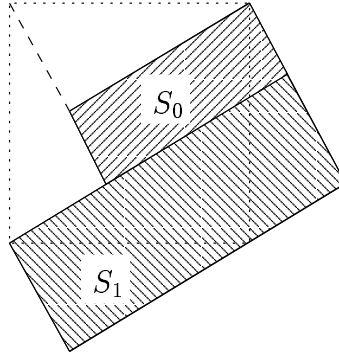
solution de $k^2 = k + 1$. Posons $v_s = (1, -k)$ et $v_u = (1, k - 1)$. Les sous-espaces stables et instables de M sont $E_s = \mathbb{R}v_s$ et $E_u = \mathbb{R}v_u$, voir dessin ci-dessous. La norme, définie par $\|v\| = \sup\{|x|, |y|\}$ pour $v = xv_s + yv_u$, est adaptée à M , la constante d'hyperbolicité de M est $ch(M) = \frac{1}{k^2}$, et $\|M\| = \|M^{-1}\| = k^2$. Le rectangle de côtés les vecteurs v_s, v_u est une boule de rayon $\frac{1}{2}$ pour cette norme. Un petit calcul montre que la projection canonique est injective sur la boule de rayon r si et seulement si $r < \epsilon'_1$, avec $\epsilon'_1 = \frac{1}{2(3-k)} \simeq 0,3618$. Donc la constante ϵ_0 du paragraphe 12.1 doit être prise strictement inférieure à

$$\frac{\epsilon'_1(1 - ch(M))}{2 \sup\{\|M\|, \|M^{-1}\|\}} = \frac{k - 1}{4(k + 2)} \simeq 0,042 .$$

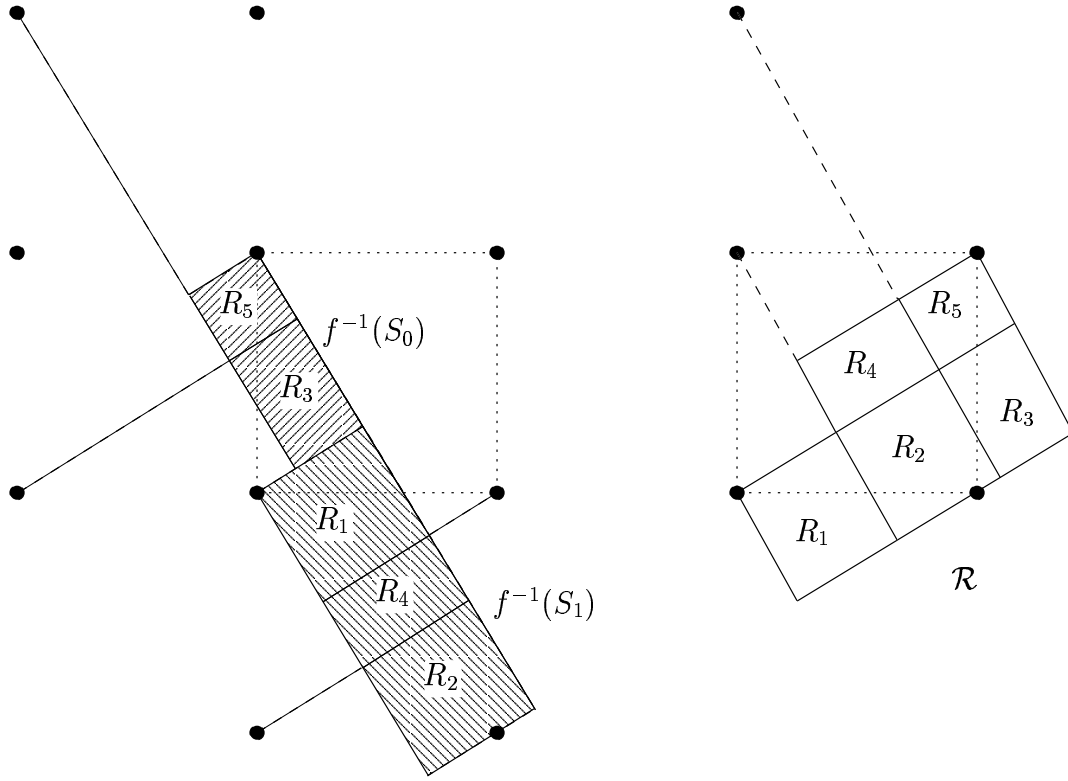
Le recouvrement par rectangles $\mathcal{R}$ que nous allons construire aura 5 rectangles, alors qu'une (vraie) partition de Markov (i.e. dont les rectangles sont de diamètre strictement inférieur à ϵ'_0) aurait de l'ordre d'une centaine de rectangles.



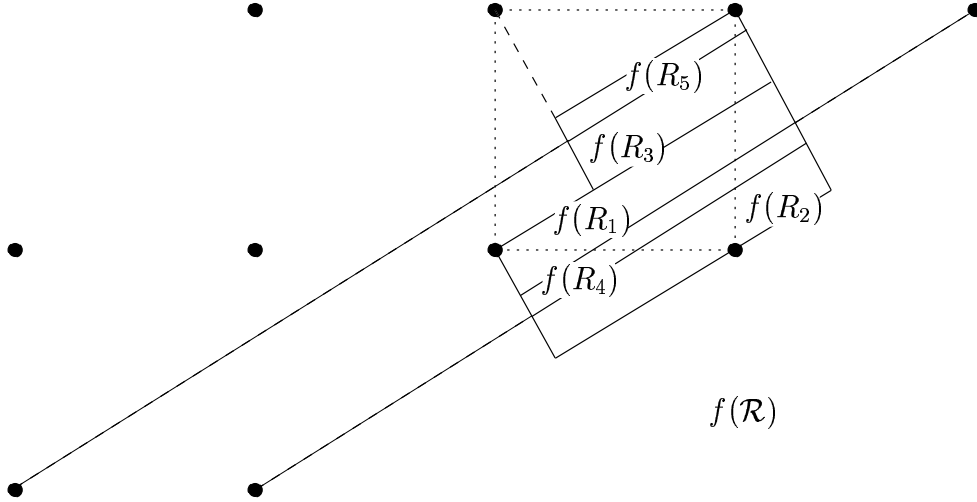
On part d'un premier découpage de $\mathbb{T}^2$ en deux rectangles S_0, S_1 aux côtés parallèles à E^s et E^u , et dont deux côtés sont sur les projections de E^s, E^u . Plus précisément, S_0 est la projection du rectangle de $\mathbb{R}^2$ de côtés portés par les droites affines $(1, 1) + E_s, (1, 1) + E_u, (0, 1) + E_s$, et S_0 est la projection du rectangle de $\mathbb{R}^2$ de côtés portés par les droites affines $E_s, E_u, (1, 0) + E_u, (1, 1) + E_s$, voir dessin ci-dessous.



Comme $M^{-1} : (1, 1) \mapsto (0, 1), (0, 1) \mapsto (-1, 2), (1, 0) \mapsto (1, -1)$, on calcule facilement $f^{-1}(S_0), f^{-1}(S_1)$. On prend alors $\mathcal{R} = \{R_1, R_2, R_3, R_4, R_5\}$ où les $\overset{\circ}{R}_\alpha$ sont les composantes connexes de $f^{-1}(\overset{\circ}{S}_i) \cap \overset{\circ}{S}_j$.



On numérote les rectangles R_i de sorte que $f^{-1}(\overset{\circ}{S}_0) \cap \overset{\circ}{S}_0 = \overset{\circ}{R}_5$, $f^{-1}(\overset{\circ}{S}_0) \cap \overset{\circ}{S}_1 = \overset{\circ}{R}_3$, $f^{-1}(\overset{\circ}{S}_1) \cap \overset{\circ}{S}_0 = \overset{\circ}{R}_4$, $f^{-1}(\overset{\circ}{S}_1) \cap \overset{\circ}{S}_1 = \overset{\circ}{R}_1 \cup \overset{\circ}{R}_2$. Donc le rectangle S_0 est réunion des rectangles R_4 et R_5 , et le rectangle S_1 est réunion des rectangles R_1, R_2 et R_3 . On calcule les images par f des éléments de $\mathcal{R}$, voir figure ci-dessous.

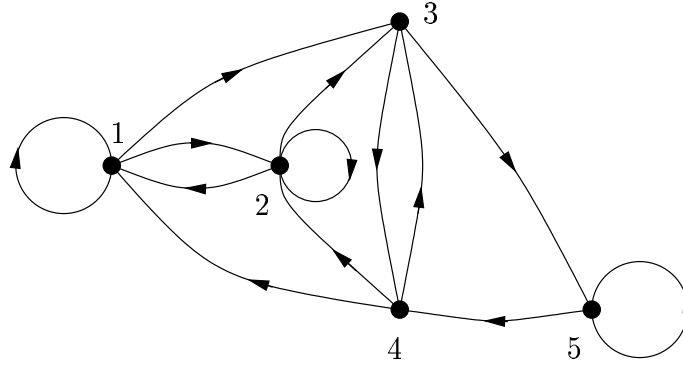


La matrice d'incidence se calcule alors facilement : lorsque $R_\alpha \subset f^{-1}(S_i) \cap S_j$ et $R_\beta \subset f^{-1}(S_k) \cap S_\ell$, alors $A_{\alpha\beta} = 1$, c'est-à-dire $f(\overset{\circ}{R}_\alpha) \cap \overset{\circ}{R}_\beta$ est non vide, si et seulement si $i = \ell$.

On a donc (et cela se lit sur les dessins précédents)

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

On peut dessiner le graphe Γ_A de matrice d'incidence A :



12.4 Constructions de partitions de Markov

Théorème 12.5 *Pour tout $\epsilon > 0$ et tout automorphisme linéaire hyperbolique f_M de $\mathbb{T}^N$, il existe une partition de Markov $\mathcal{R}$ pour f_M par des rectangles de diamètre au plus ϵ .*

Preuve. La démonstration est en deux étapes. On peut supposer $\epsilon \leq \epsilon_0$. Notons $f = f_M$.

Étape 1 : on construit tout d'abord un recouvrement $\mathcal{R}'$ par des rectangles de diamètre au plus ϵ vérifiant un analogue de la propriété *b*) des partitions de Markov.

Pour cela, on choisit un recouvrement fini $(B_i)_{i \in I}$ de $\mathbb{T}^N$ par des boules B_i de centre x_i et de rayon η où $\eta \leq \frac{\epsilon(1-ch(M))}{2(1+||M||)}$. On note $A' = (A'_{ij})_{i,j \in I}$ la matrice donnée par $A'_{ij} = 1$ si et seulement si $f(x_i) \in B(x_j, \eta(1 + ||M||))$, et on note $(X_{A'}, \sigma_{A'})$ le sous-décalage de type fini associé.

Pour tout ω dans $X_{A'}$, la suite $n \mapsto x_{\omega_n}$ est une $\eta(1 + ||M||)$ -pseudo-orbite. Donc, d'après le lemme 12.2, elle est $\frac{\eta(1+||M||)}{1-ch(M)}$ -pistée par une unique orbite d'un point $\rho(\omega)$ de $\mathbb{T}^N$. Comme dans le paragraphe 12.2, il est facile de voir que l'application $\rho : X_{A'} \rightarrow \mathbb{T}^N$ est continue et vérifie $\rho \circ \sigma_{A'} = f \circ \rho$. Cette application est surjective, car tout point x de $\mathbb{T}^N$ est l'image de toute suite ω telle que, pour tout n dans $\mathbb{Z}$, le point $f^n(x)$ est dans B_{ω_n} (une telle suite ω est toujours dans $X_{A'}$, car

$$d(f(x_{\omega_n}), x_{\omega_{n+1}}) \leq d(f(x_{\omega_n}), f(f^n(x))) + d(f^{n+1}(x), x_{\omega_{n+1}}) \leq (||M|| + 1)\eta).$$

Notons $X_i^0 = \{\omega \in X_{A'} \mid \omega_0 = i\}$ et $R'_i = \rho(X_i^0)$ et montrons que le recouvrement $\mathcal{R}' = (R'_i)_{i \in I}$ convient.

Pour cela, introduisons la structure de “produit local” sur $X_{A'}$: pour ω, ω' dans $X_{A'}$ tels que $\omega_0 = \omega'_0$, on note $[\omega, \omega']$ l’élément de $X_{A'}$ tel que

$$[\omega, \omega']_n = \begin{cases} \omega_n & \text{si } n \geq 0 \\ \omega'_n & \text{si } n \leq 0 \end{cases} .$$

Autrement dit, $[\omega, \omega']$ a le même futur que ω et le même passé que ω' (remarquons que

$$\lim_{n \rightarrow +\infty} d(\sigma_{A'}^n(\omega), \sigma_{A'}^n([\omega, \omega'])) = 0, \quad \lim_{n \rightarrow +\infty} d(\sigma_{A'}^{-n}(\omega'), \sigma_{A'}^{-n}([\omega, \omega'])) = 0 ,$$

donc $[\omega, \omega']$ est le point d’intersection de la “variété locale stable” de ω et de la “variété locale instable” de ω').

Par construction (i.e. par la relation d’équivariance $\rho \circ \sigma_{A'} = f \circ \rho$), on a $\rho([\omega, \omega']) = [\rho(\omega), \rho(\omega')]$. Ce qui prouve bien que les parties R'_i sont des rectangles.

Comme par construction (i.e. le fait que $(x_{\omega_n})_{n \in \mathbb{Z}}$ est $\frac{\eta(1+||M||)}{1-ch(M)}$ -pistée par l’orbite de $\rho(\omega)$), le point $\rho(\omega)$ appartient à la boule $B\left(x_{\omega_0}, \frac{\eta(1+||M||)}{1-ch(M)}\right)$. Donc si x, x' appartiennent à R'_i , alors, en écrivant $x = \rho(\omega), x' = \rho(\omega')$ avec ω, ω' dans X_i^0 , on a $d(x, x') \leq \epsilon$, par définition de η . Donc les parties R'_i sont de diamètre au plus ϵ .

L’analogie de la propriété b) que vérifient ces rectangles R'_i est donné par le lemme suivant.

Lemme 12.6 *Avec les notations précédentes, soient ω dans X_A , $i = \omega_0$, $j = \omega_1$ et $x = \rho(\omega)$. Alors on a*

$$f(W_{R'_i}^s(x)) \subset W_{R'_j}^s(f(x)) \quad \text{et} \quad W_{R'_i}^u(f(x)) \subset f(W_{R'_j}^u(x)) .$$

Remarque. Tous les points de $\overset{\circ}{R'_i} \cap f^{-1}(\overset{\circ}{R'_j})$ ne sont pas de la forme $\rho(\omega)$ avec $\omega_0 = i, \omega_1 = j$. Ce lemme n’est donc qu’un analogue plus faible de la propriété b) des partitions de Markov.

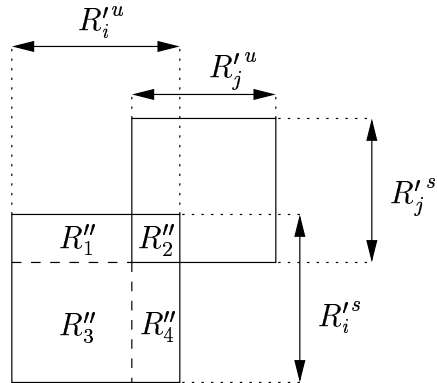
Preuve. On a $W_{R'_i}^s(\rho(\omega)) = \rho(\{\omega' \in X_{A'} / \forall n \geq 0, \omega'_n = \omega_n\})$. Donc

$$\begin{aligned} f\left(W_{R'_i}^s(\rho(\omega))\right) &= \rho(\{\omega' \in X_{A'} / \forall n \geq 0, \omega'_{n-1} = \omega_n\}) \subset \\ &\rho(\{\omega' \in X_{A'} / \forall n \geq 0, \omega'_n = \omega_{n+1}\}) = W_{R'_j}^s(f(\rho(\omega))) . \end{aligned}$$

De même avec les variétés instables. □

Étape 2 : on construit $\mathcal{R}$ en redécoupant les rectangles de $\mathcal{R}'$.

Pour chaque rectangle $\overset{\circ}{R'_j}$ qui rencontre $\overset{\circ}{R'_i}$, on découpe R'_i en (au plus) quatre rectangles comme sur le dessin ci-contre (ce sont $(R_i^s \cap R_j^s) \times (R_i^u \cap R_j^u)$, $(R_i^s \cap R_j^s) \times (R_i^u - R_j^u)$, $(R_i^s - R_j^s) \times (R_i^u \cap R_j^u)$, $(R_i^s - R_j^s) \times (R_i^u - R_j^u)$). On note $(R''_a)_{a \in \mathcal{A}}$ les rectangles qui apparaissent ainsi.

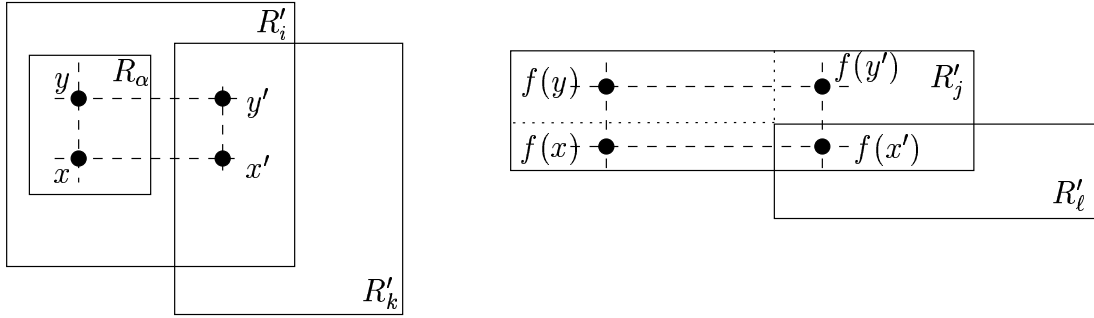


Pour chaque point x de $\mathbb{T}^N$, on note $\overset{\circ}{R}(x)$ l'intersection des $\overset{\circ}{R}''_a$ qui contiennent x , et $R(x)$ l'adhérence de $\overset{\circ}{R}(x)$, qui est un rectangle. Seul un nombre fini de rectangles apparaissent ainsi. On les note $(R_\alpha)_{\alpha \in \Lambda}$. Ils sont de diamètre au plus ϵ et vérifient $\overset{\circ}{R}_\alpha \cap \overset{\circ}{R}_\beta = \emptyset$ si $\alpha \neq \beta$.

Il reste à vérifier que cette famille $\mathcal{R} = (R_\alpha)_{\alpha \in \Lambda}$ vérifie la propriété b) des partitions de Markov.

Soient x dans $\overset{\circ}{R}_\alpha$ tel que $f(x)$ est dans $\overset{\circ}{R}_\beta$, et y dans $W_{R_\alpha}^s(x)$. Il suffit de montrer que $f(y)$ appartient à R_β , donc à $W_{R_\beta}^s(f(x))$, le raisonnement étant le même pour les variétés instables locales, en remplaçant f par f^{-1} .

Pour tout j tel que le rectangle $\overset{\circ}{R}'_j$ contient $f(x)$, soit ω dans $X_{A'}$ tel que $\rho(\omega) = x$ et $\omega_1 = j$. Notons $i = \omega_0$, de sorte que $R_\alpha \subset R'_i$. D'après le lemme 12.6 de la première étape, on sait que le point $f(y)$ appartient à $\overset{\circ}{R}'_j$. Si $f(y)$ n'est pas dans R_β , alors il existe un élément ℓ dans I tel que $f(x)$ et $f(y)$ sont dans des rectangles différents du découpage en quatre de $\overset{\circ}{R}'_j$ dû à sa rencontre avec $\overset{\circ}{R}'_\ell$.



Comme $f(W_{R'_i}^u(x))$ contient $W_{R'_j}^u(f(x))$, il existe un point x' de $W_{R'_i}^u(x)$ tel que, en notant $y' = [x', y]$, les points $f(x')$ et $f(y')$ ne sont pas simultanément dans R'_ℓ . Par exemple, supposons que $f(x')$ est dans R'_ℓ et que $f(y')$ n'y est pas. On peut trouver ω' dans X_A tel que $\rho(\omega') = x'$ et $\omega'_1 = \ell$. Notons $k = \omega'_0$. Comme x et y sont dans le même sous-rectangle du découpage en quatre de $\overset{\circ}{R}'_i$ dû à sa rencontre avec $\overset{\circ}{R}'_k$, le point y' est forcément dans R'_k . Ceci contredit l'inclusion $f(W_{R'_k}^s(x')) \subset W_{R'_\ell}^s(f(x'))$.

Remarquons que le raisonnement ci-dessus n'est valable que pour les points x tels que $W_\epsilon^u \cup W_\epsilon^u(f(x))$ ne rencontrent pas $\partial^u \mathcal{R}'$. Cela est suffisant pour conclure, car ces points forment une partie dense de $\mathbb{T}^N$. $\square$

12.5 Entropie des automorphismes hyperboliques du tore

Proposition 12.7 *Soit f_M un automorphisme linéaire hyperbolique du tore $\mathbb{T}^N$ et μ la mesure de probabilité de Haar sur $\mathbb{T}^N$. Notons $\lambda_+ = |\det(M|_{E_u})|$. Alors*

$$h_{\text{top}}(f_M) = h_\mu(f_M) = \log \lambda_+ .$$

De plus, μ est l'unique mesure de probabilité borélienne f_M -invariante sur $\mathbb{T}^N$ vérifiant cette égalité.

Remarque. On a $\log \lambda_+ = \sum \log |\lambda_i|$, où les λ_i sont les valeurs propres de M telles que $|\lambda_i| > 1$.

Preuve. Soient $\mathcal{R} = \{R_\alpha, \alpha \in \Lambda\}$ une partition de Markov pour f_M , et $\xi_0 = \{R'_\alpha, \alpha \in \Lambda\}$ une (vraie) partition mesurable de $\mathbb{T}^N$ telle que, pour tout α , on a $\overset{\circ}{R}_\alpha \subset R'_\alpha \subset R_\alpha$. La proposition résulte alors des deux lemmes suivants.

Lemme 12.8 *Il existe $c_1, c_2 > 0$ tels que, pour tout $n \geq 0$ et toute partie C non μ -négligeable de $\bigvee_{i=0}^{n-1} f_M^{-1} \xi_0$, on a*

$$c_1 \lambda_+^{-n} \leq \mu(C) \leq c_2 \lambda_+^{-1} .$$

Preuve. Notons $m = m_s \otimes m_u$ la mesure de Lebesgue de $\mathbb{R}^N = E_s \times E_u$, qui s'identifie localement à μ . On a donc, pour tout rectangle $R_\alpha = R_\alpha^s \times R_\alpha^u$, l'égalité $\mu(R_\alpha) = m_s(R_\alpha^s) m_u(R_\alpha^u)$. Par conséquent, pour tout w dans X_A , on a

$$\bigcap_{i=0}^{n-1} f^{-i}(R_{\omega_i}) = R_{\omega_0}^s \times f^{-(n-1)}(R_{\omega_{n-1}}^u) ,$$

et donc, si $C = \bigcap_{i=0}^{n-1} f^{-i}(R_{\omega_i})$, alors $\mu(C) = \lambda_+^{-n+1} m_s(R_{\omega_0}^s) m_u(R_{\omega_{n-1}}^u)$. Il suffit donc de prendre pour c_1 le minimum des $\lambda_+ m_s(R_\alpha^s) m_u(R_\beta^u)$ et pour c_2 le maximum. $\square$

Lemme 12.9 *Soient X un espace métrique compact, f un homéomorphisme de X , μ une mesure de probabilité borélienne f -invariante et ergodique sur X , et ξ_0 une partition mesurable de X génératrice pour f . On suppose qu'il existe $c_1, c_2, \lambda_+ > 0$ tels que, pour toute partie C non μ -négligeable de $\bigvee_{i=0}^{n-1} f^{-i} \xi_0$, et pour tout $n \geq 0$, on a $c_1 \lambda_+^{-n} \leq \mu(C) \leq c_2 \lambda_+^{-n}$. Alors $h_{\text{top}}(f) = h_\mu(f) = \log \lambda_+$. En outre, μ est l'unique mesure de probabilité borélienne f_M -invariante sur X vérifiant cette égalité.*

Preuve. L'hypothèse assure que, pour tout $n \geq 0$, on a (voir la remarque (3) juste avant le paragraphe 9.2)

$$H \left(\bigvee_{i=0}^{n-1} f^{-i} \xi_0 \right) \in [-\log(c_2 \lambda_+^{-n}), -\log(c_1 \lambda_+^{-n})] .$$

Le théorème de Kolmogorov-Sinai donne alors

$$h_\mu(f) = h_\mu(f, \xi_0) = \lim_{n \rightarrow \infty} \frac{1}{n} H \left(\bigvee_{i=0}^{n-1} f^{-i} \xi_0 \right) = \log(\lambda_+) .$$

D'après le principe variationnel (le théorème 10.11), il reste à montrer que si une mesure de probabilité borélienne f -invariante ν sur X vérifie $h_\nu(f) \geq h_\mu(f)$, alors $\nu = \mu$. Pour cela, il suffit de reprendre quasiment mot à mot la démonstration de la proposition 11.8. $\square$

On a vu que f_M est mélangeante, donc ergodique, pour la mesure de Haar μ (voir la proposition 6.3). La partition ξ_0 construite ci-dessus est génératrice pour f_M (qui est inversible), car par la preuve du b) du théorème 12.4, les éléments de $\bigvee_{i=-n}^n f^{-1} \xi_0$ sont des rectangles d'intérieurs disjoints et de diamètre tendant vers 0, donc qui engendrent la σ -algèbre des boréliens. En appliquant le lemme précédent, on termine la démonstration de la proposition 12.7. $\square$

12.6 Exercices

Exercice E.12.1 Soit f_M un automorphisme linéaire hyperbolique du tore $\mathbb{T}^N$. Montrer que pour tout $\eta > 0$, il existe un voisinage U de F_M dans $C^1(\mathbb{T}^N, \mathbb{T}^N)$ et $\epsilon > 0$ tels que, pour tout homéomorphisme g dans U , alors toute ϵ -pseudo-orbite de g est η -pistée par une orbite de g .

Exercice E.12.2 Soit $\mathcal{R} = (R_\alpha)_{\alpha \in \Lambda}$ un recouvrement de $\mathbb{T}^N$ par des rectangles propres d'intérieurs disjoints vérifiant la condition b)' de la section 12.3 :

b') si $x \in \overset{\circ}{R}_\alpha$ et $f(x) \in \overset{\circ}{R}_\beta$, alors

$$f(W_{R_\alpha}^s(x)) = W_{R_\beta}^s(f(x)) \cap f(R_\alpha) \text{ et } W_{R_\beta}^u(f(x)) = f(W_{R_\alpha}^u(x)) \cap R_\beta .$$

Soient A la matrice d'incidence de $\mathcal{R}$ (définie comme pour les "vraies" partitions de Markov) et (X_A, σ_A) le sous-décalage de type fini associé. Montrer que les conclusions du théorème 12.4 sont encore vérifiées :

- a) Pour tout ω dans X_A , l'intersection $\bigcap_{n \in \mathbb{Z}} f^{-n}(R_{\omega_n})$ est réduite à un point $\pi(\omega)$.
- b) L'application $\pi : X_A \rightarrow \mathbb{T}^N$ est continue, surjective et vérifie $f \circ \pi = \pi \circ \sigma_A$.
- c) Pour toute mesure de probabilité σ_A -invariante et ergodique μ , de support X_A , on a

$$\mu(\{\omega \in X_A / \text{Card } \pi^{-1}(\pi(\omega)) > 1\}) = 0 .$$

Exercice E.12.3 Soit $M = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$, f_M l'automorphisme linéaire hyperbolique du tore $\mathbb{T}^2$ associé, k le nombre d'or et $P_n = \{x \in \mathbb{T}^2 / f_M^n(x) = x\}$ l'ensemble des points périodiques de période divisant n pour f_M . Montrer que

$$\text{Card } P_n = k^{2n} + k^{-2n} - 2 .$$

Si $\mathcal{R}$ est la partition de Markov construite au paragraphe 12.3, A sa matrice d'incidence, et (X_A, σ_A) le sous-décalage de type fini de matrice A , calculer le nombre de points périodiques de période divisant n pour σ_A .

12.7 Indications pour la résolution des exercices

Exercice E.12.1 Utiliser le théorème d'Anosov 7.11, en remarquant que h est proche de l'identité.

Exercice E.12.2 Reprendre la preuve du théorème 12.4.

Exercice E.12.3 On pourra regarder l'application $f_M^n - id$ de $\mathbb{T}^2$ et le nombre $\det(M^n - id)$.

Index

- $\alpha \vee \beta$, 80
- algèbre
 - de Boole, 7
 - de quaternions, 73
- alphabet, 8
- anneau de quaternions, 73
- application
 - de Gauss, 48
 - ergodique, 28
 - expansive, 93
 - invariante, 28
 - mélangeante, 44
 - préservant la mesure, 6
 - presque partout invariante, 28
 - semi-continue supérieurement, 19
 - topologiquement mélangeante, 53
 - uniquement ergodique, 38
- asymptotiquement stable, 65
- automorphisme
 - hyperbolique, 58
- $C^1(\mathbb{T}^N, \mathbb{T}^N)$, 63
- $C^0(X, \mathbb{C})$, 33
- $C_c(X)$, 68
- $c_{v,v'}$, 69
- closing lemma, 111
- coder, 99
- coefficient
 - matriciel, 69
- condition de Renyi, 49
- conjugué, 73
- conjugués, 6, 7
- constante
 - d'expansivité, 94
 - d'hyperbolicité, 58
 - de Lipschitz, 58
- convergence
 - faible, 71
- convexe
 - symétrique, 74
- corps gauche, 73
- critère
 - de Weyl, 41
- cylindre, 9, 102
- Δ , 82
- décomposition
 - de Cartan, 71
- développement
 - en fraction continue, 48
- décalage, 9
- demi-plan supérieur, 75
- ϵ -dense, 24, 90
- différence symétrique, 82
- différentielle, 63
- distance
 - de Hilbert, 100
 - hyperbolique, 76
 - invariante à gauche, 67
- distribution, 32
- doublement de l'angle, 15, 29
- entropie, 79, 84, 90, 91
 - métrique, 79, 84
 - relative, 80
 - topologique, 91
- équirépartition, 38
- ergodique, 28
- espérance, 32
- expansive, 93
- $f_*\mu$, 8
- fibré unitaire tangent, 76
- flot
 - du billard, 12
 - géodésique, 77
- fonction
 - caractéristique, 17, 28
- fraction continue, 48
- géodésique hyperbolique, 76
- groupe
 - unimodulaire, 66
 - unitaire, 68
- $H(\alpha|\beta)$, 80
- $H_\mu(\alpha)$, 79
- $\mathbb{H}$, 75
- $h(f, \mathcal{U})$, 91
- $h(f, \alpha)$, 84

- $h_\mu(f)$, 84
- $h_{\text{top}}(f)$, 91
- hamiltonien, 11
- homéomorphisme
 - bilipschitzien, 57
- I_α , 79
- $I_{\alpha|\beta}$, 80
- $\mathbb{I}_A$, 28
- indépendantes, 32
- information, 79
 - relative, 80
- intégrale première, 11
- invariant, 6, 18, 25, 28
- inversible, 6
- joint, 80, 90
- lemme
 - de Minkowski, 74
 - de fermeture, 111
 - de Lebesgue, 92
 - de pistage, 65, 111
 - de van der Corput, 43
- lipschitzien, 33
- loi, 32
- loi forte des grands nombres, 32
- longueur hyperbolique, 76
- mélange, 44
 - faible, 44
 - fort, 44
- matrice
 - d'incidence, 99, 113
 - irréductible, 107
 - période d'une, 107
 - stochastique, 103
- mesurablement conjugués, 7
- mesure
 - d'entropie maximale, 97
 - de Gauss, 49
 - de Haar, 8, 66
 - de Liouville, 10
 - de Markov, 103
 - de Parry, 105
 - ergodique, 28
 - image, 8
 - invariante, 6
 - mélangeante, 44
 - produit, 9
 - σ -finie, 7
 - uniquement ergodique, 38
- mesures
 - étrangères, 36
- minimal, 18, 25
- moins fin, 90
- moins fine, 80
- mot, 9
- moyenne
 - orbitale, 30
 - spatiale, 30
 - temporelle, 30
- $N'_d(\epsilon)$, 91
- $N_d(\epsilon)$, 91
- norme
 - adaptée, 58
- norme hyperbolique, 76
- norme matricielle, 67
- $\mathcal{O}_\epsilon$, 90
- orbite, 6
 - positive, 6
- paradoxe de Zermelo, 22
- partie
 - ϵ -dense, 90
 - entière, 38
 - fractionnaire, 38
 - invariante, 6, 18, 25
 - minimale, 18, 25
 - ϵ -séparé, 91
- partition
 - de Markov, 112
 - mesurable, 79
 - génératrice, 86
 - indépendantes, 88
 - moins fine, 80
- périodique, 18
- phénomène de Mautner, 69
- pistage, 65, 111
- plan hyperbolique, 76
- principe variationnel, 94
- produit local, 110
- ϵ -pseudo-orbite, 110

- quasipériodique, 23
- quotient, 6
- raffinement, 80
- rayon spectral, 101
- recouvrement
 - générateur, 93
 - joint, 90
 - moins fin, 90
- rectangle, 111
 - propre, 112
- récurrent, 18
- relèvement, 62
 - des applications, 62
 - des chemins, 62
- représentation
 - unitaire, 68
- réseau, 68
 - cocompact, 68
 - uniforme, 68
- rotation, 29
- $SL_N(\mathbb{R})$, 66
- $\mathbb{S}^1$, 8
- σ , 9
- séparant les points, 33
- semi-algèbre, 8
- semi-conjugués, 6
- semi-flot mesurable, 35
- ϵ -séparé, 91
- shadowing lemma, 111
- shift, 9
- σ -algèbre
 - produit, 9
- somme de Birkhoff, 30
- sous-décalage de type fini, 99
 - transitif, 100
- sous-rectangle
 - instable, 112
 - stable, 112
- suite
 - équirépartie, 38
- système
 - de Bernoulli, 10
 - de Kronecker, 8
 - de Markov, 99
 - transitif, 100
 - hamiltonien, 11
 - symbolique, 10
- système dynamique
 - mesuré, 6
 - mesurable, 6
 - inversible, 6
 - probabilisé, 6, 84
 - inversible, 84
 - topologique, 6
 - inversible, 6
- $\mathbb{T}^N$, 29, 39
- temps de premier retour, 36
- théorème
 - d'Anosov, 64
 - de Banach-Alaoglu, 33
 - de Birkhoff, 30
 - de Borel-Bernstein, 50
 - de Carathéodory, 7
 - de Dirichlet, 47
 - de Furstenberg, 24
 - de Furstenberg-Weiss, 19
 - de Furstenberg-Weiss II, 23
 - de Gauss, 49
 - de Grobman-Hartman, 61
 - de Hindman, 24
 - de Howe-Moore, 70
 - de Kolmogorov, 32
 - de Kolmogorov-Sinaï, 86
 - de Liouville, 10
 - de Perron-Frobenius, 100
 - de récurrence
 - de Birkhoff, 18
 - de Kac, 36
 - de Poincaré, 22
 - de Riesz, 33
 - de Stone-Weierstrass, 33
 - de van der Waerden, 21
 - de Weyl, 41
 - du point fixe, 57
 - ergodique, 30
 - topologiquement conjugués, 6
 - topologiquement mélangeante, 53
 - topologiquement semi-conjugués, 6
- tore, 39
- trajectoire
 - de billard, 12
- transformation

- de Gauss, 48
- du billard, 13
- du boulanger, 53
- linéaire du tore, 45
- transformation de Fourier, 29
- transitive, 100
- translation
 - à droite, 66
 - à gauche, 66
- $U(\mathcal{H})$, 68
- uniformément distribuées, 32
- unimodulaire, 66
- unique ergodicité, 38
- uniquement ergodique, 38
- variable aléatoire, 32
- variables aléatoires
 - indépendantes, 32
 - uniformément distribuées, 32
- variété instable locale, 110
- variété stable locale, 110

Références

- [BHV] B. Bekka, P. de la Harpe, A. Valette, *Background on unitary group representations*, Notes polycopiées, Univ. Neuchâtel, 2002.
- [BM] B. Bekka, M. Mayer, *Ergodic theory and topological dynamics of group actions on homogeneous spaces*, LMS LNS **269**, Cambridge Univ. Press, 2000.
- [Bor] A. Borel, *Introduction aux groupes arithmétiques*, Hermann, Paris, 1967.
- [BHC] A. Borel, Harish-Chandra, *Arithmetic subgroups of algebraic groups*, Ann. Math. **75** (1962) 485-535.
- [Bow] R. Bowen, *Equilibrium states and the ergodic theory of Anosov diffeomorphisms*, Lect. Notes **470**, Springer Verlag, 1967.
- [Bre] H. Brezis, *Analyse fonctionnelle*, Masson, 1983.
- [Car] P. Cartier, *Représentations des groupes localement compacts*, Notes de cours, IHES, 1974.
- [CohD] D. Cohn, *Measure theory*, Birkhäuser, 1980.
- [CohP] P. Cohn, *Skew field constructions*, LMS LNS **27**, Cambridge Univ. Press, 1977.
- [CFS] I.P. Cornfeld, S.V. Fomin, Ya.G. Sinai, *Ergodic theory*, Springer Verlag, 1981.
- [Dud] R. Dudley, *Real analysis and probability*, Wadsworth & Brooks/Cole, 1989.
- [Fur] H. Furstenberg, *Recurrence in ergodic theory and combinatorial number theory*, Princeton Univ. Press, 1981.
- [GHL] S. Gallot, D. Hulin. J. Lafontaine, *Riemannian geometry*, Springer Verlag, 1990.
- [God] C. Godbillon, *Éléments de topologie algébrique*, Hermann, 1971.
- [Har] P. de la Harpe, *On Hilbert's metric for simplices*, pp 97-119, Vol. 1 de "Geometric group theory", A. Niblo, M. Roller eds, LMS LNS **182**, Cambridge Univ. Press, 1993.
- [Hat] A. Hatcher, *Algebraic topology*, Cambridge Univ. Press, 2002, <http://www.math.cornell.edu/~hatcher>.
- [HT] R. Howe, E.C. Tan *Non-abelian harmonic analysis : application of $SL(2, \mathbb{R})$* , Universitext, Springer Verlag, 1992.
- [Kat] S. Katok, *Fuchsian groups*, Univ. Chicago Press, 1992.
- [KH] A. Katok, B. Hasselblatt, *Introduction to the modern theory of dynamical systems*, Ency. Math. App. **54**, Camb. Univ. Press, 1995.
- [Kri] J.L. Krivine, *Théorie axiomatique des ensembles* PUF, 1969.
- [Laf] J. Lafontaine, *Introduction aux variétés différentielles*, Press. Univ. Grenoble, 1996.
- [Man] R. Mañé, *Ergodic theory and differentiable dynamics*, Springer Verlag, 1987.
- [Mar] G. Margulis, *Discrete subgroups of semi-simple groups*, Ergeb. Math. Grenz. **17**, Springer Verlag, 1991.
- [MT] N. Mneimné, F. Testard, *Introduction à la théorie des groupes de Lie classiques*, Hermann, 1986.

- [Neu] J. Neveu, *Calcul des probabilités*, 2nd éd., Masson, 1980.
- [Orn] D. Ornstein, *Ergodic theory, randomness and dynamical systems*, Yale Univ. Press, 1974.
- [Pau] F. Paulin, *Topologie algébrique élémentaire*, Notes de cours de magistère, 2002
<http://www.dma.ens.fr>.
- [PY] M. Pollicott, M. Yuri, *Dynamical systems and ergodic theory*, London Math. Soc. Stud. Texts **40**, Cambridge Univ. Press, 1998.
- [Rud] W. Rudin, *Real and complex analysis*, 3rd ed., Mc Graw-Hill, 1990.
- [Sin1] Ya.G. Sinai, *Introduction to ergodic theory*, Princeton Univ. Press, 1976.
- [Sin2] Ya.G. Sinai, *Topics in ergodic theory*, Princeton Math. Series **44**, Princeton Univ. Press, 1994.
- [Spa] E. Spanier, *Algebraic topology*, Tata McGraw-Hill, 1981.
- [Spi] M. Spivak, *Differential geometry*, Publish or Perish, 1979.
- [Spr] V. Sprindzuk, *Mahler's problem in metric number theory*, Trans. Math. Mono. **25**, Amer. Math. Soc., 1969.
- [Wal] P. Walters, *An introduction to ergodic theory*, GTM **79**, Springer Verlag, 1982.
- [Wei] A. Weil, *L'intégration dans les groupes topologiques et ses applications*, Hermann, 1951.
- [Yoc] J.-C. Yoccoz, *Recent developments in dynamics*, dans Proc. Inter. Cong. Math. Vol. 1, 2 (Zürich, 1994), 246–265, Birkhäuser, 1995.
- [Zim] R.J. Zimmer, *Ergodic theory and semisimple groups*, Birkhauser, 1984.

Département de Mathématique et Applications, UMR 8553 CNRS
 Ecole Normale Supérieure
 45 rue d'Ulm
 75230 PARIS Cedex 05, FRANCE
e-mail: Frederic.Paulin@ens.fr