NOTES DE COURS: THÉORIE ERGODIQUE ET SYSTÈMES DYNAMIQUES

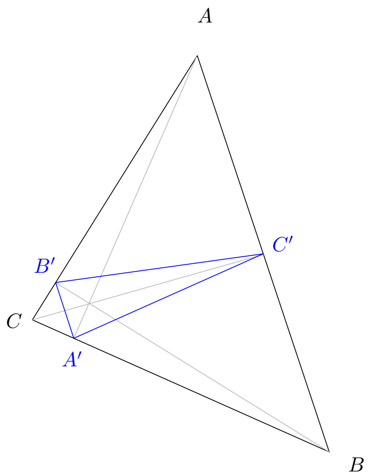
par

Romain Dujardin

Version 2021/2022.

Prologue

Soit $T = ABC$ un triangle du plan. Il y a plusieurs façons naturelles de lui associer un nouveau triangle T'. La plus simple consiste à consider le "triangle des milieux" $A'B'C'$, où A' est le milieu de $[BC]$, etc. Le théorème de Thalès montre que T' est semblable à T (avec l'orientation opposée). Si l'on itère cette transformation, on obtient ainsi une suite de triangles (T_n), tous semblables, qui converge vers le centre de gravité de T (pourquoi?). Bien sûr, si l'on s'intéresse aux triangles modulo similitude, la suite obtenue est assez pauvre. Pour gagner en richesse, il faut changer la règle de transformation $T \mapsto T'$: si A' désigne maintenant le pied de la bissectrice de l'angle $\angle BAC$ (et de même pour B' et C') on peut montrer que la suite de triangles (modulo similitude) obtenue converge vers un triangle équilatéral (exercice!).



Les choses sérieuses commencent quand le triangle T' est formé à partir des pieds des hauteurs de T (le **triangle orthique**). La suite de triangles obtenue est alors "chaotique", et a en quelque sorte la même complexité qu'une suite de variables aléatoires indépendantes identiquement distribuées sur l'ensemble $\{1, 2, 3, 4\}$. Dans le langage des systèmes dynamiques, on dit qu'elle est conjuguée à un décalage de Bernoulli sur 4 symboles(1). Pour un triangle "typique" T_0, ce caractère chaotique interdit de faire des prédictions précises sur les propriétés de T_n pour n grand **fixé**, mais permet de faire des prédictions **statistiques** sur le comportement de la suite (T_n). Par exemple, le théorème ergodique montre par exemple qu'asymptotiquement $1/4$ des triangles T_n sont aigus et permet de calculer des quantités numériques moyennes comme le quotient isopérimétrique (voir ci-dessous).

◇

On peut également à partir de T construire une suite de triangles aléatoires(2) On sait bien que les médianes T sont concourantes (et se coupent au centre de gravité de T). Si l'on découpe T selon ces médianes, on obtient 6 nouveaux triangles, (ce procédé est

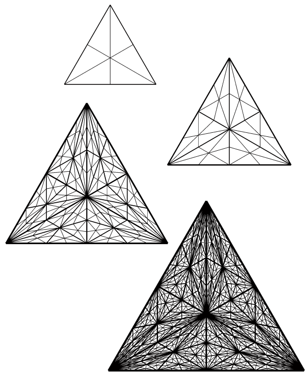
(1)Voir les articles de P. D. Lax *The Ergodic Character of Sequences of Pedal Triangles* Amer. Math. Monthly 97 (1990) pp. 377-381 et J.C. Alexander *The Symbolic Dynamics of the Sequence of Pedal Triangles* Math. Magazine 66 (1993) pp. 147-158.

(2)J'ai emprunté cet example à Amie Wilkinson: *What are Lyapunov exponents, and why are they interesting?* Bulletin of the AMS 54 (2017), 79–105.

connu sous le nom de **subdivision barycentrique**). Numérotons ces triangles de manière arbitraire en $T_1, \dots, T_6$ et itérons ce procédé: T_1 va ainsi engendrer $T_{11}, \dots, T_{16}$, et ainsi de suite. Il n'est pas difficile de se convaincre visuellement que les triangles obtenus sont de plus en plus effilés. Peut-on démontrer et quantifier cela?

Définissons le quotient isopérimétrique de T par

$$a(T) = \frac{\text{aire(T)}}{\text{périmètre}(T)^2}.$$



C'est une quantité qui mesure le taux d'aplatissement de T et qui est maximale pour le triangle équilatéral (exercice!). Soit $\varepsilon = (\varepsilon_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires de Bernouilli équiprobables dans $\{1, \dots, 6\}$. On obtient ainsi une suite de triangles aléatoires $T_{\varepsilon_1\cdots\varepsilon_n}$.

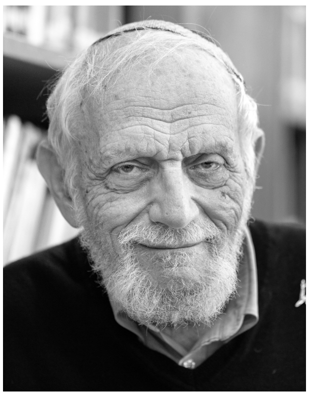
Théorème. (Bárány-Beardon-Carne, McMullen) — Presque sûrement on a

$$\frac{1}{n}\log a(T_{\varepsilon_1\cdots\varepsilon_n}) \underset{n\to\infty}{\longrightarrow} -\chi \approx -0.0894.$$

On a donc bien en général un aplatissement exponentiel des triangles itérés. Le réel χ qui apparaît ici comme taux de (dé)croissance exponentielle est un **exposant de Lyapunov**. Nous ne donnerons pas la démonstration de ce théorème dans ces notes[3] mais un des objectifs est de donner tous les éléments nécessaires à sa preuve

◇

Hillel ("Harry") Furstenberg est un mathématicien israélo-américain, né en 1935 à Berlin, qui sera sans conteste le héros de ce texte.



Il se fait remarquer très jeune par une démonstration topologique surprenante de l'infinitude de l'ensemble des nombres premiers[4], avant de faire une thèse en probabilités sous la direction de Solomon Bochner (1958). Dans les années suivantes, il écrit plusieurs articles qui deviendront des classiques de la théorie ergodique: il initie l'étude les produits aléatoires de matrices avec Harry Kesten, développe la théorie des marches aléatoires sur les groupes de Lie et la notion d'unique ergodicité. Un de ses résultats les plus célèbres est une démonstration par la théorie ergodique du théorème de Szemeredi: tout ensemble $E \subset \mathbb{N}$ de densité asymptotique strictement positive contient des progressions arithmétiques arbitrairement longues. Parmi de nombreuses récompenses internationales, il reçoit le prix

[3] Voir l'article original de I. Bárány, A. Beardon et T. Carne [Mathematika, 43 (1996), pp. 165-171] pour l'existence de χ. L'approximation numérique est due à Curt McMullen.

[4] Voir `https://images.math.cnrs.fr/Une-infinite-de-nombres-premiers-d-apres-Furstenberg.html` pour une présentation

Abel en 2020 (conjointement avec Grigory Margulis) “for pioneering the use of methods from probability and dynamics in group theory, number theory and combinatorics”.

◇

Dans ce texte nous allons dans un premier temps présenter les concepts de base de la théorie des systèmes dynamiques et de la théorie ergodique, ainsi que quelques exemples et applications emblématiques. Nous nous embarquerons ensuite pour quelques excursions dans des thématiques plus avancées, notamment en direction des systèmes dynamiques aléatoires. Un des fils conducteurs sera la notion d’exposant de Lyapunov, évoquée cidessus.

De manière générale, un système dynamique est la modélisation d’un “système” (pour nous ce sera un objet mathématique) qui évolue au cours du “temps”. On peut étudier ces systèmes de diverses façons: la théorie ergodique est l’étude des systèmes dynamiques du point de vue de la théorie de la mesure. Historiquement issu de la théorie cinétique des gaz et de “l’hypothèse ergodique” de Boltzmann, c’est maintenant un sujet à la fois bien ancré dans les mathématiques fondamentales, et également la source de nombreuses applications, en particulier pour ce qui relève des prédictions “statistiques” sur un système dynamique. C’est par exemple la théorie ergodique qui explique pourquoi on ne peut pas prévoir la météo qu’il fera dans deux semaines, mais que l’on peut savoir avec une bonne précision si la saison qui s’annonce sera plus ou moins pluvieuse, ou plus ou moins chaude(5).

Dans les chapitres I à V nous passons en revue les concepts fondamentaux du sujet: notions de système dynamique topologique et mesuré, ergodicité, mélange, les exemples classiques, les théorèmes ergodiques de Birkhoff, Von Neumann et Kingman, et la notion d’unique ergodicité. Une bonne partie de ce contenu est standard et (mieux) exposé dans les ouvrages classiques du domaine (comme par exemple [**6**], [**14**] ou [**4**], dont nous nous sommes largement inspirés). Cette partie est néanmoins parsemée de résultats moins classiques, souvent présentés sous une version simplifiée (comme par exemple les théorèmes ergodiques “non-conventionnels”).

Les chapitres VI et VII sont consacrés à l’étude de la notion d’exposant de Lyapunov et aux produits aléatoires de matrices. Dans un souci de simplicité on se concentre sur le cas de la dimension 2 ou les idées importantes apparaissent déjà, mais le formalisme est plus accessible que dans le cas général (pour lequel on pourra consulter les ouvrages [**3**] [**13**] et [**1**]).

Le chapitre VIII porte sur la notion d’entropie métrique. Son point culminant est une preuve de l’inégalité de Ruelle qui lie entropie et exposants de Lyapunov.

Au chapitre IX nous présentons la construction de la mesure “physique” pour les applications markoviennes dilatantes de l’intervalle.

(5)On pourra consulter avec profit le très joli texte de M. Hochman `http://math.huji.ac.il/%7Emhochman/research-expo.html` pour un panorama des grandes idées de la théorie générale des systèmes dynamiques.

Enfin les chapitres X et XI sont consacrés la théorie ergodique des systèmes dynamiques aléatoires, qui est un thème plus rarement abordé à ce niveau. Nous y étudierons en particulier les notions de mesure invariante et stationnaire, leurs exposants de Lyapunov, et le “principe d’invariance” qui affirme qu’une mesure stationnaire non invariante doit présenter des propriétés de contraction.

Un leitmotif de ce texte est de mettre l’accent sur les grandes idées de la théorie ergodique, en limitant autant que possible les difficultés techniques. Les résultats classiques sont pour la plupart démontrés en détail; néanmoins nous évoquerons un certain nombre de résultats plus avancés, présentés sans démonstration, ou sous des hypothèses très simplificatrices, qui sont autant d’ouvertures vers des thématiques de recherche contemporaines.

$\diamond$

Merci aux étudiants des cours 2018-2021, en particulier à Orphée Collin et Nadège Couvert, pour leurs remarques et corrections qui ont permis de grandement améliorer ce texte.

$\diamond$

Sommaire

I. GÉNÉRALITÉS SUR LES SYSTÈMES DYNAMIQUES

1. Notion de système dynamique

1.1. *Définition.* — Un système dynamique à temps discret est la donnée d'un couple (X, f) constitué d'un ensemble X et d'une application $f : X \to X$.

Étudier le système dynamique (X, f) consiste à étudier les compositions itérées $f^{\circ n} := f \circ \cdots \circ f$, ou en d'autres termes, la famille des suites récurrentes définies par $x_{n+1} = f(x_n)$, où $x_0 \in X$ est arbitraire. En pratique une compréhension exhaustive des itérées f^n est en général impossible et le principe général de la théorie ergodique est de se contenter de l'étude du comportement dynamique de **presque tout** x_0, en un sens que nous allons préciser dans les chapitres suivants.

En pratique X est en général muni de structures supplémentaires (topologique, mesurable, différentiable, algébrique, etc.) préservées par f. C'est l'interaction de la dynamique avec ces structures qui est le sel de la théorie et fournit les informations intéressantes. Un cadre qui nous intéressera particulièrement est celui des **systèmes dynamiques mesurables** $(X, \mathcal{A}, f)$ correspondant à la donnée d'une application mesurable f sur un espace mesurable $(X, \mathcal{A})$.

1.2. — Si f est inversible, pour $n \geqslant 1$ on note $f^{-n} = (f^{-1})^n$ et $f^0 = \mathrm{id}_X$ de sorte que pour tous $m, n \in \mathbb{Z}$ on a $f^{m+n} = f^m \circ f^n$. En d'autres termes on a un morphisme de groupes $(\mathbb{Z}, +) \to (\mathfrak{S}(X), \circ)$ défini par $n \mapsto f^n$.

1.3. — Pour $x_0 \in X$ on note $\mathcal{O}^+(x_0) = \{f^n(x_0),\ n \in \mathbb{N}\}$ l'**orbite positive** de x_0 (on dira en général simplement orbite; noter la terminologie liée à la mécanique céleste) et si f est inversible $\mathcal{O}(x_0) = \{f^n(x_0),\ n \in \mathbb{Z}\}$ est l'**orbite totale** de x_0. On dit que x_0 est **périodique** s'il existe $n \geqslant 1$ tel que $f^n(x_0) = x_0$. Le plus petit tel n est la **période** de x_0. Un point fixe est un point périodique de période 1. Enfin, on dit que $A \subset X$ est **(positivement) invariant** si $f(A) \subset A$ et **totalement invariant** si $f^{-1}(A) = A$.

1.4. — Dans ce cours nous considèrerons principalement des systèmes dynamiques à temps discret (plus précisément, à temps entier). Néanmoins il y a d'autres cadres intéressants:

Définition. — Un système dynamique à temps continu est la donnée d'une famille à un paramètre $(f^t)_{t \in \mathbb{R}^+}$ de transformations d'un ensemble X telle que pour tous t, t' on a $f^{t+t'} = f^t \circ f^{t'}$. Si (f^t) est défini pour tout $t \in \mathbb{R}$ on parle de **flot**.

L'exemple fondamental est donné par l'intégration des champs de vecteurs. Par exemple si $h : \mathbb{R}^d \to \mathbb{R}^d$ est une application globalement Lipschitzienne, i.e.

$$\exists C > 0,\ \forall x, y,\ \|h(x) - h(y)\| \leqslant C\,\|x - y\|$$

alors par le théorème de Cauchy-Lipschitz pour tout $x_0 \in \mathbb{R}^d$ il existe une unique solution de l'équation différentielle $x' = h(x)$, définie pour tout $t \in \mathbb{R}$, et telle que $x(0) = x_0$.

L'application qui à $(t, x_0) \in \mathbb{R}^{d+1}$ associe la valeur $\phi^t(x_0)$ en t de cette solution définit un flot de classe C^1.

1.5. — Si $(f^t)_{t\geqslant 0}$ est un système dynamique à temps continu alors pour tout t_0 $(f^{t_0 n})_{n\geqslant 0}$ est un système dynamique à temps discret qui partage beaucoup de propriétés avec (f^t). Si $t_0 = 1$ on le désigne en général sous le nom de **temps 1** du flot.

Une construction un peu plus sophistiquée, mais aussi plus utile, est celle de la **section de Poincaré**. Supposons qu'on ait u sous-ensemble $A \subset X$ avec la propriété que pour tout $x \in X$, l'ensemble des $t \in \mathbb{R}$ tel que $f^t(x) \in A$ est discret. Dans ce cas pour $a \in A$ on définit le **temps de retour** $\tau(a) = \inf t > 0,\ f^t(a) \in A$ (qui peut être infini) et **l'application de premier retour** (ou application de Poincaré) $f_A : a \mapsto f^{\tau(a)}(a)$ (qui n'est définie que là où $\tau(a) < \infty$).

En pratique, lorsque (f^t) est le flot d'une équation différentielle sur $\mathbb{R}^d$ on prend pour A une sous-variété locale de dimension $d-1$ transverse au champ de vecteur, le plus souvent près d'une orbite périodique, qui est donc un point fixe de l'application de premier retour. La compréhension des orbites voisines de cette orbite périodique se ramène à celle de l'itération de f_A au voisinage de ce point fixe.

1.6. — Réciproquement, si $f : X \to X$ est une application bijective, on introduit l'ensemble $Y = X \times \mathbb{R}/\sim$, où $\sim$ est la relation d'équivalence définie de la façon suivante: $(x, t) \sim (x', t')$ si et seulement s'il existe $k \in \mathbb{Z}$ tel que $t' = t - k$ et $x' = f^k(x)$. La translation selon la deuxième variable définit alors un flot sur Y dont le temps 1 restreint à l'image de $X \times \{0\}$ est (conjugué à) f. Ce flot s'appelle la **suspension** de f.

Exercice. — Démontrer ce fait.

Exercice. — Est ce que tout système dynamique à temps discret sur $\mathbb{R}^n$ est le temps 1 d'un champ de vecteurs? (réponse: non)

1.7. — Plus généralement si G est un groupe infini et ϕ est une **action** de G sur X, c'est-à-dire un morphisme $G \to \mathfrak{S}(X)$, on peut étudier les propriétés "asymptotiques" de cette action par des techniques issues des systèmes dynamiques. C'est un immense champ d'application de la théorie ergodique, que nous aborderons notamment dans la deuxième partie de ces notes.

1.8. — Il y a une notion d'isomorphisme entre deux systèmes dynamiques, définie comme suit: $(X, f) \sim (Y, g)$ s'il existe une bijection $\varphi : X \to Y$ telle que $\varphi \circ f = g \circ \varphi$. On dit alors que (X, f) et (Y, g) sont **conjugués**. Si une telle application φ existe mais n'est que surjective, on dit que (Y, g) est un **facteur** de (X, f) (resp. (X, f) est une **extension** de (Y, g)).

Exercice. — Montrer qu'une conjugaison envoie les orbites de f sur les orbites de g.

Cette notion n'a en fait d'intérêt véritable que si la régularité de la conjugaison est précisée: on parle alors de conjugaison mesurable, topologique, etc. Noter que la conjugaison φ est bien souvent beaucoup moins régulière que les systèmes dynamiques de départ: l'existence d'une conjugaison mesurable entre deux systèmes dynamiques différentiables est déjà en soi une information intéressante.

1.9. *Exercice.* — Soit (X, f) un système dynamique non inversible, tel que f est surjective. Montrer qu'il admet une extension inversible $(\widehat{X}, \widehat{f})$, qui est minimale au sens où toute extension inversible (Y, g) se factorise à travers $(\widehat{X}, \widehat{f})$. (indication: poser $\widehat{X} \subset X^{\mathbb{Z}}$ l'ensemble des suites $(x_n)_{n\in\mathbb{Z}}$ telles que pour tout $n \in \mathbb{Z}$ on a $f(x_n) = x_{n+1}$ et $\widehat{f}$ le décalage défini par $\widehat{f}((x_n)) = (y_n)$, où $y_n = x_{n+1}$). Ceci s'appelle l'**extension naturelle** de f. Montrer que si en outre X est compact et f est continue, alors il en est de même pour $(\widehat{X}, \widehat{f})$.

2. Vocabulaire des systèmes dynamiques topologiques

Dans cette section (X, f) est un **système dynamique topologique**, c'est à dire que X est un espace topologique et f est continue. En pratique nous ne chercherons pas la généralité maximale et supposerons toujours par commodité X métrisable, et le plus souvent localement compact.

2.1. — Soit $x \in X$. On dit que $y \in X$ est un **point ω-limite** de x s'il existe une suite (n_j) infinie telle que $f^{n_j}(x) \to y$ quand $j \to \infty$. On notera $\omega_f(x)$ ou plus simplement $\omega(x)$ l'ensemble des points ω-limites de x. Si f est inversible on définit de même l'ensemble **α-limite** des points tels que $f^{-n_j}(x) \to y$. Un point x est dit **récurrent** si $x \in \omega(x)$.

Exercice. — Montrer que $\omega_f(x)$ est un fermé invariant.

2.2. — On dit que x est **non-errant** si pour tout voisinage U de x il existe $n \geqslant 1$ tel que $f^n(U) \cap U \neq \varnothing$, et **errant** sinon. On note en général $NW(f)$ l'ensemble des points non-errants.

Exercice. — Montrer que $NW(f)$ est un fermé invariant et que récurrent implique non errant. (La réciproque est fausse: chercher un contre exemple ou attendre le §3.2.2.)

2.3. *Définition.* — On dit que (X, f) est topologiquement transitif s'il existe $x \in X$ dont l'orbite est dense.

Exercice. — Montrer que si X n'a pas de point isolé alors $\mathcal{O}^+(x)$ dense équivaut à $\omega(x) = X$.

2.4. *Théorème.* — *Supposons que X est un espace métrique complet séparable sans point isolé. Alors:*

1. *(X, f) est topologiquement transitif ssi pour toute paire d'ouverts U, V non vides il existe $n \geqslant 0$ tel que $f^{-n}(U) \cap V \neq \varnothing$.*
2. *Si (X, f) est topologiquement transitif alors $\left\{x,\ \overline{\mathcal{O}^+(x)} = X\right\}$ est dense.*

Démonstration. — Pour le premier point, prenons x ayant une orbite dense. D'après l'exercice précédant l'énoncé, $\omega(x)$ est dense. En particulier x admet une infinité d'itérés dans U et une infinité d'itérés dans V. Il existe donc $k \geqslant 0$ tel que $f^k(x) \in U$ et $\ell \geqslant k$ tel que $f^\ell(x) \in V$, et ainsi $f^{-(\ell-k)}(V) \cap U$ contient $f^k(x)$ et est donc non-vide.

Réciproquement, nous allons montrer que si toute paire d'ouverts U, V non vides il existe $n \geqslant 0$ tel que $f^{-n}(U) \cap V \neq \varnothing$, alors $\left\{x,\ \overline{\mathcal{O}^+(x)} = X\right\}$ est dense, ce qui démontrera la réciproque du point 1. ainsi que le point 2. Pour cela on considère une base dénombrable d'ouverts $(U_k)_{k\in\mathbb{N}}$. Une orbite est dense si et seulement si elle visite tous les U_k. Ainsi

$$\left\{x, \overline{\mathcal{O}^+(x)} = X\right\} = \{x,\ \forall k \geqslant 0,\ \exists n \geqslant 0,\ f^n(x) \in U_k\} = \bigcap_{k\geqslant 0} \bigcup_{n\geqslant 0} f^{-n}(U_k).$$

Notre hypothèse dit que pour tout ouvert V et tout entier k, il existe n tel que $V \cap f^{-n}(U_k) \neq \varnothing$, c'est-à-dire que $V \cap \bigcup_{n\geqslant 0} f^{-n}(U_k) \neq \varnothing$. Ceci signifie que l'ouvert $\bigcup_{n\geqslant 0} f^{-n}(U_k)$ est dense. Ainsi par le théorème de Baire, $\bigcap_{k\geqslant 0} \bigcup_{n\geqslant 0} f^{-n}(U_k)$ est un G_δ dense, ce qui termine la preuve. □

2.5. *Exercice.* — On dit que f est exacte si pour tout ouvert U non vide alors pour n assez grand $f^n(U) = X$. Montrer qu'alors f est topologiquement transitive.

2.6. — Supposons maintenant que X est un espace métrique(1) compact.

Définition. — On dit que (X, f) est minimal s'il n'y a pas de fermé invariant non-trivial, i.e. tout fermé invariant est soit vide, soit X.

2.7. *Proposition.* — (X, f) *est minimal si et seulement si toutes ses orbites sont denses. De même pour tout* x *on a* $\omega(x) = X$.

Démonstration. —

$\Rightarrow$. Pour tout $x \in X$, $\overline{\mathcal{O}^+(x)}$ (resp. $\omega(x)$) est un fermé invariant non trivial, donc égal à X.

$\Leftarrow$. Si $Y \subset X$ est un fermé invariant non vide, alors pour tout $x \in Y$ on a $X = \overline{\mathcal{O}^+(x)} \subset Y$. □

2.8. — Plus généralement on dit que $Y \subset X$ est un **sous-ensemble minimal** si c'est un fermé invariant non vide minimal pour l'inclusion.

2.9. *Proposition.* — *Si* (X, f) *est un système dynamique topologique avec* X *compact alors il existe un sous-ensemble minimal.*

Démonstration. — Considérons l'ensemble $\mathcal{C}$ des sous ensembles fermés invariants non-vides. C'est un ensemble non-vide et ordonné par l'inclusion. Si $\mathcal{K} \subset \mathcal{C}$ est un sous-ensemble totalement ordonné, on a par compacité que $\bigcap_{K\in\mathcal{K}} K \neq \varnothing$ (dans un compact, une intersection décroissante de fermés non-vides est non-vide). Cet ensemble est également fermé et invariant, c'est donc un minorant de $\mathcal{K}$. On peut ainsi appliquer le lemme de Zorn qui fournit l'élément minimal souhaité. □

2.10. — Le corollaire suivant est parfois désigné sous le nom de "Théorème de récurrence de Birkhoff".

Corollaire. — *Si* X *est compact et* $f : X \to X$ *alors il existe un point récurrent.*

(1) En fait topologique suffirait, comme pour beaucoup de résultats invoquant la compacité dans la suite.

Démonstration. — On prend un fermé invariant minimal $Y \subset X$. Pour tout $y \in Y$ on a $\omega(y) = Y \ni y$, d'où le résultat. □

2.11. — Il est clair que si (X, f) est minimal, tout point est récurrent. On peut en fait être plus précis: on dit que $x \in X$ est **presque périodique** si pour tout $\varepsilon > 0$, l'ensemble $\{n \in \mathbb{N},\ d(f^n(x), x) < \varepsilon\}$ est **syndétique**, c'est à dire qu'il existe N tel que tout intervalle de son complémentaire est de longueur majorée par N. De manière équivalente, il existe une suite d'entiers (n_i) avec $n_{i+1} - n_i \leqslant N + 1$, tels que pour tout i, $d(f^{n_i}(x), x) < \varepsilon$.

Exercice. — Construire un exemple de système dynamique possédant un point récurrent qui n'est pas presque périodique.

2.12. *Proposition. — Si (X, f) est un système dynamique compact et minimal alors tout point est presque périodique.*

Démonstration. — Raisonnons par l'absurde et supposons qu'il existe un point x non presque périodique. Alors il existe $\varepsilon > 0$ tel que pour tout N il existe n_N tel que pour tout $j \in \{n_N - N, \ldots, n_N + N\}$, on a $d(f^j(x), x) \geqslant \varepsilon$. Soit y une valeur d'adhérence de la suite $(f^{n_N}(x))_{N\geqslant 1}$. Par continuité de f, on voit que pour tout $j \in \mathbb{Z}$, $d(f^j(y), x) \geqslant \varepsilon$, et ainsi $\overline{\mathcal{O}(y)} \not\ni x$, ce qui contredit la minimalité de f. □

3. Exemples classiques

3.1. Les rotations du cercle. — Considérons le cercle $\mathbb{S}^1$, vu comme le cercle unité du plan complexe (avec la topologie induite) et définissons $R_\alpha : \mathbb{S}^1 \to \mathbb{S}^1$ par $R_\alpha(z) = e^{2i\pi\alpha} z$.

Exercice. — Montrer que cette application est conjuguée à la translation $x \mapsto x + \alpha$ dans le groupe abélien $\mathbb{R}/\mathbb{Z}$.

3.1.1. *Théorème. —*

- *Si $\alpha \in \mathbb{Q}$ alors R_α est périodique: $R_\alpha^q = \mathrm{id}$ où $\alpha = \frac{p}{q}$.*
- *Si $\alpha \notin \mathbb{Q}$ alors R_α est minimale.*

Démonstration. — Le premier point est évident. Pour le deuxième point on remarque d'abord que comme α est irrationnel, si n et m sont des entiers relatifs distincts on a pour tout z, $R_\alpha^n(z) \neq R_\alpha^m(z)$. Fixons un entier naturel N grand et découpons $\mathbb{S}^1$ en N parties par les racines N^{e} de l'unité. Par le principe des tiroirs il existe $n \neq m$ tels que $R_\alpha^n(z)$ et $R_\alpha^m(z)$ sont dans la même pièce de cette partition, i.e. $|\mathrm{Arg}(2i\pi(n-m)\alpha| < 1/N$. Pour $k = n - m$ on voit donc que R_α^k est une rotation d'angle $< 1/N$ et ainsi l'orbite de tout $z \in \mathbb{S}^1$ pour R_α^k –et donc **a fortiori** pour R_α– est $1/N$ dense. Comme ceci est vrai pour tout N on conclut que toutes les orbites sont denses. □

3.2. Multiplication de l'angle sur le cercle. — Pour $m \geqslant 2$ un entier on pose M_m l'application définie par

$$\begin{aligned} M_m : \mathbb{S}^1 &\longrightarrow \mathbb{S}^1 \\ z &\longmapsto z^m \end{aligned}$$

Exercice. — Montrer que cette application est conjuguée à la multiplication par m: $x \mapsto mx$ sur le tore $\mathbb{R}/\mathbb{Z}$.

Remarquer que contrairement aux rotations qui préservent les distances, M_m est "dilatante" (sa dérivée en tout point est supérieure à 1).

3.2.1. *Théorème.* — *M_m est topologiquement transitive mais pas minimale.*

Démonstration. — M_m n'est pas minimale car elle a $m-1$ points fixes solutions de $z^m = z$ sur $\mathbb{S}^1$.

Pour établir la transitivité topologique, donnons nous deux intervalles I et J non-triviaux sur le cercle et montrons que $M_m^{-n}(I) \cap J$ est non vide pour tout n assez grand[(2)] En effet si l'on pose $J =]e^{ia}, e^{ib}[$ on a

$$\begin{aligned} M_m^{-n}(I) &= \left\{z \in \mathbb{S}^1,\ z^{m^n} \in I\right\} \\ &= \{\theta \in \mathbb{R}/\mathbb{Z},\ m^n\theta \in]a, b[\} \\ &= \bigcup_{k=0}^{m^n-1} \left]\frac{a}{m^n} + \frac{k}{m^n}, \frac{b}{m^n} + \frac{k}{m^n}\right[, \end{aligned}$$

soit m^n intervalles régulièrement espacés sur le cercle. On voit donc que $M^{-n}(I)$ intersecte J dès que $\frac{1}{m^n} < b - a$. □

3.2.2. *Exercice.* — Montrer qu'il existe des points non-errants pour M_m qui ne sont pas récurrents.

3.2.3. *Exercice.* — Soit $f : [0,1] \to [0,1]$ l'application "tente" définie par

$$\begin{cases} f(x) = 2x \text{ si } x \leqslant 1/2 \\ f(x) = 2 - 2x \text{ si } x \geqslant 1/2 \end{cases}.$$

Montrer que f est topologiquement transitive et non minimale.

3.3. Systèmes dynamiques symboliques. — On fixe un entier $d \geqslant 2$ et on pose $\mathfrak{A} = \{0, \dots, d-1\}$ un ensemble de cardinal d appelé **alphabet**. On pose $\Sigma_+ = \mathfrak{A}^{\mathbb{N}}$ (resp. $\Sigma = \mathfrak{A}^{\mathbb{Z}}$) l'ensemble des suites infinies (resp. bi-infinies) de lettres de $\mathfrak{A}$.

3.3.1. *Proposition.* — *Les espaces Σ et Σ_+ munis de la topologie produit de la topologie discrète sur $\mathfrak{A}$ sont compacts.*

(2) Remarquer que cette propriété est plus forte que celle du Théorème 2.4, on l'appelle **mélange topologique**. La terminologie sera claire au chapitre II. En fait M_m est même **topologiquement exacte**: pour tout intervalle I on a $M_m^n(I) = \mathbb{S}^1$ pour n assez grand.

Démonstration. — C'est le théorème de Tychonoff. On peut le démontrer dans ce cas précis de manière élémentaire en introduisant la distance d définie sur Σ_+ par $d(x,y) = 2^{-\min\{i,\ x_i \neq y_i\}}$ (i.e. $d(x,y) = 2^{-n}$ si x et y ont un préfixe commun de longueur exactement n). Alors d métrise la topologie produit sur Σ_+ et l'extraction diagonale montre que (Σ_+, d) est compact. (Le cas de Σ est analogue). □

Remarquer qu'une base d'ouverts de Σ_+ (resp. Σ) est donnée par les **cylindres** $\Sigma_{\alpha_0,\ldots,\alpha_n} = \{x \in \Sigma_+,\ x_0 = \alpha_0, \ldots, x_n = \alpha_n\}$(resp. $\Sigma_{\alpha_{-m},\ldots,\alpha_n} = \{x_{-m} = \alpha_{-m}, \ldots, x_n = \alpha_n\}$).

3.3.2. — On introduit les **décalages** (**shifts**) unilatéral et bilatéral, respectivement définis par

$$\begin{aligned} \sigma : \quad \Sigma_+ &\longrightarrow \Sigma_+ \\ x = (x_i)_{i \geqslant 0} &\longmapsto (x_{i+1})_{i \geqslant 0} \end{aligned} \qquad \text{et} \qquad \begin{aligned} \sigma : \quad \Sigma &\longrightarrow \Sigma \\ (x_i) &\longmapsto (y_i) \text{ où } y_i = x_{i+1}. \end{aligned}$$

3.3.3. *Théorème.* — *Les systèmes dynamiques* (Σ_+, σ) *et* (Σ, σ) *sont topologiquement transitifs et non-minimaux.*

Démonstration. — Comme précédemment on n'a pas minimalité car il y a des orbites périodiques. La transitivité se vérifie avec des cylindres: soient $\Sigma_{\alpha_0,\ldots,\alpha_n}$ et $\Sigma_{\beta_0,\ldots,\beta_m}$ des cylindres arbitraires alors $x \in \sigma^{-N}(\Sigma_{\beta_0,\ldots,\beta_m})$ ssi $x_N = \beta_0, \ldots, x_{N+m} = \beta_m$. On voit que si $N > n$, toute suite ayant $\alpha_0, \ldots, \alpha_n$ comme préfixe et $\beta_0, \ldots, \beta_m$ en positions respectives $N, \ldots, N+m$ appartient à $\Sigma_{\beta_0,\ldots,\beta_m} \cap \sigma^{-N}(\Sigma_{\beta_0,\ldots,\beta_m})$. □

3.3.4. *Exercice.* —

1. Calculer le nombre de points de période N de (Σ_+, σ).
2. Montrer que si $d \neq d'$ les décalages sur d et d' symboles ne sont pas topologiquement conjugués.

3.3.5. *Exercice.* — Soit μ_0 la mesure équidistribuée sur $\mathfrak{A}$ et $\mu = \mu_0^{\mathbb{N}}$ la mesure produit correspondante sur Σ_+. Montrer que μ-presque tout point a une orbite dense. Montrer que l'ensemble des x dont l'orbite n'est pas dense est non-dénombrable.

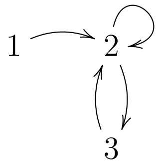
3.3.6. — Pour $x \in [0,1]$ on considère son développement en base d (si x est de la forme $\frac{q}{d^k}$ il y a deux développements possibles). On obtient ainsi une surjection $\tilde{h} : \Sigma_+ \to [0,1]$ définie par $\tilde{h}(x) = \sum \frac{x_i}{d^i}$. Soit $h = \pi \circ \tilde{h}$ où π est l'application naturelle $[0,1] \to \mathbb{R}/\mathbb{Z}$ (identifié à $\mathbb{S}^1$). On a un diagramme commutatif

$$\begin{tikzcd} \Sigma_+ \arrow[r, "\sigma"] \arrow[d, "h"'] & \Sigma_+ \arrow[d, "h"] \\ \mathbb{S}^1 \arrow[r, "M_d"'] & \mathbb{S}^1 \end{tikzcd}$$

qui montre que $(\mathbb{S}^1, M_d)$ est un facteur de (Σ_+, σ). Comme h est injective hors d'un ensemble dénombrable on peut "presque" le considérer comme une conjugaison. Ainsi on voit que la multiplication par d sur le cercle est **aussi imprévisible qu'une suite de variables aléatoires équidistribuée sur** $\{0, \ldots, d-1\}$!

3.3.7. *Exercice.* — Montrer que presque tout point de $\mathbb{S}^1$ (pour la mesure de Lebesgue) a une orbite dense pour M_d.

3.3.8. — Sous décalages de type fini. Si l'on se donne un graphe Γ à d sommets numérotés de 1 à d comme par exemple:



alors on peut définir un sous-ensemble Σ_Γ de (Σ, σ) (ou (Σ_+, σ)) en décrétant que $x \in \Sigma_\Gamma$ ssi pour tout $i \in \mathbb{Z}$ il y a une arête dans le graphe Γ joignant x_i et x_{i+1}.

Exercice. — Montrer que Σ_Γ est un fermé invariant.

3.3.9. *Proposition. — Soit A la matrice d'adjacence du graphe (i.e. $a_{i,j} = 1$ si $i \to j$, 0 sinon). Alors pour tous i, j, A^N_{ij} est le nombre de chemins de longueur N dans le graphe démarrant en i et terminant en j.*

Exercice. — Montrer que le nombre de points périodiques de période N de (Σ_Γ, σ) est $\mathrm{tr}(A^N)$.

3.3.10. *Définition.* — On dit que:

- A est irréductible si $\forall i, j,\ \exists N,\ A^N_{ij} > 0$;
- A est primitive si $\exists N,\ \forall i, j,\ A^N_{ij} > 0$.

Exercice. — Montrer que si A est irréductible, (Σ_Γ, σ) est topologiquement transitif. Montrer également que sous cette hypothèse les points périodiques sont denses.

3.4. Applications linéaires des tores.

— On ne va pas traiter la théorie générale mais se concentrer sur un exemple fameux: le **chat d'Arnol'd**[3].

3.4.1. — Soit $\mathbb{T}^2 = \mathbb{R}^2/\mathbb{Z}^2$ le tore bidimensionnel et $M \in M_2(\mathbb{Z})$. Alors M induit une application $f_M : \mathbb{T}^2 \to \mathbb{T}^2$: en effet pour tout $x \in \mathbb{R}^2$ on a $M(x + \mathbb{Z}^2) = Mx + M\mathbb{Z}^2 = Mx + \mathbb{Z}^2$. Si en outre $M \in \mathrm{SL}_2(\mathbb{Z})$ alors $M^{-1} \in M_2(\mathbb{Z})$ et f_M est inversible. On note π la projection canonique $\mathbb{R}^2 \to \mathbb{T}^2$.

On fixe dorénavant $M = \left(\begin{smallmatrix} 2 & 1 \\ 1 & 1 \end{smallmatrix}\right)$. C'est une matrice de $\mathrm{SL}_2(\mathbb{Z})$ dont les valeurs propres sont $\lambda^u = \frac{3+\sqrt{5}}{2} > 1$ et $\lambda^s = \frac{3-\sqrt{5}}{2} < 1$. Soient e^u et e^s des vecteurs propres associés. Noter que leurs pentes sont irrationnelles.

3.4.2. *Proposition. — $x \in \mathbb{T}^2$ est périodique si et seulement si $x \in \mathbb{Q}^2/\mathbb{Z}^2$. En particulier les points périodiques sont denses dans $\mathbb{T}^2$.*

---

(3) Ainsi nommé à cause de son illustration dans le livre de Arnol'd et Avez *Problèmes ergodiques de la mécanique classique.*

Démonstration. — $\Leftarrow$. Si $X \in \mathbb{Q}^2$ alors la suite $M^n X$ est à dénominateurs bornés donc sa projection x dans $\mathbb{T}^2$ est finie. Ainsi x est prépériodique, donc périodique car f est inversible (exercice).

$\Rightarrow$. Si x est périodique alors un représentant X de x dans $\mathbb{R}^2$ doit vérifier une équation du type $M^n X = X + U$, avec $U \in \mathbb{Z}^2$. On en déduit que $X = (M^n - I_2)^{-1} U$ donc comme $M^n - I_2$ est à coefficients entiers il vient $X \in \mathbb{Q}^2$. □

3.4.3. *Théorème. — f_M est topologiquement transitive.*

Démonstration. — Nous allons montrer que si U et V sont des ouverts de $\mathbb{T}^2$ alors $f^n(U) \cap V \neq \varnothing$ pour n assez grand. Il suffit de considérer le cas où U et V sont des projections de petits carrés de cotés parallèles à e^u et e^s, i.e. de la forme $\{x_0 + ae^u + be^s,\ |a| < r, |b| < r\}$. M^n agit sur un tel carré en l'envoyant sur

$$M^n(Q) = (\text{la projection de})\ M^n x_0 + a\left(\frac{3+\sqrt{5}}{2}\right)^n e^u + b\left(\frac{3-\sqrt{5}}{2}\right)^n e^s,\ |a| < r, |b| < r.$$

C'est un rectangle très fin et étiré dans la direction de e^u.

3.4.4. *Lemme. — Si $D \subset \mathbb{R}^2$ est une droite de vecteur directeur e^u alors $\pi(D)$ est dense dans $\mathbb{T}^2$.*

Démonstration. — Les points d'intersection de $\pi(D)$ avec $\pi(\{0\}\times\mathbb{R})$ sont les orbites d'une rotation R_α, où α est la pente de la droite, ici $\frac{\sqrt{5}-1}{2}$, donc ils sont denses sur $\pi(\{0\} \times \mathbb{R})$. Ensuite si l'on se place dans un domaine fondamental $[0,1[^2$ de π, on voit que $\pi(D)$ est l'image réciproque de cet ensemble dense par la projection parallèlement à e^u, donc $\pi(D)$ est également dense. □

Pour conclure la démonstration de la transitivité topologique, il faut préciser ce lemme de la manière suivante: pour tout $\varepsilon > 0$, l'image d'un segment de longueur ℓ dans une droite de vecteur directeur e^u est ε-dense dans $\mathbb{T}^2$ pour ℓ assez grand. On en déduit ainsi que $f^n(U)$ est ε-dense dans $\mathbb{T}^2$ pour n assez grand, et finalement $f^n(U) \cap V \neq \varnothing$ si ε est choisi tel que V contienne une boule de rayon ε. □

Exercice. — Remplir tous les détails du raisonnement ci dessus.

3.5. Polynômes quadratiques. — Contrairement à ce qu'on pourrait imaginer, l'étude des suites récurrentes sur $\mathbb{R}$ du type $u_{n+1} = f(u_n)$ où f une une fonction "élémentaire", est en général un problème très difficile et profond. Le premier cas non-trivial est celui où f est un polynôme de degré 2. L'exercice suivant montre qu'on peut toujours se ramener à une forme simple (les experts disent "forme normale") pour f.

3.5.1. *Exercice.* — Montrer que si f est un polynôme de degré 2, le système dynamique défini par f sur $\mathbb{R}$ est conjugué à $f_c : \mathbb{R} \to \mathbb{R}$, où $f_c(x) = x^2 + c$. Si en outre $c \leqslant 1/4$ et $c \neq 0$, montrer que f_c est conjugué à $x \mapsto \alpha x(1-x)$.

En particulier la famille des polynômes quadratiques vus comme des systèmes dynamiques est de dimension 1. L'exercice suivant montre que pour espérer une dynamique intéressante il faut se placer dans le cas $c \leqslant 1/4$.

3.5.2. *Exercice.* — Montrer que si $c > 1/4$, pour tout $x \in \mathbb{R}$, $f_c^n(x) \to \infty$.

Pour $c \leqslant 1/4$ notons $\alpha \leqslant \beta$ les points fixes de f_c (qui sont confondus pour $c = 1/4$), et remarquer que $f_c(-\beta) = \beta$. On note I_β l'intervalle $[-\beta, \beta]$. L'exercice suivant regroupe un certain nombre de propriétés élémentaires de f_c.

3.5.3. *Exercice.* —

(i) Montrer que pour tout $c \leqslant 1/4$, si $x \notin I_\beta$, alors $f^n(x) \to \infty$.
(ii) Montrer que pour $-2 \leqslant c \leqslant 1/4$, l'intervalle I_β est stable par f.
(iii) Montrer que pour $-3/4 \leqslant c < 1/4$, pour tout $x \in]-\beta, \beta[$, $f^n(x)$ converge vers un point fixe attractif de f. Que se passe-t-il pour $c = 1/4$.

On pourra remarquer qu'avec la normalisation $\alpha x(1-x)$, dans la plage $0 < \alpha \leqslant 4$, l'intervalle $[0,1]$ est stable.

Lorsque c traverse le paramètre $-3/4$, on montre assez simplement que le point fixe attractif “explose” en un cycle attractif de période 2, qui attire tous les points de $]-\beta, \beta[$. Si on continue à diminuer c, on constate une succession de telles bifurcations en des paramètres c_n, où la période du cycle attractif limite double successivement. La suite (c_n) converge vers un paramètre $c_\infty \simeq -1,401155$ nommé **point de Feigenbaum**. L'étude de la dynamique de f_c dans I_β pour $-2 < c \leqslant c_\infty$ est un monument de la théorie des systèmes dynamiques, qui fait appel à des idées d'une beauté et d'une sophistication tout à fait inattendues [(4)]. Un aspect essentiel de cette étude est de considérer f_c comme une transformation **holomorphe** du plan complexe. Dans la suite de cette section, nous allons montrer qu'inversement dans le cas $c \leqslant 2$ on peut décrire complètement la dynamique de f_c.

3.5.4. — Le cas $c = -2$.

Pour $c = -2$ l'intervalle $I = [-2, 2]$ est stable par f. Soit π l'application de $\mathbb{R}/\mathbb{Z}$ dans I définie par $\pi(\theta) = 2\cos(2\pi\theta)$. La formule de doublement du cosinus $\cos(2\alpha) = 2\cos^2(\alpha) - 1$ montre que

$$f_{-2} \circ \pi(\theta) = 2(2\cos^2(2\pi\theta) - 1) = 2\cos(2\pi \cdot 2\theta) = \pi \circ M_2(\theta),$$

où comme précédemment M_2 désigne la multiplication par 2 sur le cercle. Ainsi $z^2 - 2$ est un facteur de M_2, et on en déduit que $z^2 - 2$ est topologiquement transitive, et également semi-conjuguée à un décalage complet sur deux symboles. La dynamique est donc **chaotique** sur $[-2, 2]$.

3.5.5. — Le cas $c < -2$.

Dans le cas où $c < -2$, l'intervalle I n'est plus stable par f. Nous avons par ailleurs vu dans la question (i) de l'exercice précédent que si $x \notin I$ alors $f^n(x) \to \infty$. On définit

$$K = K(f) = \{x \in \mathbb{R} : \ \forall n \geqslant 0, \ f^n(x) \in I\} = \bigcap_{n \geqslant 0} f_c^{-n}(I),$$

(4) Pour une première approche sur ce sujet, on peut consulter l'article de M. Lyubich *The Quadratic Family as a Qualitatively Solvable Model of Chaos* Notices of the AMS 47 (2000).

qui est donc exactement l'ensemble des points d'orbite bornée, en particulier l'ensemble non-errant est inclus dans K. Pour décrire la structure de K remarquons que $f^{-1}I \cap I = I_0 \cup I_1$, où les I_j sont des intervalles symétriques par rapport à l'origine (pour fixer les notations on peut décider que $I_0 \subset [-2, 0]$ et $I_1 \subset [0, 2]$). En outre, pour $j = 0, 1$, f réalise un homéomorphisme $I_j \to I$, qui sera noté f_j. Ainsi $f^{-1}I_0 = f_0^{-1}(I_0) \cup f_1^{-1}(I_0) =: I_{00} \cup I_{01}$ et de même pour $f^{-1}(I_1)$. Plus généralement, si $(\varepsilon_n) \in \{0,1\}^{\mathbb{N}}$ on pose par récurrence

$$I_{\varepsilon_0\cdots\varepsilon_n} = f_{\varepsilon_n}^{-1}(I_{\varepsilon_0\cdots\varepsilon_{n-1}}) = f_{\varepsilon_n}^{-1}\cdots f_{\varepsilon_0}^{-1}(I),$$

et $K_\varepsilon = \bigcap_{n\geqslant 0} I_{\varepsilon_0\cdots\varepsilon_n}$. Inversement, à tout $x \in K$ on peut associer une suite $\iota(x) = (\iota_n(x)) \in \{0,1\}^{\mathbb{N}}$ qui est l'**itinéraire** de l'orbite de x dans la partition $K = (K \cap I_0) \cup (K \cap I_1)$. Plus précisément $\iota_n(x)$ est l'unique $j \in \{0, 1\}$ tel que $f^j(x) \in I_j$. Ceci définit une application

$$\iota : K \to \{0,1\}^{\mathbb{N}},$$

dont on vérifie simplement continue pour la topologie produit, et qui semi-conjugue le décalage σ à f.

Pour simplifier la discussion , supposons que $c < -9/4$. On vérifie que sous cette hypothèse on a que pour $j = 0, 1$, $\left|f_j'\right| \geqslant \alpha > 1$ sur I_j. Dans ce cas $\operatorname{diam}(I_{\varepsilon_0\cdots\varepsilon_n}) \leqslant \alpha^{-n}$, et il s'ensuit que K_ε est réduit à un point.

Remarque. — Le cas général est un peu plus délicat. On peut montrer que pour tout $c < -2$ il existe une fonction ρ définie au voisinage de K telle que $\rho(f(x))\,|f'(x)|\,/\rho(x) > 1$. En pratique on peut prendre $\rho(x) = (1 - x^2/c^2)^{-1/2}$. Si maintenant on pense à $\rho(x)\,|dx|$ comme une métrique riemannienne sur l'intervalle, on en déduit qu'il existe une distance au voisinage de K équivalente à la distance usuelle et qui est uniformément dilatée par f. Ainsi $\operatorname{diam}(I_{\varepsilon_0\cdots\varepsilon_n})$ décroît exponentiellement vers 0.

Modulo cette difficulté, nous avons ainsi démontré:

Théorème. — Pour tout $c < -2$, $f_c|_{K(f_c)}$ est conjuguée à un décalage sur 2 symboles.

On peut ainsi décrire très précisément la dynamique topologique de f_c:

- l'ensemble non-errant de f_c est égal à $K(f_c)$;
- $f_c|_{K(f_c)}$ est topologiquement transitive et même exacte;
- les points périodiques sont denses dans $K(f_c)$, et pour tout n, f_c admet 2^n orbites périodiques de période n.

On voit également que, même si $f_c|_{K(f_c)}$ est chaotique, pour $c \neq c' < -2$, $f_{c'}|_{K(f_{c'})}$ est topologiquement conjuguée à $f_c|_{K(f_c)}$. On dit que $f_c|_{K(f_c)}$ est **structurellement stable**. Cette propriété fondamentale (et quelque peu contre-intuitive) est liée de manière essentielle à l'estimation $|f_c'|\,|_{K(f_c)} > 1$ (propriété dite d'**hyperbolicité**). L'opération consistant à introduire une application "itinéraire" pour se ramener à une dynamique symbolique s'appelle un **codage** de la dynamique.

Exercice. — Montrer que pour $c \neq c'$ la conjugaison topologique ainsi construite entre $f_c|_{K(f_c)}$ et $f_{c'}|_{K(f_{c'})}$ est Hölder mais jamais dérivable (i.e. n'admet pas d'extension dérivable à un voisinage de K).

Exercice. — Montrer que si $|c|$ est assez grand, $K(f_c)$ est de mesure de Lebesgue nulle. Cette propriété est en fait vraie pour tout $c < -2$.

3.5.6. — Retour sur le cas $c = -2$.

On a vu que $([-2,2], f_{-2})$ est un facteur de $(\mathbb{R}/\mathbb{Z}, M_2)$, et également qu'il existe une semi-conjugaison $(\{0,1\}^{\mathbb{N}}, \sigma) \to (\mathbb{R}/\mathbb{Z}, M_2)$, et donc une semi-conjugaison $(\{0,1\}^{\mathbb{N}}, \sigma) \to ([-2,2], f_{-2})$. On peut également interpréter cette semi-conjugaison en termes d'itinéraire. En effet posons $I_0 = [-2,0]$ et $I_1 = [0,2]$. Pour tout $x \in [0,1]$ on pose $\iota(x) = (\varepsilon_n)$, où $f^n(x) \in I_{\varepsilon_n}$. On voit que cette définition est bien posée si $x \notin \bigcup_{n\geqslant 0} f^{-n}(\{0\})$, et présente une ambigüité sinon. Cette ambigüité est de même nature que celle donnée par le développement dyadique, et en développant les définitions on voit que ι inverse bien la semiconjugaison $(\{0,1\}^{\mathbb{N}}, \sigma) \to ([-2,2], f_{-2})$ (détails en exercice).

◇

II. MESURES INVARIANTES ET NOTION D'ERGODICITÉ

1. Définitions et premières propriétés

1.1. *Définition.* — Soit $(X, \mathcal{A}, \mu)$ un espace probabilisé[(1)] et $f : (X, \mathcal{A}) \to (X, \mathcal{A})$ mesurable (i.e. $f^{-1}(\mathcal{A}) \subset \mathcal{A}$). La mesure μ est f-invariante si $f_\star\mu = \mu$, c'est à dire pour tout $A \in \mathcal{A}$, $\mu(f^{-1}(A)) = \mu(A)$. On dira alors que $(X, \mathcal{A}, \mu, f)$ est un **système dynamique mesuré**.

1.2. *Proposition. — La mesure μ est f-invariante si et seulement si pour toute fonction $\varphi \in L^1(X, \mathcal{A}, \mu)$ on a $\int \varphi \, \mathrm{d}\mu = \int \varphi \circ f \, \mathrm{d}\mu$.*

Démonstration. — C'est très classique. L'égalité est claire pour une fonction indicatrice par définition de l'invariance et la relation $\mathbf{1}_A \circ f = \mathbf{1}_{f^{-1}A}$, puis on l'étend par linéarité aux fonctions étagées, puis aux fonctions positives par le théorème de convergence monotone, puis à toutes les fonctions L^1 en écrivant φ comme différence de sa partie positive et de sa partie négative. $\square$

1.3. — On a comme précédemment une notion d'isomorphisme entre systèmes dynamiques mesurés. $(X, \mathcal{A}, \mu, f) \sim (Y, \mathcal{B}, \nu, g)$ s'il existe une application bimesurable $\varphi : (X, \mathcal{A}) \to (Y, \mathcal{B})$ telle que $\varphi_\star\mu = \nu$ et $\varphi \circ f = g \circ \varphi$.

1.4. — Le résultat suivant est fondamental: c'est le **théorème de récurrence de Poincaré**.

Théorème. — Soit f une application préservant la mesure sur un espace probabilisé $(X, \mathcal{A}, \mu)$. Si $A \in \mathcal{A}$ est tel que $\mu(A) > 0$, alors pour presque tout $x \in A$ il existe une infinité de n tels que $f^n(x) \in A$.

Démonstration. — Posons

$$B = \left\{x \in A, \ \forall k \geqslant 1 \ f^k(x) \notin A\right\} = A \setminus \bigcup_{k \geqslant 1} f^{-k}(A) :$$

ce sont les points de A qui ne reviennent jamais dans A. Évidemment B est mesurable et on observe que pour tous $m \neq n$ on a $f^{-n}(B) \cap f^{-m}(B) = \varnothing$. En effet si par exemple $n > m$ on a $f^{-n}(B) \cap f^{-m}(B) = f^{-m}(B \cap f^{-(n-m)}(B))$. Maintenant si $x \in B$ on a nécessairement $f^{n-m}(x) \notin B$ car sinon comme $B \subset A$ on aurait $f^{n-m}(x) \in A$, une contradiction. Donc les ensembles $f^{-n}(B)$, $n \in \mathbb{N}$ sont disjoints et comme μ est f-invariante ils sont tous de même mesure. On conclut que $\mu(B) = 0$, c'est à dire que presque tout point de A revient au moins une fois dans A.

Posons maintenant $A' = A \setminus \bigcup_{n \geqslant 0} f^{-n}(B)$. Alors $\mu(A') = \mu(A)$ et si $x \in A'$ alors $f^{n(x)}(x) \notin B$, où $n(x)$ est le plus petit temps de retour de x dans A. Ainsi il existe $m \geqslant 1$

(1) Dans ce cours les mesures seront **toujours** positives et de masse finie.

tel que $f^{n(x)+m}(x) \in A$ et on construit facilement par récurrence une suite strictement croissante $(n_k(x))_{k\geqslant 1}$ telle que pour tout k, $f^{n_k(x)}(x)$ appartient à A. $\square$

1.5. — Plaçons nous maintenant dans un cadre topologique: X est un espace métrique muni de sa tribu borélienne $\mathcal{B}(X)$ et $f : X \to X$ est supposée continue

Les théorèmes usuels de densité des fonctions continues bornées dans L^1 impliquent facilement le résultat suivant.

Proposition. — Soit X un espace métrique localement compact, $f : X \to X$ continue et μ une probabilité borélienne sur X. S'il existe un sous espace dense $V \subset \mathcal{C}_b(X, \mathbb{R})$ (pour la topologie de la convergence uniforme) tel que pour toute $\varphi \in V$ on a $\int \varphi \, \mathrm{d}\mu = \int \varphi \circ f \, \mathrm{d}\mu$, alors μ est f-invariante.

1.6. *Théorème (Krylov-Bogolyubov). — Si X est compact et $f : X \to X$ est continue alors il existe une mesure de probabilité f-invariante.*

Démonstration. — L'espace $\mathcal{P}(X)$ des mesures de probabilités boréliennes sur X muni de sa topologie faible-$*$ (ou ce qui revient au même de la topologie de la convergence vague) est compact, et $f_* : \mathcal{P}(X) \to \mathcal{P}(X)$ est continue. On prend $\mu_0 \in \mathcal{P}(X)$ quelconque et on pose

$$\mu_n = \frac{1}{n} \sum_{k=0}^{n-1} f_*^k \mu_0.$$

En appliquant f_* on voit que $f_*\mu_n - \mu_n = \frac{1}{n}(f_*^n \mu_0 - \mu_0)$, et ainsi si ν est une valeur d'adhérence de (μ_n), par continuité on obtient que $f_*\nu = \nu$. $\square$

1.7. *Théorème. — Soit X un espace métrique séparable, μ une probabilité borélienne sur X et $f : X \to X$ mesurable et préservant μ. Alors μ-presque tout point est récurrent.*

Démonstration. — Soit $(V_n)_{n\in\mathbb{N}}$ une base de voisinages de X. Pour $n \in \mathbb{N}$ on pose $F_n \subset V_n$ l'ensemble des points de V_n ne revenant pas dans V_n. Si $\mu(V_n) > 0$, d'après le théorème de récurrence de Poincaré on a $\mu(F_n) = 0$. Ceci reste vrai tautologiquement si $\mu(V_n) = 0$. Posons $F = \bigcup_n F_n$. Alors $\mu(F) = 0$ et si $x \notin F$, x est récurrent: en effet si $(V_{n_k})_{k\geqslant 0}$ est une base de voisinages de x extraite de la famille (V_n), alors pour tout k, x revient dans V_{n_k}, ce qui est précisément la définition de la récurrence. $\square$

2. Exemples et constructions classiques (bis)

On reprend les notations du §3 du chapitre précédent.

2.1. Rotations. — Posons pour A un borélien de $\mathbb{T} = \mathbb{R}/\mathbb{Z}$, $\mathrm{Leb}_{\mathbb{T}}(A) = \mathrm{Leb}_{\mathbb{R}}(\pi^{-1}(A) \cap [0, 1[)$, où $\mathrm{Leb}_{\mathbb{R}}$ est la mesure de Lebesgue sur $\mathbb{R}$ et $\pi : \mathbb{R} \to \mathbb{R}/\mathbb{Z}$ est la projection naturelle. L'invariance de $\mathrm{Leb}_{\mathbb{R}}$ par translation implique que $\mathrm{Leb}_{\mathbb{T}}$ est invariante par les translations de $\mathbb{T}$ (exercice!). Il s'agit de la **mesure de Haar** du groupe abélien compact $\mathbb{R}/\mathbb{Z}$. Par conjugaison, on voit ainsi que les rotations préservent la mesure angulaire sur le cercle $\mathbb{S}^1$.

2.2. Multiplication sur le cercle. — La transformation $z \mapsto z^m$ sur le cercle $\mathbb{S}^1$ est conjuguée à $x \mapsto mx$ sur $\mathbb{T}$. Si $I = [a, b]$ est un intervalle de $[0, 1)$, que l'on identifie à un intervalle de $\mathbb{T}$, on voit que

$$M_m^{-1}(I) = \bigcup_{k=0}^{m-1} \left[\frac{a}{m} + \frac{k}{m}, \frac{b}{m} + \frac{k}{m}\right]$$

de sorte que $\mathrm{Leb}_{\mathbb{T}}(I) = \mathrm{Leb}_{\mathbb{T}}(M_m^{-1}(I))$. Par un argument de classe monotone, il s'ensuit que la même relation vaut pour tout borélien, i.e. que $\mathrm{Leb}_{\mathbb{T}}$ est M_m-invariante.

2.3. Systèmes dynamiques symboliques. — On ne parlera que du décalage unilatéral $\sigma : \Sigma^+ \to \Sigma^+$, où $\Sigma^+ = \mathfrak{A}^{\mathbb{N}}$. L'adaptation au cas bilatéral est laissée comme exercice au lecteur. La tribu borélienne de Σ^+ est engendrée par les cylindres

$$\Sigma_{a_0,\ldots,a_n} = \left\{x \in \Sigma^+,\ x_0 = a_0, \ldots, x_n = a_n\right\},$$

donc pour définir une mesure borélienne sur Σ^+ ou vérifier son invariance, il suffit de travailler avec les cylindres.

2.3.1. — Pour le décalage (complet) $\sigma : \Sigma^+ \to \Sigma^+$ on a

$$\sigma^{-1}\left(\Sigma_{a_0,\ldots,a_n}\right) = \bigcup_{a \in \mathfrak{A}} \Sigma_{a,a_0,\ldots,a_n}$$

donc μ sera σ-invariante si et seulement si pour tous $a_0, \ldots, a_n$ on a

$$\mu\left(\Sigma_{a_0,\ldots,a_n}\right) = \sum_{a \in \mathfrak{A}} \mu\left(\Sigma_{a,a_0,\ldots,a_n}\right).$$

Un exemple important est la **mesure équilibrée** définie par

$$\mu_{\mathrm{eq}}\left(\Sigma_{a_0,\ldots,a_n}\right) = \frac{1}{d^{n+1}}, \text{ où } d = \#\mathfrak{A},$$

mais il y a beaucoup d'autres exemples comme les mesures produit

$$\left(\sum_{a \in \mathfrak{A}} p_a \delta_a\right)^{\otimes \mathbb{N}} \quad \text{où} \quad \sum_{a \in \mathfrak{A}} p_a = 1.$$

2.3.2. — Soit $\sigma : \Sigma_A^+ \to \Sigma_A^+$ un sous décalage de type fini associé à une matrice $A \in M_d(\{0, 1\})$. On rappelle que A est la matrice d'adjacence d'un graphe Γ d'ensemble de sommets $\mathfrak{A}$. On se donne des **probabilités de transition** $(p_{ij})_{1 \leqslant i,j \leqslant d}$, c'est à dire que pour tous i, j on a $p_{ij} \geqslant 0$, $p_{ij} = 0$ si $a_{ij} = 0$ et pour tout i on a $\sum_j p_{i,j} = 1$. On note $P = (p_{ij})_{1 \leqslant i,j \leqslant d}$ et on remarque que le vecteur (colonne) $(1, \ldots, 1)$ est un vecteur propre associé à la valeur propre 1. On a donc également $1 \in \mathrm{Sp}({}^tP)$, on peut ainsi trouver un vecteur (ligne) v tel que $vP = v$, $v_i \geqslant 0$ et $\sum v_i = 1$. On définit alors une mesure μ sur Σ^+ en décrétant sa valeur sur les cylindres[(2)]

$$\mu\left(\Sigma_{a_0,\ldots,a_{n-1}}\right) = v_{a_0} p_{a_0 a_1} \cdots p_{a_{n-2} a_{n-1}}.$$

[(2)] Évidemment *a priori* cela ne suffit pas! Pour étendre cette mesure aux boréliens de Σ il faut un **théorème d'extension** non trivial: ce peut être le théorème d'extension de Carathéodory ou celui de Kolmogorov (voir par exemple le livre de Bogachev [**2**, §§1.5 et 3.5]).

Il s'agit bien d'une mesure de probabilité car la stochasticité de P (i.e. la relation $\sum_{k=1}^d p_{a_{n-1}k} = 1$) implique que

$$\begin{aligned}\mu\left(\Sigma_{a_0,\dots,a_{n-1}}\right) &= v_{a_0}p_{a_0a_1}\cdots p_{a_{n-2}a_{n-1}}\\ &= \sum_{a\in\mathfrak{A}} v_{a_0}p_{a_0a_1}\cdots p_{a_{n-2}a_{n-1}}p_{a_{n-1}a} = \sum_{a\in\mathfrak{A}}\mu\left(\Sigma_{a_0,\dots,a_{n-1},a}\right),\end{aligned}$$

et elle est invariante car elle satisfait les relations

$$\begin{aligned}\mu\left(\sigma^{-1}\left(\Sigma_{a_0,\dots,a_{n-1}}\right)\right) &= \sum_{k=1}^d \mu\left(\Sigma_{k,a_0,\dots,a_{n-1}}\right) = \sum_{k=1}^d v_k p_{ka_0}p_{a_0a_1}\cdots p_{a_{n-2}a_{n-1}}\\ &= v_{a_0}p_{a_0a_1}\cdots p_{a_{n-2}a_{n-1}}\ , \text{ car } \sum_{k=1}^d v_k p_{ka_0} = v_{a_0}\\ &= \mu\left(\Sigma_{a_0,\dots,a_{n-1}}\right)\end{aligned}$$

(ici encore il faut appliquer un argument de classe monotone).

2.4. Transformation du chat d'Arnol'd. — Comme précédemment le tore $\mathbb{T}^2$ admet une mesure de Haar, notée $\mathrm{Leb}_{\mathbb{T}^2}$, que l'on peut simplement définir par $\mathrm{Leb}_{\mathbb{T}^2}(A) = \mathrm{Leb}_{\mathbb{R}^2}(\pi^{-1}(A)\cap[0,1]^2)$. Par ailleurs la matrice $M = \left(\begin{smallmatrix}2&1\\1&1\end{smallmatrix}\right)$ est de déterminant 1, donc $M_\star\mathrm{Leb}_{\mathbb{R}^2} = \mathrm{Leb}_{\mathbb{R}^2}$. Rappelons que M induit une transformation f du tore et on vérifie simplement que si $A\subset\mathbb{T}^2$ est un borélien on a $\mathrm{Leb}_{\mathbb{T}^2}(f^{-1}(A)) = \mathrm{Leb}_{\mathbb{T}^2}(A)$. En effet si A admet un relevé $\widetilde{A}$ dans $[0,1]^2$ tel que $M^{-1}(\widetilde{A})\subset[0,1]^2$ c'est clair, et sinon on procède par découpage et translations.

2.5. Transformation induite. — Soient $(X,\mathcal{A},\mu,f)$ un système dynamique mesuré et $A\in\mathcal{A}$ tel que $\mu(A)>0$. Pour $x\in A$ posons $n_A(x) = \min\{n\geqslant 1,\ f^n(x)\in A\}$ Par récurrence de Poincaré, on a $n_A(x)<\infty$ p.s. On définit alors

$$\begin{aligned} f^{n_A} : A &\longrightarrow A\\ x &\longmapsto f^{n_A}(x)\end{aligned}$$

c'est *l'application induite* par f sur A.

Exercice. — Montrer que si f est inversible, f^{n_A} préserve $\mu_A = \frac{1}{\mu(A)}\mu|_A$.

2.6. Tours de Rokhlin. — C'est en quelque sorte l'opération inverse de la construction précédente. On se donne une transformation inversible $f : X\to X$ préservant une mesure μ et une fonction $n : X\to\mathbb{N}^*$ intégrable (dite fonction "toît"). On définit alors l'espace $\widetilde{X}_n$ comme

$$\widetilde{X}_n = {\{(x,k)\in X\times\mathbb{N},\ 0\leqslant k\leqslant n(x)\}}\big/{\sim}$$

où $\sim$ identifie $(x,n(x))$ et $(f(x),0)$. On peut définir une mesure $\widetilde{\mu}$ sur $\widetilde{X}_n$ qui est "localement le produit de μ par n": plus précisément $\widetilde{X}_n|_{n(\cdot)=k}$ est bimesurablement isomorphe à $\{n(\cdot)=k\}^k$, et on définit $\widetilde{\mu}$ sur $\widetilde{X}_n|_{n(\cdot)=k}$ comme image de $\left(\mu|_{\{n(\cdot)=k\}}\right)^k$ via cet isomorphisme. Comme n est intégrable, μ est finie. Finalement, on définit une transformation

$\widetilde{f}$ sur $\widetilde{X}$ par $f(x,k) = (f, k+1)$ pour tout $0 \leqslant k \leqslant n(x) - 1$, et on vérifie simplement que $\widetilde{f}$ préserve $\widetilde{\mu}$.

Exercice. — Montrer que si $(X, \mathcal{A}, \mu, f)$ est un système inversible et ergodique, et $A \in \mathcal{A}$ est de mesure strictement positive, alors $(X, \mathcal{A}, \mu, f)$ est mesurablement conjugué à la tour de Rokhlin définie par (A, n_A).

2.7. Extension naturelle. — Il est souvent très utile lorsqu'on travaille avec un système dynamique non-inversible de se ramener au cas inversible par extension. La construction de l'extension naturelle présentée à l'exercice 1.9 du chapitre précédent dans un cadre topologique peut se faire dans un cadre mesurable (et sans hypothèse de surjectivité), modulo une petite hypothèse sur $(X, \mathcal{A}, \mu)$, qui est en pratique essentiellement toujours satisfaite. On demande en effet que $(X, \mathcal{A}, \mu)$ soit un **espace probabilisé standard** (encore appelé espace de Lebesgue), c'est à dire qu'il est mesurablement isomorphe à la réunion disjointe d'un intervalle muni de sa mesure de Lebesgue et d'une quantité au plus dénombrable d'atomes.

Théorème. — *Soit* $(X, \mathcal{A}, \mu)$ *un un espace probabilisé standard et* f *une transformation sur* X *préservant la mesure. Alors il existe une extension inversible* $(\widehat{X}, \widehat{\mathcal{A}}, \widehat{\mu}, \widehat{f})$, *où* $(\widehat{X}, \widehat{\mathcal{A}}, \widehat{\mu})$ *est de Lebesgue) vérifiant la propriété universelle suivante: toute extension inversible* $(Y, \mathcal{B}, \nu, g)$ *se factorise à travers* $\widehat{X}$.

La construction est similaire au cas topologique, mais pose des problèmes de mesurabilité assez délicats: voir le livre de Przytycki et Urbanski [**12**, §2.7] pour les détails.

3. Ergodicité

3.1. *Définition.* — Soit $(X, \mathcal{A}, \mu, f)$ un système dynamique mesuré. On dit qu'il est ergodique si tout tout ensemble A tel que $\mu(f^{-1}(A)\Delta A) = 0$ (A est “essentiellement f^{-1}-invariant”) est de mesure 0 ou 1.

3.2. *Exercice.* — Montrer que l'ergodicité de $(X, \mathcal{A}, \mu, f)$ est équivalente aux deux conditions suivantes:

(i) Tout ensemble $A \in \mathcal{A}$ tel que $f^{-1}(A) = A$ est de mesure 0 ou 1.
(ii) Tout ensemble $A \in \mathcal{A}$ tel que $f(A) \subset A$ est de mesure 0 ou 1.

3.3. *Proposition.* — *Un système dynamique mesuré* $(X, \mathcal{A}, \mu, f)$ *est ergodique si et seulement si pour toute fonction mesurable (resp. pour toute fonction mesurable bornée)* $\varphi : X \to \mathbb{R}$ *on a*

$$\varphi \circ f = \varphi \ \textit{p.p.} \implies \varphi \textit{ est constante p.p.}$$

Démonstration. — Si f est ergodique et φ est une fonction mesurable telle que $\varphi \circ f = \varphi$, alors pour tout $M \in \mathbb{R}$, l'ensemble $\{\varphi < M\}$ est essentiellement invariant, donc de mesure

0 ou 1. Comme φ est finie p.p. $\mu(\{\varphi < M\})$ est non-nul et vaut donc 1 pour M assez grand. Posons $c = \inf\{M,\ \mu(\{\varphi < M\}) = 1\}$. Alors on vérifie simplement que $\varphi = c$ p.p.

Réciproquement, supposons que toute fonction essentiellement invariante est égale p.p. à une constante. Alors si A est essentiellement invariant 1_A est essentiellement constante et on conclut. □

3.4. *Proposition. — Soient μ_1 et μ_2 deux mesures de probabilité ergodiques pour un système dynamique mesurable $(X, \mathcal{A}, f)$. Montrer qu'on a soit $\mu_1 \perp \mu_2$ soit $\mu_1 = \mu_2$.*

Démonstration. — Exercice. □

3.5. *Théorème. — Soit X un espace métrique compact et $f : X \to X$ une application continue. Alors f admet une mesure de probabilité ergodique.*

Démonstration. — On sait déjà que f admet des probabilités invariantes. Soit $\mathcal{M}$ l'ensemble des probas f-invariantes, que l'on peut voir comme un sous-ensemble de $\mathcal{C}(X, \mathbb{R})^*$. C'est un convexe fermé borné (pour sa topologie d'espace normé), donc il est compact pour la topologie faible-$\star$. Rappelons que $\mu \in \mathcal{M}$ est dit extrémal si

$$\forall \mu_1, \mu_2 \in \mathcal{M},\ \forall \alpha \in]0,1[,\ (\mu = \alpha\mu_1 + (1-\alpha)\mu_2 \Rightarrow \mu_1 = \mu_2 = \mu)\,.$$

Admettons le lemme suivant pour le moment.

3.6. *Lemme. — $\mathcal{M}$ admet un point extrémal.*

Noter que ce lemme peut également être vu comme conséquence d'un résultat plus général: tout compact convexe d'un espace localement convexe séparé est l'enveloppe convexe-fermée de l'ensemble de ses points extrémaux (théorème de Krein-Milman). Le théorème découle alors directement du lemme suivant (noter que la compacité n'est pas nécessaire pour la démonstration). □

3.7. *Lemme. — Si μ est un point extrémal de $\mathcal{M}$ alors μ est ergodique*

Démonstration. — Soit μ un point extrémal de $\mathcal{M}$ et supposons qu'il existe $B \in \mathcal{A}$ essentiellement invariant tel que $0 < \mu(B) < 1$: il en est alors de même pour $X \backslash B$. On en déduit alors que la probabilité $\nu := \frac{1}{\mu(B)}\mu|_B$ est f-invariante. En effet

$$\begin{aligned}\nu(f^{-1}(A)) &= \frac{1}{\mu(B)}\mu(f^{-1}(A) \cap B) = \frac{1}{\mu(B)}\mu(f^{-1}(A) \cap f^{-1}(B)) \\ &= \frac{1}{\mu(B)}\mu(f^{-1}(A \cap B)) = \frac{1}{\mu(B)}\mu(A \cap B) = \nu(A).\end{aligned}$$

Ainsi

$$\mu = \mu|_B + \mu|_{X\backslash B} = \mu(B)\left(\frac{1}{\mu(B)}\mu|_B\right) + \mu(X\backslash B)\left(\frac{1}{\mu(X\backslash B)}\mu|_{X\backslash B}\right)$$

est une combinaison convexe non-triviale de μ dans $\mathcal{M}$: contradiction. □

Démonstration du Lemme 3.6. — Soit $(\varphi_k)_{k \geqslant 0}$ une suite dense dans $\mathcal{C}(X, \mathbb{R})$. Posons

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

Par compacité faible de $\mathcal{M}$, $\mathcal{M}_0$ est non vide et fermé (faiblement ou fortement). Par récurrence définissons pour $k \geqslant 1$

$$\mathcal{M}_k = \left\{\nu \in \mathcal{M}_{k-1},\ \int \varphi_k \mathrm{d}\nu = \sup_{\mu \in \mathcal{M}_{k-1}} \int \varphi_k \mathrm{d}\nu\right\}.$$

On obtient ainsi une suite décroissante de fermés non vides de $\mathcal{M}$. Par compacité, l'intersection $\mathcal{M}_\infty := \bigcap_{k\geqslant 0} \mathcal{M}_k$ est non vide.

Considérons $\mu \in \mathcal{M}_\infty$ et montrons que μ est un point extrémal de $\mathcal{M}$. En effet si $\mu = \alpha\mu_1 + (1-\alpha)\mu_2$, alors

$$\int \varphi_0 \mathrm{d}\mu = \alpha\varphi_0 \mathrm{d}\mu_1 + (1-\alpha)\varphi_0 \mathrm{d}\mu_2,$$

mais par définition de de φ_0 on a

$$\int \varphi_0 \mathrm{d}\mu_1 \leqslant \int \varphi_0 \mathrm{d}\mu \text{ et } \int \varphi_0 \mathrm{d}\mu_2 \leqslant \int \varphi_0 \mathrm{d}\mu.$$

Ainsi

$$\int \varphi_0 \mathrm{d}\mu_1 = \int \varphi_0 \mathrm{d}\mu_2 = \int \varphi_0 \mathrm{d}\mu,$$

et en particulier μ_1 et μ_2 appartiennent à $\mathcal{M}_1$. Ainsi de suite on voit que μ_1 et μ_2 sont dans $\mathcal{M}_k$ et $\int \varphi_k \mathrm{d}\mu_1 = \int \varphi_k \mathrm{d}\mu_2 = \int \varphi_k \mathrm{d}\mu$. Finalement comme la suite (φ_k) est dense on conclut que $\mu_1 = \mu_2 = \mu$. □

3.8. *Exercice.* — Montrer que la réciproque est vraie: si μ est ergodique alors μ est un point extrémal de l'ensemble des mesures invariantes.

3.9. *Remarque.* — Le théorème de décomposition de Choquet dit que pour toute $\mu \in \mathcal{M}$ il existe une mesure de probabilité ρ_μ sur $\mathcal{M}$, donnant masse totale aux points extrémaux, telle que $\mu = \int \nu \, d\rho_\mu(\nu)$. Ainsi toute mesure invariante est une moyenne de mesures ergodiques. Sur ce sujet, voir le livre de Phelps [**11**] (en particulier le chapitre 12). Nous expliquerons un résultat plus précis au § IV.4.

3.10. — Dans le cas ergodique on a un renforcement du théorème de récurrence de Poincaré.

Proposition. — Soit $(X, \mathcal{A}, \mu, f)$ un système dynamique ergodique. Si $A \in \mathcal{A}$ est tel que $\mu(A) > 0$, alors pour presque tout x il existe une infinité de n tels que $f^n(x) \in A$.

Démonstration. — Par récurrence de Poincaré il suffit de démontrer l'existence d'un entier n tel que $f^n(x) \in A$. Pour cela, posons $B = \bigcup_{n\geqslant 0} f^{-n}(A)$. Alors $f^{-1}(B) \subset B$ donc $f^{-1}(B) = B$ mod. 0 (exercice) et donc $\mu(B)$ vaut 0 ou 1. Comme $B \supset A$, on a $\mu(B) = 1$ et on conclut. □

3.11. *Corollaire. — Si X est un espace métrique séparable et μ une mesure borélienne ergodique. Alors pour μ presque tout x, l'orbite de x est dense dans* $\mathrm{Supp}(\mu)$.

Exercice. — Combiner ce corollaire à la Remarque 3.9 pour donner une nouvelle preuve du Théorème 1.7.

Démonstration. — En effet pour tout ouvert U tel que $U \cap \mathrm{Supp}(\mu) \neq \varnothing$ on a $\mu(U) > 0$, donc presque tout x visite U, i.e. $\mu\left(\bigcup_{n\geqslant 0} f^{-n}(U) = 1\right) = 1$. Si $(U_k)_{k\geqslant 0}$ est une base de voisinages de $\mathrm{Supp}(\mu)$ on a alors $\mu\left(\bigcap_{k\geqslant 0}\bigcup_{n\geqslant 0} f^{-n}(U_k) = 1\right) = 1$ ce qui est le résultat souhaité. □

3.12. — On peut également estimer le temps de retour moyen de A, c'est le **Théorème de Kač**.

Théorème. — Soit $(X, \mathcal{A}, \mu, f)$ un système dynamique ergodique, et $A \in \mathcal{A}$ tel que $\mu(A) > 0$. Pour $x \in A$ soit $n_A(x)$ le temps de premier retour dans A: $n_A(x) = \inf\{n \geqslant 1,\ f^n(x) \in A\}$. Alors

$$\mathbb{E}(n_A) = \frac{1}{\mu(A)},\ \textit{i.e.}\ \int_A n_A(x)\frac{\mathrm{d}\mu(x)}{\mu(A)} = \frac{1}{\mu(A)}.$$

Démonstration. — On doit montrer que $\int_A n_A(x)\mathrm{d}\mu(x) = 1$. Pour tout $n \geqslant 1$ on pose $A_n = \{x \in A,\ n_A(x) = n\}$. Par récurrence de Poincaré on a $A = \bigcup_{n\geqslant 1} A_n$ mod. 0. En outre les A_n sont disjoints donc $\mu(A) = \sum_{n=1}^{\infty} \mu(A_n)$.

La preuve du théorème **dans le cas inversible** est très intuitive et rappelle le principe des tours de Rokhlin. En effet posons pour $k < n$, $A_{n,k} = f^k(A_n)$. Alors $\mu(A_{n,k}) = \mu(A_n)$. On vérifie simplement que les ensembles $(A_{n,k})_{n\geqslant 1, 0\leqslant k<n}$ sont tous disjoints, et par ergodicité on a

$$\mu\left(\bigcup_{k\geqslant 0} f^k(A)\right) = \mu\left(\bigcup_{n\geqslant 1}\bigcup_{k=0}^{n-1} A_{n,k}\right) = 1.$$

Ainsi

$$\int_A n_A(x)\mathrm{d}x = \sum_{n=1}^{\infty} n\mu(A_n) = \sum_{n=1}^{\infty}\sum_{k=0}^{n-1} \mu(A_{n,k}) = 1$$

et le résultat est démontré.

Pour adapter l'argument au **cas général**, on pose pour $n \geqslant 1$

$$B_n = \left\{x \in X,\ f(x) \notin A, \dots f^{n-1}(x) \notin A, f^n(x) \in A\right\}$$

l'ensemble des points de X tombant dans A en exactement n itérations. Les B_n sont disjoints et par ergodicité on a $\mu\left(\bigcup_{n\geqslant 1} B_n\right) = \sum_{n=1}^{\infty} \mu(B_n) = 1$ (Proposition 3.10). Par ailleurs

$$\int_A n_A(x)\mathrm{d}\mu(x) = \sum_{n=1}^{\infty} n\mu(A_n) = \sum_{n=1}^{\infty}\sum_{k=1}^{n} \mu(A_n) = \sum_{k=1}^{\infty}\sum_{n=k}^{\infty} \mu(A_n).$$

Nous allons montrer par récurrence que $\sum_{n=k}^{\infty} \mu(A_n) = \mu(B_k)$, ce qui montrera le résultat voulu $\int_A n_A \mathrm{d}\mu = 1$. En effet pour $k = 1$, $B_1 = f^{-1}(A)$ donc $\mu(B_1) = \mu(A) = \sum_{n=1}^{\infty} \mu(A_n)$. Ensuite on a la relation $f^{-1}(B_k) = B_{k+1} \cup f^{-1}(A_k)$ car $f^{-1}(B_k)$ est constitué des points tombant en $(k+1)$ itérations dans A, et tels que $f^2(x), \dots, f^k(x)$ ne sont pas dans A (mais $f(x)$ est peut être dans A). En outre c'est une réunion disjointe donc $\mu(f^{-1}(B_k)) = \mu(B_{k+1}) + \mu(A_k)$ et la récurrence est démontrée. □

3.13. *Exercice.* — Montrer que si (X, μ, f) est inversible et ergodique et A est de mesure strictement positive, alors (X, μ, f) est mesurablement conjugué à la tour de Rokhlin définie à partir de (A, n_A).

3.14. *Remarque.* — On peut montrer que l'extension naturelle (définie au §2.7) d'un système ergodique est ergodique.

4. Mélange

4.1. *Définition.* — Soit $(X, \mathcal{A}, f, \mu)$ un système dynamique mesuré.

– On dit que μ est **mélangeante** si

$$\text{pour tous } A, B \in \mathcal{A} \text{ on a } \mu(f^{-n}(A) \cap B) \underset{n\to\infty}{\longrightarrow} \mu(A)\mu(B).$$

– On dit que μ est **faiblement mélangeante** si

$$\text{pour tous } A, B \in \mathcal{A} \text{ on a } \frac{1}{n}\sum_{k=0}^{n-1} \left|\mu(f^{-k}(A) \cap B) - \mu(A)\mu(B)\right| \underset{n\to\infty}{\longrightarrow} 0.$$

4.2. — La propriété de mélange est une propriété d'indépendance asymptotique: pour tout $B \in \mathcal{A}$, la proportion de points de B tombant dans A en n itérations i.e. $\frac{\mu(f^{-n}(A)\cap B)}{\mu(B)}$ converge vers la proportion de A dans X. La terminologie est ainsi claire: dans un shaker contenant 10% de sirop et 90 % d'eau (ou de tout autre breuvage...), après n mouvements du shaker, la proportion de sirop et d'eau dans n'importe quelle région du liquide tend vers 10%/90%.

4.3. — Le mélange implique le mélange faible d'après le théorème de Cesarò. Plus précisément, on dit qu'une suite (u_n) converge au sens de Cesarò vers ℓ si

$$\frac{1}{n}\sum_{k=0}^{n-1} |u_n - \ell| \underset{n\to\infty}{\longrightarrow} 0.$$

Donc la définition du mélange faible est que pour tous A, B, $\mu(f^{-k}(A) \cap B)$ tend vers $\mu(A)\mu(B)$ au sens de Cesarò. Intuitivement, cela signifie que pour “la plupart” des entiers k, $\left|\mu(f^{-k}(A) \cap B) - \mu(A)\mu(B)\right|$ est petit. Pour formaliser cela, rappelons qu'un sous ensemble $E \subset \mathbb{N}$ est dit de densité asymptotique 1 si

$$\lim_{n\to\infty} \#E \cap [0, n-1] = 1.$$

On a alors l'interprétation suivante de la convergence au sens de Cesarò d'une suite de réels positifs:

Lemme. — Soit (a_n) une suite bornée de réels positifs. Montrer que

$$\frac{1}{n}\sum_{k=0}^{n} a_k \underset{n\to\infty}{\longrightarrow} 0 \iff \exists E \subset \mathbb{N} \textit{ de densité 1 tel que } \lim_{\substack{n\to\infty \\ n\in E}} a_n = 0.$$

On en déduit:

Proposition. — $(X, \mathcal{A}, f, \mu)$ est faiblement mélangeant ssi pour tous $A, B \in \mathcal{A}$ il existe $E \subset \mathbb{N}$ de densité asymptotique 1 tel que

$$\lim_{\substack{n\to\infty \\ n\in E}} \mu(f^{-k}(A) \cap B) = \mu(A)\mu(B)$$

Preuve du lemme. — Le sens réciproque est assez simple: on écrit

$$0 \leqslant \frac{1}{n}\sum_{k=0}^{n} a_k = \frac{1}{n}\sum_{k\notin E} a_k + \frac{1}{n}\sum_{k\in E} a_k \leqslant \frac{M}{n}\#(E^{\complement} \cap [0,n]) + \frac{1}{n}\sum_{k\notin E} a_k$$

(où M est un majorant de (a_n)) et on traite le deuxième terme comme dans la démonstration du théorème de Cesarò.

Pour le sens direct, on pose pour tout $m \geqslant 1$, $F_m = \{k \in \mathbb{Z}, a_k \geqslant 1/m\}$. Il est clair que F_m est de densité 0 car

$$\frac{1}{n}\#(F_m \cap [0,n]) \leqslant \frac{m}{n}\sum_{k=0}^{n} a_k \underset{n\to\infty}{\longrightarrow} 0.$$

On définit alors une suite strictement croissante d'entiers $(n_m)_{m\geqslant 1}$ tels que pour tout $n \geqslant n_m$ on a $\frac{1}{n}\#(F_m \cap [0,n]) \leqslant 1/m$, et on définit $F \subset \mathbb{N}$ par

$$F \cap [n_m, n_{m+1}[= F_m \cap [n_m, n_{m+1}[$$

et $E = F^{\complement}$. Il est clair que $a_n \to 0$ si n tend vers l'infini dans E, et il reste à voir que la densité asymptotique de F est nulle. Pour cela on remarque que comme (F_m) est croissante, on a pour tout $i \leqslant m$, $F_i \cap [n_i, n_{i+1}[\subset F_m \cap [n_i, n_{i+1}[$. Ainsi on obtient que pour tout n tel que $n_m \leqslant n < n_{m+1}$, $F \cap [0,n] \subset F_m \cap [0,n]$ et donc par définition de n_m, $\#(F \cap [0,n]) \leqslant \#(F_m \cap [0,n]) \leqslant 1/m$. Nous avons donc montré que pour tout $m \geqslant 1$, il existe n_m tel que pour $n \geqslant n_m$, $\#(F \cap [0,n]) \leqslant 1/m$, et la densité asymptotique de F est donc nulle, comme annoncé. □

4.4. *Proposition. — Si un système dynamique mesuré $(X, \mathcal{A}, f, \mu)$ est (resp. faiblement) mélangeant, alors il est ergodique.*

Démonstration. — Si A est essentiellement invariant, alors pour tout n, $\mu(f^{-n}(A) \cap A) = \mu(A)$. Mais si μ mélange (resp. faiblement) alors $\mu(f^{-n}(A) \cap A)$ converge vers $\mu(A)^2$ (resp. converge vers $\mu(A)^2$ le long d'un sous-ensemble de densité 1). Dans tous les cas on conclut que $\mu(A)^2 = \mu(A)$ et donc $\mu(A)$ vaut 0 ou 1. □

La réciproque est fausse! Un exemple typique pour lequel on a ergodicité mais pas mélange est celui d'une orbite périodique: si x est un point de période n pour f, alors la mesure $\mu = \frac{1}{n}\sum_{k=0}^{n-1}\delta_{f^k(x)}$ est ergodique mais pas mélangeante ni faiblement mélangeante. De même, comme la terminologie le suggère, le mélange faible n'implique pas le mélange. Il est néanmoins plus délicat de trouver des exemples (voir [**10**, §5.7]).

4.5. — Un intérêt des notions de mélange et mélange faible est qu'elles se prêtent bien à des caractérisations "fonctionnelles" (on verra au chapitre suivant que l'on peut faire de même pour l'ergodicité). Remarquer que si $\varphi \in L^2(X, \mu, \mathbb{R})$ (ou $L^2(X, \mu, \mathbb{C})$) on a $\int |\varphi|^2 \, d\mu = \int |\varphi \circ f|^2 \, d\mu$. On introduit alors l'opérateur dit **de Koopman**:

$$\begin{aligned} U_f : \ L^2(X, \mu) &\longrightarrow L^2(X, \mu) \\ \varphi &\longmapsto \varphi \circ f \ , \end{aligned}$$

qui est une isométrie linéaire (on notera souvent $U = U_f$). Remarquer que par l'inégalité de Cauchy-Schwarz pour tous $\varphi, \psi \in L^2(X, \mu)$ on a

$$\left| \int \varphi \, U^n \psi \, d\mu \right| \leqslant \|\varphi\|_{L^2} \, \|U^n \psi\|_{L^2} = \|\varphi\|_{L^2} \, \|\psi\|_{L^2} \, . \tag{1}$$

4.6. *Proposition. — Les assertions suivantes sont équivalentes:*

(i) $(X, \mathcal{A}, f, \mu)$ *est mélangeant*

(ii) Pour tous $\varphi, \psi \in L^2(X, \mu)$ *on a*

$$\int \varphi \, U^n \psi \, d\mu = \langle \varphi, U^n \psi \rangle \underset{n \to \infty}{\longrightarrow} \left(\int \varphi \, d\mu \right) \left(\int \psi \, d\mu \right) = \langle \varphi, \mathbf{1} \rangle \cdot \langle \psi, \mathbf{1} \rangle, \tag{2}$$

(iii) Il existe un sous-ensemble dense $D \subset L^2$ *tel que pour tous* φ, ψ *dans* D*, la convergence* (2) *a lieu.*

Le même résultat vaut pour le mélange faible en remplaçant la convergence dans (2) *par la convergence au sens de Cesarò.*

Corollaire. — Si X *est un espace métrique localement compact, une mesure borélienne* μ *est mélangeante ssi*

$$\forall \varphi, \psi \in \mathcal{C}_c(X), \ \int \varphi \, (\psi \circ f^n) \, d\mu \longrightarrow \left(\int \varphi \, d\mu \right) \left(\int \psi \, d\mu \right).$$

Remarque. — Il est facile de voir que dans (2) et dans le corollaire on peut se restreindre au cas où les fonctions φ et ψ sont d'intégrale nulle, ou de manière équivalentes, orthogonales aux fonctions constantes.

Démonstration de la proposition. — L'implication *(ii)⇒(i)* est claire: il suffit d'appliquer la convergence (2) à $\varphi = \mathbf{1}_A$ et $\psi = \mathbf{1}_B$. Supposons maintenant *(i)*. Alors par définition du mélange, la convergence (2) vaut pour $\varphi = \mathbf{1}_A$ et $\psi = \mathbf{1}_B$. On l'étend alors aux fonctions étagées par linéarité sur φ puis ψ. Comme le sous espace des fonctions étagées est dense dans L^2 on conclut que *(iii)* est satisfaite. Pour *(iii)⇒(ii)* remarquons d'abord que par (1)

$$\Phi_n : (\varphi, \psi) \longmapsto \langle \varphi, U^n \psi \rangle = \int \varphi \, U^n \psi \, d\mu$$

est bilinéaire et continue. Considérons maintenant φ et ψ des fonctions arbitraires dans L^2. Par bilinéarité, comme la convergence (2) est évidente si φ est constante, il suffit de considérer le cas où $\int \varphi \, d\mu = 0$. Si (φ_k) et (ψ_k) sont des suites de fonctions de D telles que $\|\varphi_k - \varphi\| \leqslant \varepsilon_k$ et $\|\psi_k - \psi\| \leqslant \varepsilon_k$, où $\varepsilon_k \to 0$, on a

$$\begin{aligned}\Phi_n(\varphi, \psi) &= \Phi_n((\varphi - \varphi_k) + \varphi_k, (\psi - \psi_k) + \psi_k) \\ &= \Phi_n(\varphi - \varphi_k, \psi - \varphi_k) + \Phi_n(\varphi - \varphi_k, \psi) + \Phi_n(\varphi_k, \psi - \psi_k) + \Phi_n(\varphi_k, \psi_k).\end{aligned}$$

Si on fixe $\varepsilon > 0$ petit et k tel que $\varepsilon_k \leqslant \varepsilon$ on aura alors

$$\begin{aligned}|\Phi_n(\varphi, \psi)| &\leqslant |\Phi_n(\varphi_k, \psi_k)| + \|\varphi - \varphi_k\| \cdot \|\psi\| + \|\varphi_k\| \cdot \|\psi - \psi_k\| + \|\varphi - \varphi_k\| \cdot \|\psi - \psi_k\| \\ &\leqslant |\Phi_n(\varphi_k, \psi_k)| + (2\|\varphi\| + 2\|\psi\| + 1)\varepsilon,\end{aligned}$$

et comme $\Phi_n(\varphi_k, \psi_k) \to \left(\int \varphi_k \, d\mu\right)\left(\int \psi_k \, d\mu\right)$, et $\left|\int \varphi_k \, d\mu\right| = \left|\int (\varphi - \varphi_k) \, d\mu\right| \leqslant \varepsilon$, on en déduit que $\limsup_{n\to\infty} \Phi_n(\varphi, \psi) \leqslant (2\|\varphi\| + 3\|\psi\| + 2)\varepsilon$. Comme ε est arbitraire on conclut que $\lim_{n\to\infty} \Phi_n(\varphi, \psi) = 0$. Le cas faiblement mélangeant est similaire.

4.7. — L'étude des liens entre propriétés spectrales de l'opérateur de Koopman U_f et les propriétés ergodiques de (f, μ) est un sujet classique et important en théorie ergodique. Comme U est une isométrie son spectre est contenu dans le cercle unité de $\mathbb{C}$. On peut remarquer que 1 est toujours une valeur propre de U puisque si φ est constante alors $U\varphi = \varphi$. De même (f, μ) est ergodique si et seulement si l'espace propre associé à la valeur propre 1 est de dimension 1 (i.e. réduit aux constantes). Pour la culture on pourra noter le résultat suivant: (f, μ) *est faiblement mélangeante si et seulement si 1 est la seule valeur propre de U et l'espace propre associé est de dimension 1* (voir par exemple le livre de Brin et Stuck [**4**]). L'existence d'une telle caractérisation spectrale est un des intérêts principaux de la notion de mélange faible, comparativement au mélange dont la définition est plus naturelle mais qui ne se laisse pas caractériser de cette façon.

4.8. *Exercice.* — Soit $k \in \mathbb{N}^*$. Montrer que f est mélangeante (resp. faiblement mélangeante) si et seulement si f^k est mélangeante (resp. faiblement mélangeante).

5. Exemples (ter)

5.1. Rotations. —

5.1.1. *Théorème. — Si α est irrationnel, la rotation $R_\alpha : z \mapsto e^{2i\pi\alpha}z$ est ergodique pour la mesure de Lebesgue sur le cercle, et non mélangeante.*

Démonstration. — Soit A un borélien essentiellement $(R_\alpha)^{-1}$-invariant. La fonction $\mathbf{1}_A$ est dans $L^2(\mathbb{S}^1, \mathrm{Leb})$ donc elle est somme dans L^2 de sa série de Fourier, i.e. on a l'égalité $\mathbf{1}_A(x) = \sum_{n\in\mathbb{Z}} c_n e^{2i\pi nx}$ dans L^2. Toujours dans L^2 on a

$$\mathbf{1}_A \circ R_\alpha(x) = \sum_{n\in\mathbb{Z}} c_n e^{2i\pi n(x+\alpha)x} = \sum_{n\in\mathbb{Z}} c_n e^{2i\pi n\alpha} e^{2i\pi nx}.$$

Par unicité des coefficients de Fourier on en déduit que pour tout $n \in \mathbb{Z}$, $c_n = c_n e^{2i\pi n\alpha}$, et comme $\alpha \notin \mathbb{Q}$ ceci implique que pour tout $n \neq 0$, $c_n = 0$, ainsi $\mathbf{1}_A$ est égale p.p. à une constante et l'ergodicité est établie.

Pour le mélange, prenons pour B le demi cercle $[0, 1/2]$ et posons $A = [0, 1/10]$. On sait par minimalité de R_α^{-1} qu'il existe une infinité d'entiers n tels que $R_\alpha^{-n}(0) \in [3/5, 4/5]$.

Comme R_α est une isométrie on a dans ce cas $R_\alpha^{-n}(A) \cap B = \varnothing$, ce qui est antagonique au mélange. □

Remarque. — En utilisant le théorème ergodique de Birkhoff, on peut aisément adapter le raisonnement pour montrer que R_α n'est pas faiblement mélangeante.

5.2. Multiplication sur le cercle. —

5.2.1. *Théorème. — Pour $m \geqslant 2$ l'application $M_m : z \mapsto z^m$ est mélangeante pour la mesure de Lebesgue sur le cercle $\mathbb{S}^1$.*

Démonstration. — Comme les fonctions constantes par morceaux sont denses dans $L^2(\mathbb{S}^1, \mathrm{Leb})$ il suffit de montrer que si A et B sont des intervalles de $\mathbb{S}^1$ on a $\mathrm{Leb}(M_m^{-n}(A) \cap B) \to \mathrm{Leb}(A)\mathrm{Leb}(B)$. En effet, $M_m^{-n}(A)$ est la réunion de m^n intervalles de longueur $\frac{1}{m^n}\mathrm{Leb}(A)$ régulièrement répartis sur le cercle, donc la proportion de ceux ci contenus dans B converge vers $\mu(B)$ (et les effets de bord sont négligeables car ceux-ci concernent au plus deux de ces petits intervalles), d'où le mélange. □

5.2.2. *Corollaire. — Le décalage unilatéral sur m symboles est mélangeant pour la mesure équilibrée sur Σ^+.*

Démonstration. — En effet on a établi au §3.3.6 que le décalage sur m symboles est semiconjugué à $(\mathbb{S}^1, M_m)$. En outre la semi-conjugaison envoie la mesure équilibrée sur la mesure de Lebesgue et est bimesurable sur un ensemble de mesure totale, d'où le résultat. □

5.2.3. *Exercice.* — Soit f la transformation définie sur l'intervalle $[-1, 1]$ par

$$f(x) = 2x^2 - 1.$$

(1) Montrer que la mesure μ sur $[-1, 1]$ définie par

$$\mathrm{d}\mu = \frac{1}{\pi} \frac{\mathrm{d}x}{\sqrt{1 - x^2}}$$

est une mesure de probabilité f-invariante.

(2) En utilisant une formule bien connue pour $\cos(2\theta)$ montrer que f est semi-conjuguée à l'application de doublement de l'angle M_2 et retrouver le résultat précédent.

(3) Montrer que μ est mélangeante.

5.3. Le chat d'Arnol'd. — On rappelle que f désigne l'application induite par

$$M = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$$

sur le tore $\mathbb{T}^2$.

5.3.1. *Théorème. — f est mélangeante pour la mesure de Lebesgue sur $\mathbb{T}^2$.*

Démonstration. — Nous allons utiliser quelques éléments d'analyse de Fourier sur $\mathbb{T}^2$. Pour $(k, \ell) \in \mathbb{Z}^2$ on pose $e_{k,\ell}(x, y) = e^{2i\pi(kx+\ell y)}$. Alors toute fonction $\varphi = L^2(\mathbb{T}^2, \mathrm{Leb}_{\mathbb{T}^2})$ admet un unique développement en série de Fourier de la forme $\varphi = \sum_{(k,\ell)\in\mathbb{Z}^2} \varphi_{k,\ell} e_{k,\ell}$ et l'application $L^2(\mathbb{T}^2, \mathrm{Leb}_{\mathbb{T}^2}) \to \ell^2(\mathbb{Z}^2)$ définie par $\varphi \mapsto (\varphi_{k,\ell})$ est une isométrie.

Montrons d'abord l'ergodicité. Soit φ une fonction mesurable bornée telle que $\varphi\circ f = \varphi$. Alors $\varphi \in L^2(\mathbb{T}^2)$ et donc les développements de Fourier de φ et $\varphi \circ f$ coincident. Si $\varphi = \sum \varphi_{(k,\ell)} e_{k,\ell}$ on a

$$\begin{aligned}\varphi \circ f(x, y) &= \sum_{(k,\ell)\in\mathbb{Z}^2} \varphi_{(k,\ell)} e^{ik(2x+y)} e^{i\ell(x+y)} = \sum_{(k,\ell)\in\mathbb{Z}^2} \varphi_{(k,\ell)} e^{i(2k+\ell)x} e^{i(k+\ell)y)} \\ &= \sum_{(k,\ell)\in\mathbb{Z}^2} \varphi_{A^{-1}(k,\ell)} e_{k,\ell}(x, y)\end{aligned}$$

en changeant la variable de sommation. Ainsi pour tous (k, ℓ) on a $\varphi_{A^{-1}(k,\ell)} = \varphi_{(k,\ell)}$, ou ce qui revient au même $\varphi_{A(k,\ell)} = \varphi_{(k,l)}$ Mais pour tout $(k, \ell) \neq (0, 0)$ on a $A^n(k, \ell) \to \infty$. En effet pour tout $(x, y) \in \mathbb{R}^2$ tel que $(x, y) \notin \mathbb{R}e^s$ on a $A^n(x, y) \to \infty$ et comme la pente de $\mathbb{R}e^s$ est irrationnelle, cette droite n'intersecte $\mathbb{Z}^2$ qu'à l'origine. Par ailleurs comme $\varphi \in L^2$ ses coefficients de Fourier tendent vers 0 à l'infini. On conclut donc que pour tout $(k, \ell) \neq (0, 0)$, $\varphi_{(k,\ell)} = 0$, c'est à dire que φ est constante.

Une variante de cet argument va donner le mélange. Pour $\varphi, \psi \in L^2(\mathbb{T}^2, \mathbb{C})$, considérons

$$\langle U^n\varphi, \psi\rangle = \int \varphi \circ f^n\, \overline{\psi}\, \mathrm{dLeb} = \sum_{(k,\ell)\in\mathbb{Z}^2} \varphi_{A^{-n}(k,\ell)} \overline{\psi}_{(k,\ell)}.$$

Comme précédemment, pour $(k, \ell) \neq (0, 0)$ on a $\varphi_{A^{-n}(k,\ell)} \to 0$, et on en déduit sans difficulté que

$$\sum_{(k,\ell)\in\mathbb{Z}^2\setminus\{(0,0)\}} \varphi_{A^{-n}(k,\ell)} \overline{\psi}_{(k,\ell)} \underset{n\to\infty}{\longrightarrow} 0.$$

Finalement

$$\langle U^n\varphi, \psi\rangle \underset{n\to\infty}{\longrightarrow} \varphi_{(0,0)} \overline{\psi}_{(0,0)} = \langle\varphi, \mathbf{1}\rangle \cdot \langle\psi, \mathbf{1}\rangle$$

et on conclut que (f, Leb) mélange. □

Remarque. — Il existe une démonstration "géométrique" du mélange dans l'esprit de la preuve de la transitivité topologique donnée au §3.4 du chapitre précédent, connue sous le nom d'**argument de Hopf** (voir le livre de Coudène [**5**]).

5.4. Translations sur le tore. — Le résultat du paragraphe 5.1 se généralise en dimension supérieure comme suit:

5.4.1. *Théorème. — Soit $\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbb{R}^n$. La translation $\tau_\alpha : x \mapsto x + \alpha$ sur le tore $\mathbb{T}^n$ est ergodique pour la mesure de Lebesgue si et seulement si la famille $(1, \alpha_1, \ldots, \alpha_n)$ est linéairement indépendante sur $\mathbb{Q}$.*

La démonstration est laissée en exercice au lecteur. Pour le sens direct on pourra démontrer que si $(1, \alpha_1, \ldots, \alpha_n)$ est $\mathbb{Q}$-liée alors τ_α n'est pas topologiquement transitive, et pour la réciproque on s'inspirera de la section précédente. Noter que la transitivité topologique de τ_α dans le cas où $(1, \alpha_1, \ldots, \alpha_n)$ est libre n'est pas évidente et peut se déduire de l'ergodicité.

5.5. Applications linéaires du tore, cas général. — Soit A une matrice $d \times d$ à coefficients entiers et de déterminant ± 1. Alors comme en dimension 2, A induit un difféomorphisme f_A de $\mathbb{T}^d := \mathbb{R}^d/\mathbb{Z}^d$ dans lui même, préservant la mesure de Lebesgue. On peut caractériser complètement l'ergodicité de f_A.

5.5.1. *Théorème. — Sous les hypothèses précédentes, f_A est ergodique si et seulement si aucune valeur propre de A n'est une racine de l'unité.*

La démonstration est laissée en exercice au lecteur[3], qui pourra s'appuyer sur le lemme suivant:

Lemme. — Les deux assertions suivantes sont équivalentes:

(i) f_A est ergodique;
(ii) si $m \in \mathbb{Z}^d$ est tel que $e^{2i\pi\langle m, A^p x\rangle} = e^{2i\pi\langle m, x\rangle}$ Leb-presque partout alors $m = 0$ (où l'on a noté $\langle m, x\rangle = \sum_{i=1}^d m_i x_i$).

5.6. Systèmes dynamiques symboliques. — Considérons un sous décalage de type fini induit par une matrice d'adjacence A sur un graphe Γ et soit P une matrice de transition comme au §2.3.2. On a construit au §2.3.2 une mesure μ invariante par le sous-décalage (Σ_A^+, σ). On rappelle que P est **primitive** s'il existe N tel que toutes les entrées de P^N sont strictement positives. Dans ce cas A est également primitive.

5.6.1. *Proposition. — Si P est primitive alors la mesure μ est mélangeante.*

Démonstration. — (esquisse, voir [**6**] *pour les détails)* Par le théorème de Perron-Frobenius P admet un unique vecteur propre (à homothétie près) associé à la valeur propre 1 (il s'agit de $(1, \ldots, 1)$) et toutes les autres valeurs propres sont de module < 1. Il en est de même pour ${}^tP$, et on note $(v_1, \ldots, v_n)$ le vecteur propre associé. On en déduit que P^n converge vers une matrice de rang 1 de la forme

$$\begin{pmatrix} v_1 & \cdots & v_n \\ \vdots & & \vdots \\ v_1 & \cdots & v_n \end{pmatrix}.$$

Ceci signifie qu'asymptotiquement, la probabilité de passer du sommet i au sommet j en n coups converge vers v_j, et la propriété de mélange suit aisément. □

5.6.2. *Remarque.* — Si P n'est qu'irréductible alors on n'a pas forcément mélange: penser par exemple au graphe suivant:

$$\bullet \rightleftarrows \bullet\ .$$

[3] Voir [**14**, Cor. 1.10.1] pour la solution, et [**5**, §3.5] pour le mélange.

6. Complément: exactitude

Nous introduisons ici une notion plus forte d'ergodicité qui sera utile au Chapitre IX. Fixons un système dynamique mesuré $(X, \mathcal{A}, \mu, f)$.

6.1. *Proposition-Définition. — Les conditions suivantes sont équivalentes:*

(i) L'intersection des tribus $f^{-n}(\mathcal{A})$ coincide mod. 0 avec la tribu triviale, autrement dit, si $A \in \mathcal{A}$ est tel que $A \in f^{-n}(\mathcal{A})$ pour tout n alors $\mu(A)$ vaut 0 ou 1.
(ii) Si $A \in \mathcal{A}$ est tel que $\mu(A) > 0$ et pour tout n, $f^n(A) \in \mathcal{A}$, alors $\mu(f^n(A)) \underset{n\to\infty}{\longrightarrow} 1$.

Si l'une de ces conditions a lieu on dit que $(X, \mathcal{A}, \mu, f)$ est ***exact****.*

Démonstration. — Pour $(ii) \Rightarrow (i)$ on on prend $A \in \bigcap_{n\geqslant 0} f^{-n}(\mathcal{A})$, c'est à dire que pour pour tout n il existe A_n tel que $A = f^{-n}(A_n)$, et tel que $\mu(A) > 0$. Alors $A_n = f^n(A) \in \mathcal{A}$ et donc par (ii) on a $\mu(A_n) \underset{n\to\infty}{\longrightarrow} 1$. Mais par invariance $\mu(A_n) = \mu(A)$ donc $\mu(A) = 1$.

Pour l'implication réciproque, considérons $A \in \mathcal{A}$ tel que $\mu(A) > 0$ et pour tout n, $f^n(A) \in \mathcal{A}$. On remarque que pour tout n, $f^{-n}(f^n(A)) \subset f^{-(n+1)}(f^{n+1}(A))$, donc par invariance de μ, la suite $n \mapsto \mu(f^n(A))$ est croissante. Soit ℓ sa limite. Remarquons également que pour $n \geqslant n_0$, on a $f^{-n}(f^n(A)) = f^{-n_0}(f^{-(n-n_0)}(f^n(A))$, et donc $f^{-n}(f^n(A)) \in f^{-n_0}(\mathcal{A})$. Comme la suite $f^{-n}(f^n(A))$ est croissante, on en déduit que $B := \bigcup_n f^{-n}(f^n(A))$ appartient à $f^{-n_0}(\mathcal{A})$. Comme ceci est vrai pour tout n_0 on conclut que $B \in \bigcap_{n\geqslant 0} f^{-n}(\mathcal{A})$, et comme $\mu(B) > 0$, par (i) on a $\mu(B) = 1$. Mais par ailleurs on a

$$\mu(B) = \lim_{n\to\infty} \mu(f^{-n}(f^n(A))) = \lim_{n\to\infty} \mu(f^n(A)) = \ell$$

donc $\ell = 1$ et le résultat est démontré. □

Remarque. — Noter que la notion d'exactitude n'a d'intérêt que lorsque f n'est pas inversible. Il y a une notion analogue à la propriété d'exactitude pour les isomorphismes qui s'appelle la propriété K (ou de K-mélange; voir par exemple [**12**, §2.10]).

6.2. *Théorème. — Si $(X, \mathcal{A}, \mu, f)$ est exact, alors il est mélangeant.*

Avant de démontrer ce théorème, faisons quelques rappels sur la topologie faible dans un espace de Hilbert $\mathcal{H}$ (dont le produit scalaire est noté $\langle\cdot,\cdot\rangle$). On dit qu'une suite de vecteurs u_n converge faiblement vers u (noté $u_n \rightharpoonup u$) si pour tout $v \in \mathcal{H}$, on a $\langle u_n, v\rangle \underset{n\to\infty}{\longrightarrow} \langle u, v\rangle$. Toute suite bornée admet une sous suite faiblement convergente, et tout sous espace vectoriel fermé pour la topologie forte est fermé pour la topologie faible (preuve: soit V ce sous espace et $v \in V^\perp$, alors si $(u_n) \in V^{\mathbb{N}}$ est telle que $u_n \to u$ on a $0 = \langle u_n, v\rangle \to \langle u, v\rangle$ donc $u \in (V^\perp)^\perp = V$).

Démonstration. — Commençons par observer qu'une fonction $\varphi \in L^2$ est $f^{-n}(\mathcal{A})$-mesurable s'il existe $\tilde{\varphi} \in L^2(X, \mathcal{A}, \mu)$ telle que $\varphi = \tilde{\varphi} \circ f^n$, ou de manière équivalente si φ est p.s. constante sur les fibres de f^n. On en déduit que $L^2(X, f^{-n}\mathcal{A}, \mu)$ est un sous-espace vectoriel fermé de $L^2(X, \mathcal{A}, \mu)$. En effet si (φ_k) est une suite de $L^2(X, f^{-n}\mathcal{A}, \mu)$ qui converge vers φ (pour la topologie forte), on peut extraire une sous-suite convergeant p.p., et donc φ est p.s. constante sur les fibres de f^n.

Soit maintenant $\varphi \in L^2(X, \mathcal{A}, \mu)$ d'intégrale nulle. On veut montrer que pour toute $\psi \in L^2(X, \mathcal{A}, \mu)$ on a $\int (\varphi \circ f^n)\, \psi \, \mathrm{d}\mu \underset{n\to\infty}{\longrightarrow} 0$. Soit φ_∞ une valeur d'adhérence de la suite $(\varphi \circ f^n)$ pour la topologie faible. On remarque que pour tout $m \geqslant n_0$, on a $\varphi \circ f^n \in L^2(X, f^{-n_0}\mathcal{A}, \mu)$. D'après ce qui précède on en déduit que φ_∞ appartient à $L^2(X, f^{-n_0}\mathcal{A}, \mu)$ pour tout n_0, et donc par exactitude, φ_∞ est égale à une constante dans L^2. Comme $\int \varphi_\infty \mathrm{d}\mu = \langle \varphi_\infty, 1\rangle = \langle \varphi, 1\rangle = 0$ on en déduit que $\varphi_\infty = 0$ et finalement

$$\int \varphi \circ f^n \, \psi \, \mathrm{d}\mu = \langle \varphi \circ f^n, \psi \rangle \underset{n\to\infty}{\longrightarrow} \langle \varphi_\infty, \psi\rangle = 0,$$

ce qu'il fallait démontrer. □

◇

III. THÉORÈMES ERGODIQUES

Le concept d'ergodicité a été historiquement introduit par Boltzmann dans le cadre de la théorie cinétique des gaz. L'idée est que la moyenne temporelle de certaines quantités physiques pour l'évolution une particule "typique" est égale à la moyenne spatiale de cette quantité pour le système. C'est exactement cette idée qu'exprime le **théorème ergodique de Birkhoff**:

Théorème (Birkhoff, 1931). — Soit $(X, \mathcal{A}, \mu, f)$ un système dynamique ergodique ($\mu(X) = 1$). Alors pour toute fonction $\varphi \in L^1(X, \mu)$ on a presque sûrement et dans L^1

$$\frac{1}{n}\sum_{k=0}^{n-1} \varphi(f^k(x)) \underset{n\to\infty}{\longrightarrow} \int \varphi \, \mathrm{d}\mu.$$

La démonstration est assez délicate (voir le §2 ci-après). Nous allons commencer par le **théorème ergodique en moyenne**, établi quelques mois avant celui de Birkhoff par **Von Neumann**, qui repose sur des méthodes d'analyse fonctionnelle [(1)].

1. Le théorème ergodique en moyenne de Von Neumann

1.1. — Pour un système dynamique mesuré $(X, \mathcal{A}, f, \mu)$ on introduit la tribu $\mathcal{I}$ des sous-ensembles essentiellement invariants

$$\mathcal{I} = \big\{A \in \mathcal{A},\ f^{-1}(A) = A \text{ mod. } 0\big\}.$$

L'ensemble des fonctions $\mathcal{I}$-mesurables est l'ensemble des fonctions mesurables essentiellement invariantes, i.e. $\varphi \circ f = \varphi$ μ-p.p. On note $L^2_{\mathcal{I}}$ l'ensemble des fonctions L^2 essentiellement invariantes, qui est un sous-espace fermé, et l'espérance conditionnelle $\mathbb{E}(\varphi|\,\mathcal{I})$ est la projection orthogonale de f sur $L^2_{\mathcal{I}}$.

Si μ est ergodique $L^2_{\mathcal{I}}$ est le sous-espace des fonctions p.s. constantes et on a dans ce cas

$$\mathbb{E}(\varphi|\,\mathcal{I}) = \int \varphi \, \mathrm{d}\mu.$$

Si $\varphi \in L^2(X, \mu)$ on peut considérer son espérance conditionnelle $\mathbb{E}(\varphi|\mathcal{I})$ qui n'est autre que la projection orthogonale de φ sur $L^2_{\mathcal{I}}$. On peut étendre par densité cette définition au cas des fonctions L^1.

1.2. — On rappelle que l'on peut associer à $(X, \mathcal{A}, f, \mu)$ l'opérateur de composition $U = U_f : \varphi \mapsto \varphi \circ f$ dans L^2 qui est un opérateur linéaire borné (c'est même une isométrie $\|U\varphi\|_{L^2} = \|\varphi\|_{L^2}$), multiplicatif, i.e. $U(\varphi\psi) = (U\varphi)(U\psi)$, et qui préserve la positivité.

[(1)] Méthodes que Von Neumann a beaucoup développé, on lui doit par exemple la terminologie "espace de Hilbert".

1.3. *Théorème ergodique de Von Neumann. — Si* $(X, \mathcal{A}, f, \mu)$ *est un système dynamique mesuré, alors pour toute* $\varphi \in L^2(X, \mu)$ *on a*

$$\frac{1}{n}\sum_{k=0}^{n-1} U^k\varphi \underset{n\to\infty}{\longrightarrow} \mathbb{E}(\varphi|\mathcal{I}) \text{ dans } L^2.$$

1.4. — En dimension finie, ce théorème est très simple: U est une isométrie, donc quitte à complexifier l'espace $\mathcal{H}$, U est diagonalisable en base orthonormale et ses valeurs propres sont des nombres complexes de module 1, c'est à dire que

$$U = PDP^*, \text{ où } D = \mathrm{diag}(1, \dots, 1, \lambda_1, \dots, \lambda_q), \text{ avec } \lambda_i \neq 1.$$

Pour les $\lambda_i \neq 1$ on a $\lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} \lambda_i^k = 0$ et donc

$$\frac{1}{n}\sum_{k=0}^{n-1} U^k \underset{n\to\infty}{\longrightarrow} P\,\mathrm{diag}(1, \dots, 1, 0, \dots, 0)P^*$$

cette dernière matrice étant exactement celle de la projection sur $\mathrm{Ker}(U - \mathrm{id})$.

La démonstration originale de Von Neumann implémente cette idée dans un espace de Hilbert arbitraire, via le théorème spectral. La preuve que nous allons donner, due à F. Riesz, est plus élémentaire et montre un résultat un peu plus général.

1.5. *Théorème. — Soit* $\mathcal{H}$ *un espace de Hilbert et* $U : \mathcal{H} \to \mathcal{H}$ *une contraction faible, c'est à dire un opérateur linéaire de norme inférieure ou égale à 1 (* $\forall x \in \mathcal{H}$, $\|Ux\| \leqslant \|x\|$*), alors*

$$\frac{1}{n}\sum_{k=0}^{n-1} U^k \underset{n\to\infty}{\longrightarrow} \Pi,$$

où Π *est la projection orthogonale sur* $\mathrm{Ker}(U - \mathrm{id})$.

1.6. — Nous aurons besoin de quelques éléments de géométrie hilbertienne:

Lemme. — Soit $\mathcal{H}$ *un espace de Hilbert.*

1. *Soit* F *est un sous-espace vectoriel de* H*, dont l'orthogonal est noté* $F^\perp$*. Alors* $F^\perp$ *est fermé,* $F^\perp = \overline{F}^\perp$ *et* $\overline{F} \oplus F^\perp = \mathcal{H}$*.*
2. *Si* U *est une application linéaire continue, alors* $(\mathrm{Im}(U))^\perp = \mathrm{Ker}(U^*)$*.*
3. *Si* U *est une contraction faible, alors* U^* *également et* $\mathrm{Ker}(U - \mathrm{id}) = \mathrm{Ker}(U^* - \mathrm{id})$*.*

Démonstration. — Les deux premiers points sont standard. Pour le troisième, observons que pour tous $x, y \in \mathcal{H}$ on a

$$|\langle U^*x, y\rangle| = |\langle x, Uy\rangle| \leqslant \|x\|\,\|Uy\| \leqslant \|x\|\,\|y\|.$$

En maximisant sur y puis sur x on en déduit que $\|U^*\| \leqslant 1$. Ensuite, si $U^*x = x$ on a

$$\|x\|^2 = \langle x, x\rangle = \langle U^*x, x\rangle = \langle x, Ux\rangle \leqslant \|x\|^2,$$

ainsi on est dans le cas d'égalité de l'inégalité de Cauchy-Schwarz, et on en déduit que $Ux = x$. On a donc $\mathrm{Ker}(U^* - \mathrm{id}) \subset \mathrm{Ker}(U - \mathrm{id})$, et comme $U^{**} = U$, l'inclusion réciproque a également lieu. □

Démonstration du Théorème 1.5. — On commence par écrire

$$\begin{aligned}\mathcal{H} &= \overline{\operatorname{Im}(\operatorname{id}-U)} \oplus (\operatorname{Im}(\operatorname{id}-U))^{\perp}\\ &= \overline{\operatorname{Im}(\operatorname{id}-U)} \oplus \operatorname{Ker}(\operatorname{id}-U^{*})\\ &= \overline{\operatorname{Im}(\operatorname{id}-U)} \oplus \operatorname{Ker}(\operatorname{id}-U).\end{aligned}$$

Soit maintenant $x \in \mathcal{H}$. Si $x \in \operatorname{Ker}(\operatorname{id}-U)$ alors on a clairement

$$\frac{1}{n}\sum_{k=0}^{n-1} U^k x = x \underset{n\to\infty}{\longrightarrow} \Pi(x).$$

Ensuite si $x \in \operatorname{Im}(\operatorname{id}-U)$, en écrivant $x = y - Uy$ on a

$$\frac{1}{n}\sum_{k=0}^{n-1} U^k x = \frac{1}{n}\sum_{k=0}^{n-1} U^k(y-Uy) = \frac{1}{n}(y-U^n y) \underset{n\to\infty}{\longrightarrow} 0.$$

Pour conclure il reste à voir que le même résultat vaut pour $x \in \overline{\operatorname{Im}(\operatorname{id}-U)}$. Pour un tel x on a $x = \lim x_q$, avec $x_q \in \operatorname{Im}(\operatorname{id}-U)$. Posons $S_n = \frac{1}{n}\sum_{k=0}^{n-1} U^k$, qui est une contraction faible. On a donc

$$\|S_n x\| \leqslant \|S_n x_q\| + \|S_n(x-x_q)\| \leqslant \|S_n x_q\| + \|x-x_q\|,$$

donc si q est tel que $\|x-x_q\| \leqslant \varepsilon$ et n est assez grand, on a $\|S_n x\| \leqslant 2\varepsilon$, d'où le résultat. □

1.7. *Remarque.* — Les fonctions de $\operatorname{Im}(\operatorname{id}-U)$, i.e de la forme $\varphi - \varphi\circ f$ qui apparaissent dans la preuve s'appellent des **cobords**. L'argument ci-dessus montre que *si φ est orthogonale à $L^2_{\mathcal{I}}$ alors φ est limite dans L^2 d'une suite de cobords.*

1.8. — Le théorème ergodique en moyenne implique une caractérisation fonctionnelle de l'ergodicité.

Corollaire. — Un système dynamique mesuré $(X, \mathcal{A}, f, \mu)$ est ergodique si et seulement si

$$\forall \varphi, \psi \in L^2(X,\mu),\ \frac{1}{n}\sum_{k=0}^{n-1}\int \varphi\circ f^k\,\psi\,\mathrm{d}\mu \underset{n\to\infty}{\longrightarrow} \left(\int \varphi \mathrm{d}\mu\right)\left(\int \psi \mathrm{d}\mu\right). \tag{1}$$

Démonstration. —

$\Rightarrow$: on applique le théorème ergodique de Von Neumann:

$$\frac{1}{n}\sum_{k=0}^{n-1} \varphi\circ f^k \underset{n\to\infty}{\longrightarrow} \left(\int \varphi \mathrm{d}\mu\right)\mathbf{1} \text{ dans } L^2,$$

et on fait le produit scalaire par ψ.

$\Leftarrow$: si A est essentiellement invariant, on applique (1) à $\varphi = \psi = \mathbf{1}_A$ et on obtient que $\mu(A) = \mu(A)^2$ et donc $\mu(A) = 0$ ou 1. □

La variante suivante rend le lien avec le mélange faible particulièrement clair.

Corollaire. — Un système dynamique mesuré $(X, \mathcal{A}, f, \mu)$ *est ergodique si et seulement si*

$$\forall A, B \in \mathcal{A}, \frac{1}{n} \sum_{k=0}^{n-1} \left(\mu(f^{-k}(A) \cap B) - \mu(A)\mu(B)\right) \underset{n\to\infty}{\longrightarrow} 0.$$

1.9. *Corollaire (Théorème ergodique en moyenne* L^1 *). — Si* $(X, \mathcal{A}, f, \mu)$ *est un système dynamique mesuré, alors pour toute fonction* $\varphi \in L^1(X, \mu)$ *on a*

$$\frac{1}{n} \sum_{k=0}^{n-1} U^k \varphi \underset{n\to\infty}{\longrightarrow} \mathbb{E}(\varphi | \mathcal{I}) \textit{ dans } L^1.$$

Démonstration. — En reprenant les notations précédentes, considérons dans L^1 le sous espace $F = \mathrm{Im}(\mathrm{id} - U) + \mathrm{Ker}(\mathrm{id} - U)$. Comme précédemment dans F on a clairement

$$\frac{1}{n} \sum_{k=0}^{n-1} U^k \varphi \underset{n\to\infty}{\longrightarrow} \mathbb{E}(\varphi | \mathcal{I}).$$

En outre, comme F est dense dans L^2 pour la norme L^2 et comme L^2 est dense dans L^1, on voit que F est dense dans L^1 pour la norme L^1. En effet si $\varphi \in L^1$ on a une suite φ_n dans L^2 telle que $\varphi_n \xrightarrow{L^1} \varphi$. Ensuite pour tout n il existe une suite $(\varphi_{m,n})$ de vecteurs de F convergeant vers φ_n dans L^2, donc dans L^1 et on conclut par extraction diagonale.

Finalement comme U est une contraction faible dans L^1, $S_n = \frac{1}{n} \sum_{k=0}^{n-1} U^k$ en est également une, de même que l'espérance conditionnelle $\varphi \mapsto \mathbb{E}(\varphi | \mathcal{I})$, ainsi en écrivant $\varphi = \lim \varphi_n$ dans L^1, où $\varphi_n \in F$, on conclut que $\|S_n \varphi - \mathbb{E}(\varphi | \mathcal{I})\|_{L^1} \to 0$, ce qu'il fallait démontrer. □

2. Le théorème ergodique ponctuel de Birkhoff

En voici la version générale:

2.1. *Théorème. — Soit* $(X, \mathcal{A}, \mu, f)$ *un système dynamique mesuré, et* $\varphi \in L^1(X, \mu)$. *Alors pour presque tout* x *on a*

$$\frac{1}{n} \sum_{k=0}^{n} \varphi(f^k(x)) \underset{n\to\infty}{\longrightarrow} \mathbb{E}(\varphi | \mathcal{I})(x).$$

En particulier si μ *est ergodique on a presque sûrement*

$$\frac{1}{n} \sum_{k=0}^{n-1} \varphi(f^k(x)) \underset{n\to\infty}{\longrightarrow} \int \varphi \, \mathrm{d}\mu.$$

En fait le point est de montrer que la suite $\frac{1}{n} \sum_{k=0}^{n-1} \varphi(f^k(x))$ converge p.s. En effet alors la limite sera une fonction une fonction f-invariante $\varphi_{\mathcal{I}}$, et celle ci sera automatiquement égale à $\mathbb{E}(\varphi | \mathcal{I})$).

Le schéma de la démonstration –classique quand on veut démontrer une convergence presque sûre (cf. la loi forte des grands nombres ou le théorème de densité de Lebesgue)– consiste à combiner le résultat de convergence en moyenne avec une **inégalité maximale** (dont la preuve est souvent peu éclairante...).

2.2. *Remarque.* — Le théorème de Birkhoff est plus généralement valable pour une observable φ positive, sans hypothèse d'intégrabilité. On peut ainsi relaxer l'hypothèse d'intégrabilité sur φ dans le théorème en remplaçant $\varphi \in L^1$ par $\min(\varphi, 0) \in L^1$.

2.3. *Théorème ergodique maximal. — Soit $(X, \mathcal{A}, \mu, f)$ un système dynamique mesuré, et $\varphi \in L^1(X, \mu)$. Posons*

$$M\varphi : x \longmapsto M\varphi(x) = \sup_{n>0} \left(\frac{1}{n} \sum_{k=0}^{n-1} \varphi(f^k(x)) \right).$$

Alors

1. $\displaystyle\int_{\{M\varphi>0\}} \varphi \, \mathrm{d}\mu \geqslant 0.$
2. *Pour tout $\lambda > 0$ on a* $\displaystyle\mu\left(\{M\varphi > \lambda\}\right) \leqslant \frac{1}{\lambda} \int_{\{M\varphi>\lambda\}} \varphi \, \mathrm{d}\mu.$

Le corollaire suivant s'obtient en appliquant l'inégalité du 2. à $|\varphi|$.

2.4. *Corollaire (Inégalité maximale de Wiener). — Pour $(X, \mathcal{A}, \mu, f, \varphi)$ comme dans le théorème maximal, on a*

$$\mu\left(\left\{ \sup_{n>0} \frac{1}{n} \sum_{k=0}^{n-1} \left|\varphi \circ f^k\right| > \lambda \right\}\right) \leqslant \frac{1}{\lambda} \left\|\varphi\right\|_{L^1}.$$

Démonstration du théorème ergodique maximal. — Commençons déjà par observer que les assertions 1. et 2. sont équivalentes. En effet si 2. est satisfaite, alors on a

$$\int_{\{M\varphi>\lambda\}} \varphi \, \mathrm{d}\mu \geqslant \lambda\mu\left(\{M\varphi > \lambda\}\right),$$

qui implique 1. en faisant tendre λ vers 0. Réciproquement, on a

$$\left(\lambda\mu\left(\{M\varphi > \lambda\}\right) \leqslant \int_{\{M\varphi>\lambda\}} \varphi \, \mathrm{d}\mu\right) \Longleftrightarrow \left(\int_{\{M(\varphi-\lambda)>0\}} (\varphi - \lambda) \, \mathrm{d}\mu \geqslant 0\right)$$

donc quitte à remplacer φ par $\varphi - \lambda$ il suffit de s'intéresser à 1.

Pour démontrer 1. on introduit l'opérateur de composition $U = U_f$ et on pose

$$\widetilde{M}_N\varphi = \sup_{n\leqslant N} \sum_{k=0}^{n-1} U^k\varphi$$

et $\widetilde{M}_N^+\varphi = (\widetilde{M}_N\varphi)^+ = \max\left(0, \widetilde{M}_N\varphi\right)$. On a la relation

$$\widetilde{M}_{N+1}\varphi = \varphi + (U\widetilde{M}_N\varphi)^+.$$

En effet

$$\begin{aligned}
\widetilde{M}_{N+1}\varphi &= \max\big(\varphi, U\varphi+\varphi, \dots, U^{N+1}\varphi+\dots+\varphi\big)\\
&= \varphi + \max\big(0, U\varphi, \dots, U^{N+1}\varphi+\dots+U\varphi\big)\\
&= \varphi + \max\big(0, \max\big(U\varphi, \dots, U^{N+1}\varphi+\dots+U\varphi\big)\big)\\
&= \varphi + \max\Big(0, U\widetilde{M}_N\varphi)\Big) = \varphi + (U\widetilde{M}_N\varphi)^+
\end{aligned}$$

Comme $\widetilde{M}_N\varphi \leqslant \widetilde{M}_{N+1}\varphi$ on en déduit que

$$\widetilde{M}_N\varphi \leqslant \varphi + (U\widetilde{M}_N\varphi)^+$$

et ainsi

$$\begin{aligned}
\int_{\{\widetilde{M}_N\varphi>0\}} \varphi &\geqslant \int_{\{\widetilde{M}_N\varphi>0\}} \widetilde{M}_N\varphi - \int_{\{\widetilde{M}_N\varphi>0\}} (U\widetilde{M}_N\varphi)^+\\
&= \int_X (\widetilde{M}_N\varphi)^+ - \int_{\{\widetilde{M}_N\varphi>0\}} (U\widetilde{M}_N\varphi)^+\\
&\geqslant \int_X (\widetilde{M}_N\varphi)^+ - \int_X (U\widetilde{M}_N\varphi)^+\\
&= \int_X (\widetilde{M}_N\varphi)^+ - \int_X U(\widetilde{M}_N\varphi)^+ \text{ car } U \text{ préserve la positivité}\\
&= \left\|(\widetilde{M}_N\varphi)^+\right\|_{L^1} - \left\|U(\widetilde{M}_N\varphi)^+\right\|_{L^1}\\
&\geqslant 0 \text{ car } U \text{ est une contraction faible.}
\end{aligned}$$

Pour conclure, on remarque que la suite $\left\{\widetilde{M}_N\varphi > 0\right\}$ croît vers $\{M\varphi > 0\}$ et on obtient le résultat souhaité par convergence dominée. (Cette démonstration rapide et astucieuse est dûe à A.M. Garsia.) □

Démonstration du théorème ergodique de Birkhoff. — Le théorème est évident si φ est une fonction essentiellement invariante. On peut ainsi remplacer φ par $\varphi - \mathbb{E}(\varphi|\mathcal{I})$ et supposer que $\mathbb{E}(\varphi|\mathcal{I}) = 0$. Il faut montrer que

$$S_n\varphi(x) := \frac{1}{n}\sum_{k=0}^{n-1}\varphi(f^k(x)) \underset{n\to\infty}{\longrightarrow} 0 \text{ p.p.}$$

Remarquons d'abord que si φ est un cobord, i.e. $\varphi = \psi\circ f - \psi$ pour un certaine fonction $\psi \in L^1$, alors la conclusion est satisfaite. En effet on a dans ce cas

$$\frac{1}{n}\sum_{k=0}^{n-1}\varphi(f^k(x)) = \frac{1}{n}\left(\psi(f^{n+1}(x)) - \psi(x)\right).$$

Bien sûr $\frac{1}{n}\psi$ tend vers 0 p.s., et on a également que $\frac{1}{n}\psi\circ f^n$ tend vers 0 p.s. Pour le voir on applique le lemme de Borel-Cantelli: pour tout $\varepsilon > 0$,

$$\mu\left(\left\{\frac{1}{n}|\psi|\circ f^n > \varepsilon\right\}\right) = \mu\left(|\psi| > n\varepsilon\right).$$

Le fait suivant est classique: pour toute fonction positive intégrable h on a

$$\sum_{n\geqslant 1} \mu(\{h > n\}) \leqslant \int h\mathrm{d}\mu,$$

(faire une transformation d'Abel) d'où l'on déduit que

$$\sum_{n\geqslant 1} \mu\left(\left\{\frac{1}{n}|\psi|\circ f^n > \varepsilon\right\}\right) \leqslant \frac{1}{\varepsilon}\int |\psi|\ \mathrm{d}\mu < \infty$$

et le résultat suit.

On déduit de la Remarque 1.7 que les cobords sont denses dans le sous espace des fonctions telles que $\mathbb{E}(\varphi|\mathcal{I}) = 0$, pour la topologie L^2, donc pour la topologie L^1. Soit donc φ telle que $\mathbb{E}(\varphi|\mathcal{I}) = 0$ et (φ_k) une suite de cobords convergeant vers φ dans L^1. Alors pour tout x et pour tout k on a

$$\limsup_{n\to\infty} |S_n\varphi(x)| \leqslant \limsup_{n\to\infty} |S_n(\varphi-\varphi_k)(x)| + \limsup_{n\to\infty} |S_n\varphi_k(x)|\,.$$

Le deuxième terme est nul p.s. par ce qui précède, et pour le premier terme on utilise le fait que

$$\limsup_{n\to\infty} |S_n(\varphi-\varphi_k)(x)| \leqslant \limsup_{n\to\infty} (S_n|\varphi-\varphi_k|) \leqslant \sup_n (S_n|\varphi-\varphi_k|) = M|\varphi-\varphi_k|\,,$$

où $M\cdot$ est la fonction maximale de Wiener. Ainsi le Corollaire 2.4 implique que

$$\mu\left(\limsup_{n\to\infty} |S_n(\varphi-\varphi_k)(x)| > \lambda\right) \leqslant \frac{1}{\lambda}\|\varphi-\varphi_k\|_{L^1}\,.$$

Ainsi si $\lambda > 0$ est fixé, $\limsup_{n\to\infty}|S_n\varphi(x)| \leqslant \lambda$ sur un ensemble de mesure $1-\varepsilon_k$, avec $\varepsilon_k \to 0$, mais comme le membre de gauche ne dépend pas de k, cette propriété a lieu presque partout, ce qui termine la preuve. □

3. Le théorème ergodique sous-additif de Kingman

Nous allons ici présenter une généralisation du théorème de Birkhoff, qui jouera un rôle important dans la suite du cours. En guise de hors d'œuvre, voici un résultat classique:

3.1. *Lemme (dit "de Fekete"). — Soit* $(a_n)_{n\geqslant 1}$ *une suite sous-additive de nombres réels, i.e. pour tous* n, m, $a_{n+m} \leqslant a_n + a_m$. *Alors la suite* $\left(\frac{a_n}{n}\right)$ *converge vers* $\inf_{p\geqslant 1}\frac{a_p}{p}$.

Démonstration. — Posons $\ell = \inf_{p\geqslant 1}\frac{a_p}{p}$. Soit $\varepsilon > 0$ et p un entier tel que $\frac{a_p}{p} \leqslant \ell + \varepsilon$. Pour tout $n \in \mathbb{N}$ on peut alors faire la division euclidienne $n = kp + r$ avec $0 \leqslant r < p$. On a alors

$$\begin{aligned}
\frac{a_n}{n} = \frac{a_{kp+r}}{kp+r} \leqslant \frac{a_{kp}}{kp+r} + \frac{a_r}{kp+r} &= \frac{a_{kp}}{kp}\frac{kp}{kp+r} + \frac{a_r}{kp+r} \\
&\leqslant \frac{a_p}{p}\frac{kp}{kp+r} + \frac{a_r}{kp+r} \text{ car } \frac{a_{kp}}{kp} \leqslant \frac{a_p}{p} \text{ par sous-additivité} \\
&\longrightarrow \frac{a_p}{p} \text{ lorsque } n \text{ (i.e. } k) \text{ tend vers l'infini.}
\end{aligned}$$

Ainsi pour n assez grand $\frac{a_n}{n} \leqslant \ell+\varepsilon$ et on conclut que $\limsup \frac{a_n}{n} \leqslant \ell$ et donc $\lim \frac{a_n}{n} = \ell$. □

3.2. *Théorème ergodique sous-additif de Kingman. — Soit $(X, \mathcal{A}, \mu, f)$ un système dynamique mesuré, et (ϕ_n) une suite de fonctions mesurables telles que*

$$\forall n, m,\ \phi_{n+m} \leqslant \phi_n + \phi_m \circ f^n$$

et $\phi_1^+ \in L^1(X,\mu)$ (on parle de "cocycle sous-additif" (2)). Alors il existe une fonction $\phi \in L^1(X,\mu)$ f invariante telle que

$$\frac{\phi_n}{n} \underset{n\to\infty}{\longrightarrow} \phi \text{ p.p. et en moyenne}$$

et en outre

$$\int \phi \, \mathrm{d}\mu = \inf_n \frac{1}{n} \int \phi_n \mathrm{d}\mu = \lim_n \frac{1}{n} \int \phi_n \mathrm{d}\mu.$$

3.3. — Remarquer que le dernier point découle du lemme de Fekete, qui n'est autre que le théorème de Kingman dans le cas du système dynamique trivial où X est réduit à un point. Noter également que si $\varphi \in L^1(X,\mu)$ si on note $\overline{S}_n$ les sommes de Birkhoff non-normalisées, i.e. $\overline{S}_n = \sum_{k=0}^{n-1} \varphi \circ f^n$, alors on a la **relation de cocycle** $\overline{S}_{n+m} = \overline{S}_n + \overline{S}_m \circ f^n$. Ainsi le théorème de Kingman est bien une généralisation du théorème ergodique de Birkhoff.

3.4. — Nous allons présenter une preuve du théorème de Kingman due à Katznelson et Weiss, où pour simplifier nous supposerons que le cocycle est uniformément borné, c'est à dire qu'il existe $B > 0$ tel que pour tout $n \geqslant 1$, $\frac{1}{n}|\phi_n| \leqslant B$. Cette propriété est satisfaite en pratique dans dans de nombreux cas. Voir par exemple le livre de Viana [**13**] pour le cas général.

Démonstration. — Pour tout $n \geqslant 1$, commençons par appliquer le théorème de Birkhoff à ϕ_n. On obtient ainsi pour presque tout(3) x l'existence d'une limite

$$A_n(x) = \lim_{k\to\infty} \frac{1}{k} \sum_{j=0}^{k-1} \phi_n(f^j(x))$$

qui est f-invariante et bornée par nB. On vérifie sans peine en utilisant la propriété de sous-additivité de ϕ_n et l'invariance de A que presque sûrement

$$A_{n+m}(x) \leqslant A_n(x) + A_m(x).$$

Donc d'après le lemme de Fekete, la suite $A_n(x)/n$ converge p.p. vers une fonction mesurable Φ, invariante et bornée par B. Nous allons montrer que ϕ_n/n tend vers Φ presque sûrement, c'est à dire que $\Phi = \phi$. Posons

$$\overline{\phi} = \limsup_{n\to\infty} \frac{\phi_n}{n} \text{ et } \underline{\phi} = \liminf_{n\to\infty} \frac{\phi_n}{n}.$$

(2) Comme me l'a fait remarquer Sébastien Gouëzel, la terminologie "sous-cocycle" serait probablement meilleure.

(3) Dans toute la suite de la démonstration nous allons à de nombreuses reprises sans plus de précisions extraire un sous-ensemble de mesure totale où les diverses propriétés de convergence sont satisfaites.

Pour démontrer le théorème il suffit de démontrer que $\overline{\phi} \leqslant \underline{\phi}$ p.p.

Commençons par observer que $\overline{\phi}$ et $\underline{\phi}$ sont invariantes. En effet d'après la relation $\phi_{k+1}(x) \leqslant \phi_k(f(x)) + \phi_1(x)$ on déduit que $\overline{\phi}(x) \leqslant \overline{\phi}(f(x))$. Mais par ailleurs par l'invariance de la mesure on a $\int \overline{\phi}\, \mathrm{d}\mu = \int \overline{\phi} \circ f\, \mathrm{d}\mu$ et donc $\overline{\phi} = \overline{\phi} \circ f$ p.p. L'argument pour $\underline{\phi}$ est similaire.

Le théorème découle alors des Lemmes 3.5 et 3.6 ci-dessous. □

3.5. *Lemme. — Presque sûrement on a* $\overline{\phi}(x) \leqslant \Phi(x)$.

3.6. *Lemme. — On a* $\displaystyle\int \Phi\, \mathrm{d}\mu \leqslant \int \underline{\phi}\, \mathrm{d}\mu$.

Démonstration du Lemme 3.5. — On fixe un entier $n \geqslant 1$ et un entier $1 \leqslant i \leqslant n$. Alors pour tout $k \geqslant 1$ en faisant la division euclidienne de $k-i$ par n on aura donc $k = r+nm+i$ avec $0 \leqslant r < n$. En appliquant de manière répétée la sous additivité on voit que

$$\phi_k(x) \leqslant \phi_r\left(x_{mn+i}\right) + \sum_{\ell=0}^{m-1} \phi_n\left(x_{\ell n+i}\right) + \phi_i(x),$$

où l'on a posé $x_q = f^q(x)$. En sommant cette relation de $i = 1$ à n, en divisant par k et prenant la limite sup pour $k \to \infty$ (et donc $m \to \infty$) on obtient :

$$n\overline{\phi}(x) \leqslant 0 + A_n(x) + 0.$$

Finalement en divisant par n et en faisant tendre n vers l'infini on obtient que $\overline{\phi}(x) \leqslant \Phi(x)$, comme annoncé. □

Démonstration du Lemme 3.6. — Fixons $\varepsilon > 0$. Pour tout x il y a une infinité d'entiers n tels que $\frac{\phi_n(x)}{n} \leqslant \underline{\phi}(x) + \varepsilon$. Pour $N \geqslant 1$ fixé, on pose E_N l'ensemble des x tels qu'il existe $1 \leqslant n \leqslant N$ tel que $\frac{\phi_n(x)}{n} \leqslant \underline{\phi}(x) + \varepsilon$. On remarque de suite que $\mu(E_N)$ tend vers 1 quand $N \to \infty$. Étant donné x, on définit un entier $n(x)$ et un réel $\psi(x)$ comme suit:

- si $x \in E_N$, $n(x)$ est le plus petit entier tel que $\frac{\phi_n(x)}{n} \leqslant \underline{\phi}(x) + \varepsilon$ et on pose alors $\psi(x) = \underline{\phi}(x) + \varepsilon$;
- si $x \notin E_N$ on pose $n(x) = N$ et $\psi(x) = B + \varepsilon$ (rappelons que Bn est la borne sur ϕ_n).

On vérifie aisément que pour tout $x \in X$, $\psi(x) \leqslant \psi(f(x))$, ainsi

$$\phi_{n(x)}(x) \leqslant n(x)\psi(x) \leqslant \sum_{i=0}^{n(x)-1} \psi\left(f^i(x)\right).$$

Partant de tout point $x_0 \in X$, on définit une suite croissante infinie (i_k) en posant $i_0 = 0$ et $i_{k+1} = i_k + n(x_{i_k})$, et en posant $n_k = n(x_{i_k})$, l'inégalité précédente se réécrit

$$\phi_{i_k}(x_{i_k}) \leqslant n_k\psi(x_{i_k}) \leqslant \sum_{j=i_k}^{i_{k+1}-1} \psi(x_j).$$

En utilisant la sous-additivité on obtient alors

$$\phi_{i_k}(x_0) \leqslant \phi_{n_0}(x_{i_0}) + \phi_{n_1}(x_{i_1}) + \cdots + \phi_{n_{k-1}}(x_{i_{k-1}}) \leqslant \sum_{j=0}^{i_k-1} \psi(x_j).$$

L'entier N étant toujours fixé, pour tout entier M (qu'il faut penser comme grand), il existe k tel que $M - N < i_k \leqslant M$, et en utilisant la borne uniforme sur ϕ_n/n et celle sur ψ on obtient que

$$\phi_M(x_0) \leqslant \sum_{j=0}^{M-1} \psi(x_j) + B'$$

pour un certain $B' = B'(N)$. En se souvenant que $x_j = f^j(x)$ et en intégrant par rapport à x ceci donne

$$\frac{1}{M} \int \phi_M \, \mathrm{d}\mu \leqslant \int \psi \, \mathrm{d}\mu + \frac{B'}{M}.$$

Par l'invariance de μ on a que $\int \phi_M \, \mathrm{d}\mu = \int A_M \mathrm{d}\mu$ donc par le théorème de convergence dominée, lorsque M tend vers l'infini le membre de gauche converge vers $\int \Phi \, \mathrm{d}\mu$ et on obtient

$$\int \Phi \, \mathrm{d}\mu \leqslant \int \psi \, \mathrm{d}\mu.$$

Pour conclure, on remarque que

$$\int \psi \, \mathrm{d}\mu = \varepsilon + \int_{E_N} \underline{\phi} \, \mathrm{d}\mu + B\mu\,(X \backslash E_N) \leqslant \varepsilon + \int \underline{\phi} \, \mathrm{d}\mu + 2B\mu\,(X \backslash E_N)$$

(il faut prendre garde ici au fait que ψ dépend de N). Si dès le départ N avait été choisi de sorte que $\mu\,(X \backslash E_N) \leqslant \varepsilon / 2B$, on voit que

$$\int \Phi \, \mathrm{d}\mu \leqslant \int \psi \, \mathrm{d}\mu \leqslant \int \underline{\phi} \, \mathrm{d}\mu + 2\varepsilon.$$

Comme ε est arbitraire, le résultat est démontré. □

4. Théorèmes ergodiques non-conventionnels

4.1. — Pour un système dynamique mesuré $(X, \mathcal{A}, f, \mu)$, il est naturel de se demander s'il existe une version des théorèmes ergodiques (en moyenne ou ponctuel) pour des sous-suites, c'est à dire s'il est possible de montrer la convergence de suites de la forme

$$\frac{1}{n} \sum_{k=0}^{n} \varphi \circ f^{p(k)},$$

où $k \mapsto p(k)$ est une fonction à valeurs entières tendant vers l'infini avec k. Une variante qui joue un rôle important dans les applications de le théorie ergodique à la combinatoire est la convergence de sommes du type

$$\frac{1}{n} \sum_{k=0}^{n} (\varphi_1 \circ f_1^k)\,(\varphi_2 \circ f_2^{2k}) \cdots (\varphi_q \circ f^{qk})\ ,$$

où les φ_i sont des fonctions bornées. Ce type d'énoncé est communément désigné sous le nom de “théorème ergodique non-conventionnel” (cette terminologie est due à Furstenberg) et représente un chapitre important de la théorie ergodique moderne.

4.2. — Dans cette section nous allons présenter le plus simple des résultats de ce type:

Théorème (Furstenberg (1980)). — Soit $(X, \mathcal{A}, f, \mu)$ *un système dynamique mesuré tel que pour tout* $n \geqslant 1$, f^n *est ergodique. Soit* $p \in \mathbb{Z}[X]$ *un polynôme dont le coefficient dominant est positif. Alors pour toute fonction* $\varphi \in L^2$, *on a la convergence*

$$\frac{1}{n}\sum_{k=0}^{n-1} \varphi \circ f^{p(k)} \underset{n\to\infty}{\longrightarrow} \left(\int \varphi \,\mathrm{d}\mu\right) \mathbf{1}$$

dans L^2.

Remarquer que nous commettons un petit abus dans l'énoncé: on a bien $p(k) \geqslant 0$ pour k assez grand, mais p peut tout de même prendre un nombre fini de valeurs négatives, et alors $\varphi \circ f^{p(k)}$ n'a pas de sens. Mais comme l'énoncé est asymptotique en k nous pouvons simplement ignorer ces valeurs (par exemple les remplacer par 0), ce qui ne retire rien à l'intérêt de l'énoncé. Une autre solution serait de ne considérer que des polynômes à coefficients positifs.

La démonstration que nous allons présenter est essentiellement élémentaire et due à Bergelson. Noter que l'hypothèse d'ergodicité est plus forte que dans le théorème de Von Neumann. Elle est indispensable car si par exemple f^2 n'est pas ergodique on ne peut pas espérer avoir convergence des sommes de Birkhoff pour $p(k) = 2k$.

Il existe aussi une version ponctuelle de ce théorème (toujours pour les fonctions L^2, ou plus généralement L^p pour $p > 1$), qui est beaucoup plus difficile et a été démontrée par Bourgain à la fin des années 1980. Une autre variante consiste à prendre pour $p(k)$ la suite des nombres premiers.

4.3. — Le point clé de la démonstration du Théorème 4.2 est le lemme suivant, inspiré d'un lemme classique de Van der Corput sur les suites équidistribuées modulo 1 (voir la Remarque 5.3 du chapitre V).

Lemme. — Soit (x_n) *une suite bornée dans un espace de Hilbert* H. *On suppose que*

$$\textit{pour tout } m \geqslant 1,\ \frac{1}{N}\sum_{n=1}^{N}\langle x_{n+m}, x_n\rangle \underset{N\to\infty}{\longrightarrow} 0.$$

Alors en norme dans H *on a*

$$\frac{1}{N}\sum_{n=1}^{N} x_n \underset{N\to\infty}{\longrightarrow} 0.$$

Démonstration. — Posons $s_N = \frac{1}{N}\sum_{n=1}^{N} x_n$ et pour $i \geqslant 1$, $s_N^{(m)} = \frac{1}{N}\sum_{n=1}^{N} x_{n+m}$. On remarque que m étant fixé la suite $\|s_N - s_N^{(m)}\|$ tend vers 0 quand $N \to \infty$, donc pour tout

$M \geqslant 1$,

$$s_N - \frac{1}{M}\sum_{m=1}^{M} s_N^{(m)} \xrightarrow[N\to\infty]{} 0.$$

Il suffit donc de montrer que pour M fixé $\frac{1}{M}\sum_{m=1}^{M} s_N^{(m)}$ tend vers 0 avec N. Plus généralement il suffit en fait de montrer que

$$(2) \qquad \limsup_{M\to\infty}\limsup_{N\to\infty}\left\|\frac{1}{M}\sum_{m=1}^{M} s_N^{(m)}\right\| = \limsup_{M\to\infty}\limsup_{N\to\infty}\left\|\frac{1}{M}\sum_{m=1}^{M}\frac{1}{N}\sum_{n=1}^{N} x_{n+m}\right\| = 0.$$

En effet si cette propriété a lieu, pour tout $\varepsilon > 0$, on a pour M fixé assez grand $\limsup_{N\to\infty}\left\|\frac{1}{M}\sum_{m=1}^{M} s_N^{(m)}\right\| \leqslant \varepsilon$, et donc $\limsup_{N\to\infty}\|s_N\| \leqslant \varepsilon$. Comme ε est arbitraire, ceci signifie que $\limsup_{N\to\infty}\|s_N\| = 0$, ce qui est la propriété voulue.

Pour cela, nous allons utiliser l'inégalité

$$\left\|\frac{1}{N}\sum_{i=1}^{N} v_i\right\|^2 \leqslant \frac{1}{N}\sum_{i=1}^{N}\|v_i\|^2,$$

qui découle de l'inégalité de Cauchy-Schwarz:

$$\left\|\frac{1}{N}\sum_{i=1}^{N} v_i\right\|^2 = \frac{1}{N^2}\sum_{i,j=1}^{N}\langle v_i, v_j\rangle \leqslant \frac{1}{N^2}\sum_{i,j=1}^{N}\|v_i\|\,\|v_j\| = \frac{1}{N^2}\left(\sum_{i=1}^{N}\|v_i\|\right)^2 \leqslant \frac{1}{N}\sum_{i=1}^{N}\|v_i\|^2.$$

Pour montrer (2) on écrit

$$\begin{aligned}
\limsup_{N\to\infty}\left\|\frac{1}{M}\sum_{m=1}^{N}\frac{1}{N}\sum_{n=1}^{N} x_{n+m}\right\|^2 &= \limsup_{N\to\infty}\left\|\frac{1}{N}\sum_{n=1}^{N}\frac{1}{M}\sum_{m=1}^{M} x_{n+m}\right\|^2 \\
&\leqslant \limsup_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\left\|\frac{1}{M}\sum_{m=1}^{M} x_{n+m}\right\|^2 \\
&= \limsup_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\frac{1}{M^2}\sum_{\ell,m=1}^{M}\langle x_{n+\ell}, x_{n+m}\rangle \\
&= \limsup_{N\to\infty}\frac{1}{M^2}\sum_{\ell,m=1}^{M}\frac{1}{N}\sum_{n=1}^{N}\langle x_{n+\ell}, x_{n+m}\rangle.
\end{aligned}$$

À la dernière ligne, par hypothèse quand $\ell \neq m$ on a $\frac{1}{N}\sum_{n=1}^{N}\langle x_{n+\ell}, x_{n+m}\rangle \to 0$, ainsi l'expression à la dernière ligne est majorée par $\frac{1}{M}\sup\|x_n\|$, ce qui conclut la démonstration de (2), et donc celle du lemme. □

4·4. —

Démonstration du Théorème 4.2. — La démonstration se fait par récurrence sur le degré de p. Si p est de degré 1, on écrit $p(k) = ak + b$ avec $a \geqslant 0$, et quitte à remplacer $\sum_{k=0}^{n}$ par

$\sum_{k=k_0}^{n}$, on peut supposer que $b \geqslant 0$. Comme f^a est ergodique, on a d'après le théorème ergodique de Von Neumann

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \varphi \circ f^{ak+b} = \left(\int \varphi \circ f^b \,\mathrm{d}\mu\right) \mathbf{1} = \left(\int \varphi \,\mathrm{d}\mu\right) \mathbf{1}$$

comme souhaité.

Supposons maintenant le résultat démontré pour tous les polynômes de degré inférieur ou égal à $d-1$ et fixons p de degré d. En remplaçant φ par $\varphi - \int \varphi \,\mathrm{d}\mu$ on peut supposer qu'elle est de moyenne nulle. Si on pose $x_n = \varphi \circ f^{p(n)}$, qui définit une suite bornée dans L^2, il s'agit de démontrer que $\frac{1}{n}\sum_{k=0}^{n} x_k$ tend vers 0 en norme L^2. D'après le lemme précédent, il suffit de démontrer que pour tout $h \geqslant 1$, $\frac{1}{n}\sum_{k=0}^{n}\langle x_{k+h}, x_k\rangle$ tend vers 0 quand $n \to \infty$. On développe:

$$(3)\quad \frac{1}{n}\sum_{k=0}^{n-1}\langle x_{k+h}, x_k\rangle = \frac{1}{n}\sum_{k=0}^{n-1}\langle \varphi \circ f^{p(k+h)}, \varphi \circ f^{p(k)}\rangle = \frac{1}{n}\sum_{k=0}^{n-1}\langle \varphi \circ f^{p(k+h)-p(k)}, \varphi\rangle + O\left(\frac{1}{n}\right),$$

où dans la deuxième égalité on a utilisé l'invariance, et on néglige dans la somme le nombre fini de termes ou $p(k+h) - p(k) \leqslant 0$ (d'où l'apparition du $O\left(\frac{1}{n}\right)$). Maintenant $k \mapsto p(k+h) - p(k)$ est un polynôme de degré inférieur ou égal à $d-1$, donc d'après l'hypothèse de récurrence, $\frac{1}{n}\sum_{k=0}^{n-1} \varphi \circ f^{p(k+h)-p(k)}$ tend vers 0 dans L^2 et on conclut que l'expression (3) tend vers 0, ce qu'il fallait démontrer. □

◇

IV. APPLICATIONS

1. Applications directes du théorème ergodique

1.1. — Le théorème ergodique de Birkhoff implique directement la **loi forte des grands nombres** de **Kolmogorov**.

Théorème. — *Soit $(X_n)_{n\geqslant 0}$ une suite de variables aléatoires i.i.d. telles que $\mathbb{E}(|X_0|) < \infty$. Posons $\mu = \mathbb{E}(X_0)$. Alors presque sûrement on a*

$$\frac{1}{n}\sum_{i=0}^{n-1} X_i \underset{n\to\infty}{\longrightarrow} \mu.$$

Démonstration. — Soit $(\Omega, \mathbb{P})$ l'espace de probabilité sur lequel sont définies les X_i et soit $\mathbb{P}_X$ leur loi commune. Considérons l'espace produit $(\mathbb{R}^{\mathbb{N}}, \mathbb{P}_X^{\mathbb{N}})$ et soit $\sigma : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}^{\mathbb{N}}$ le décalage, et notons $\pi : \Omega \to \mathbb{R}^{\mathbb{N}}$ l'application naturelle: $\pi(\omega) = (X_n(\omega))_{n\in\mathbb{N}}$. Alors $\mathbb{P}_X^{\mathbb{N}}$ est ergodique sous σ. En effet si A est un borélien de $\mathbb{R}^{\mathbb{N}}$ qui est essentiellement invariant, alors quitte à le modifier sur un ensemble de mesure nulle on peut le supposer invariant. Le point est que $\pi^{-1}(A)$ est mesurable pour la tribu asymptotique $\bigcap_{n\geqslant 0} \sigma\,(X_k,\ k \geqslant N)$: en effet $\pi(\omega)$ appartient à A si et seulement si $\pi(\sigma^N(\omega))$ est dans A, et donc $(X_n(\omega))_{n\geqslant 0} \in A$ si et seulement si $(X_{n+N}(\omega))_{n\geqslant 0} \in A$ La loi du 0–1 de Kolmogorov implique alors que $\mathbb{P}(\pi^{-1}(A)) = \mathbb{P}_X^{\mathbb{N}}(A)$ vaut 0 ou 1: c'est l'ergodicité. Soit maintenant $\varphi : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}$ définie par $\varphi(x_n) = x_0$. On a alors

$$\frac{1}{n}\sum_{i=0}^{n-1} X_i(\omega) = \frac{1}{n}\sum_{i=0}^{n-1} \varphi(\sigma^k(\pi(\omega))) \underset{n\to\infty}{\longrightarrow} \int \varphi\, d(\mathbb{P}_X^{\mathbb{N}}) = \int x_0\, d\mathbb{P}_X(x_0) = \mathbb{E}(X_0),$$

ce qui termine la preuve. □

1.2. — **Fréquence des passages dans un borélien.** Une application directe et très utile du théorème de Birkhoff est la suivante: si $(X, \mathcal{A}, \mu, f)$ est un système dynamique ergodique, et $A \in \mathcal{A}$ est de mesure positive, alors pour μ-presque tout x on a

$$\frac{1}{N}\#\{1 \leqslant n \leqslant N,\ f^n(x) \in A\} \underset{N\to\infty}{\longrightarrow} \mu(A).$$

Il suffit pour cela d'appliquer le théorème ergodique III.2.1 à $\varphi = \mathbf{1}_A$.

1.3. — **Points génériques.**

Définition. — Soit (X, f) un système dynamique topologique sur un espace métrique compact et $x \in X$. On dit que x est **μ-générique** si la suite des mesures empiriques $\frac{1}{n}\sum_{k=0}^{n-1} \delta_{f^k(x)}$ converge faiblement vers μ.

1.4. *Proposition.* — *Si (X, f) est compact et μ est une mesure borélienne ergodique alors μ-presque tout point est générique.*

Démonstration. — Soit $(\varphi_j)_{j\in\mathbb{N}}$ une suite dense de $\mathcal{C}(X)$, et pour tout j, soit E_j un borélien de mesure 1 tel que pour tout $x \in E_j$ on a

$$\frac{1}{n}\sum_{k=0}^{n-1}\varphi_j(f^k(x)) \underset{n\to\infty}{\longrightarrow} \int \varphi_j \mathrm{d}\mu.$$

Soit $E = \bigcap_j E_j$, alors pour tout $x \in E$ on a que pour tout $j \in \mathbb{N}$

$$\frac{1}{n}\sum_{k=0}^{n-1}\varphi_j(f^k(x)) = \left\langle \frac{1}{n}\sum_{k=0}^{n-1}\delta_{f^k(x)}, \varphi_j \right\rangle \underset{n\to\infty}{\longrightarrow} \int \varphi_j \mathrm{d}\mu = \langle \mu, \varphi_j \rangle$$

et en par la densité de la famille (φ_j) pour la topologie de la convergence uniforme on en déduit le même résultat pour tout $\varphi \in \mathcal{C}(X)$. Le résultat suit. □

Remarque. — Le même énoncé vaut pour X localement compact et σ-compact, par exemple $X = \mathbb{R}^n$.

1.5. *Exercice.* — Construire un point $x \in \Sigma^+ = \{0,1\}^{\mathbb{N}}$ qui n'est générique pour aucune mesure invariante relativement au décalage.

2. Retour sur le mélange faible

2.1. — On a vu que si α est un nombre irrationnel, la rotation $R_\alpha : z \mapsto e^{2i\pi\alpha}z$ sur le cercle $\mathbb{S}^1$ est ergodique pour la mesure de Lebesgue. Considérons le système dynamique produit (parfois appelé **diagonal**) $R_\alpha \times R_\alpha$ sur $\mathbb{S}^1 \times \mathbb{S}^1$ défini par

$$(R_\alpha \times R_\alpha)(z,w) = (e^{2i\pi\alpha}z, e^{2i\pi\alpha}w).$$

Est il ergodique? La réponse est non: en effet R_α est une isométrie pour la distance angulaire définie par $d(z,w) = \arg(z/w)$, ainsi pour tout $r > 0$, le sous ensemble

$$\Delta_r = \{(z,w),\ d(z,w) < r\} \subset \mathbb{S}^1 \times \mathbb{S}^1$$

est invariant par $R_\alpha \times R_\alpha$. Il se trouve que cette absence d'ergodicité est due au fait que R_α n'est pas faiblement mélangeante.

2.2. *Théorème. — Soit $(X, \mathcal{A}, \mu, f)$ un système dynamique mesuré. Les conditions suivantes sont équivalentes:*

(i) *f est faiblement mélangeant;*
(ii) *Pour tout système dynamique faiblement mélangeant $(Y, \mathcal{B}, \nu, g)$, $f \times g$ est faiblement mélangeant pour $\mu \times \nu$;*
(iii) *Pour tout système dynamique ergodique $(Y, \mathcal{B}, \nu, g)$, $f \times g$ est ergodique pour $\mu \times \nu$;*
(iv) *$f \times f$ est ergodique pour $\mu \times \mu$;*
(v) *Toute fonction $\varphi \in L^2(X, \mathbb{C})$ non nulle telle que $\varphi \circ f = \lambda\varphi$ est constante, et en particulier $\lambda = 1$.*

On admettra que la condition *(v)* implique les autres: c'est la caractérisation spectrale du mélange faible qui a déjà été évoquée au §II.4.7.

Démonstration. — Il est clair que $(ii) \Rightarrow (i)$ et $(iii) \Rightarrow (iv)$. Montrons que (i) implique (ii). Pour cela, on commence par remarquer que l'espace vectoriel engendré par les fonctions de la forme $\varphi(x)\psi(y)$ avec $\varphi \in L^2(X,\mu)$ et $\psi \in L^2(Y,\nu)$ est dense dans $L^2(X \times Y, \mu \times \nu)$. Il suffit donc de montrer que pour $\varphi_1, \varphi_2 \in L^2(X)$ et $\psi_1, \psi_2 \in L^2(Y)$,

$$\int \varphi_1 \circ f^n(x)\psi_1 \circ g^n(y)\varphi_2(x)\psi_2(y)\mathrm{d}\mu(x)\mathrm{d}\mu(y) \underset{n\to\infty}{\longrightarrow} \left(\int \varphi_1\psi_1\, d(\mu \times \nu)\right)\left(\int \varphi_2\psi_2\, d(\mu \times \nu)\right)$$
$$= \left(\int_X \varphi_1\, \mathrm{d}\mu\right)\left(\int_X \varphi_2\, \mathrm{d}\mu\right)\left(\int_Y \psi_1\, d\nu\right)\left(\int_Y \psi_2\, d\nu\right)$$

au sens de Césarò. Ceci découle du théorème de Fubini et du fait suivant, qui est laissé en exercice au lecteur: si (a_n) et (b_n) sont deux suites bornées qui convergent respectivement au sens de Césarò vers a et b, alors (a_nb_n) converge au sens de Césarò vers ab (penser en termes de densité).

La preuve de l'implication $(i) \Rightarrow (iii)$ est similaire, et est laissée en exercice (attention au fait que l'ergodicité des facteurs ne suffit pas pour l'ergodicité du produit, donc il faut bien comprendre comment intervient le mélange faible).

Pour démontrer que (iv) implique (i), nous allons utiliser la caractérisation fonctionnelle de l'ergodicité établie au §III.1.8. On remarque déjà que l'ergodicité de $f \times f$ implique celle de f (exercice!). Soient maintenant φ et ψ dans $L^2(X,\mu)$, il suffit de démontrer que $\left|\int \varphi \circ f^n\psi\, \mathrm{d}\mu - \left(\int \varphi\, \mathrm{d}\mu\right)\left(\int \psi\, \mathrm{d}\mu\right)\right|^2$ tend vers 0 au sens de Cesarò, soit

$$\frac{1}{n}\sum_{k=0}^{n-1}\left(\int \varphi \circ f^k\psi\, \mathrm{d}\mu - \left(\int \varphi\, \mathrm{d}\mu\right)\left(\int \psi\, \mathrm{d}\mu\right)\right)^2 \underset{n\to\infty}{\longrightarrow} 0.$$

Après développement cette expression peut s'exprimer sous la forme

$$\frac{1}{n}\sum_{k=0}^{n-1}\int \varphi(f^k(x))\varphi(f^k(y))\psi(x)\psi(y)\, \mathrm{d}\mu(x)\mathrm{d}\mu(y) + \left(\int \varphi\, \mathrm{d}\mu\right)^2\left(\int \psi\, \mathrm{d}\mu\right)^2$$
$$-\, 2\left(\int \varphi\, \mathrm{d}\mu\right)\left(\int \psi\, \mathrm{d}\mu\right)\frac{1}{n}\sum_{k=0}^{n-1}\int \varphi \circ f^k\psi\, \mathrm{d}\mu.$$

La première somme converge vers $\left(\int \varphi\, \mathrm{d}\mu\right)^2\left(\int \psi\, \mathrm{d}\mu\right)^2$ par ergodicité de $f \times f$ et la dernière vers $\left(\int \varphi\, \mathrm{d}\mu\right)\left(\int \psi\, \mathrm{d}\mu\right)$ par ergodicité de f, donc le tout tend vers 0 et on voit que f est faiblement mélangeante.

Pour conclure, montrons que (iv) implique (v). Soit φ une fonction non égale à 0 p.p et telle que $\varphi \circ f = \lambda\varphi$ p.p. Alors $\mathbf{1}_{\{\varphi=0\}}$ est f-invariante, et comme (iv) implique que f est ergodique, elle est constante, ainsi $\mu(\{\varphi = 0\}) = 0$. Considérons ensuite la fonction $\Phi : (x,y) \mapsto \varphi(x)/\varphi(y)$: elle est mesurable, finie p.p. et vérifie $\Phi \circ (f \times f) = \Phi$, donc par ergodicité de $f \times f$ elle est constante. Ainsi pour presque tout x, y on a $\varphi(x) = c\varphi(y)$ donc $c = 1$. En effet, si E désigne l'ensemble des couples $(x,y) \in X \times X$ tels que $\varphi(x) = c\varphi(y)$, par applications successives du théorème de Fubini on peut trouver x, y, z tels que (x,y), (y,z) et (x,z) sont dans E, d'où l'on déduit $c^2 = c$ puis $c = 1$ car $\varphi \neq 0$. Finalement φ est constante, ce qu'il fallait démontrer. □

3. Systèmes totalement ergodiques

3.1. — Un système dynamique mesuré $(X, \mathcal{A}, \mu, f)$ est dit **totalement ergodique** si pour tout $n \geqslant 1$, $(X, \mathcal{A}, \mu, f^n)$ est ergodique. Nous avons déjà rencontré cette notion au §III.4 sur les théorèmes ergodiques non-conventionnels. Il est clair que cette propriété est plus forte que l'ergodicité: le système dynamique g induit sur $\{1, \ldots, n\}$ par un cycle de période n est ergodique pour la mesure équilibrée (pourquoi?) mais g^n n'est pas ergodique.

3.2. *Exercice.* — Montrer qu'un système faiblement mélangeant est totalement ergodique. La réciproque est elle vraie?

3.3. — Comme on l'a vu ci-dessus, il est facile de fabriquer des systèmes ergodiques et non totalement ergodiques sur des ensembles finis. Plus généralement si $(X, \mathcal{A}, \mu, f)$ est tel qu'il existe un facteur fini du type (Y, ν, g) non-trivial, au sens où Y est un ensemble fini (muni de sa tribu discrète), g est une bijection $Y \to Y$ et ν ne coincide pas avec une masse de Dirac alors (X, μ, f) n'est pas totalement ergodique. En effet si π désigne l'application mesurable induite $X \to Y$, et $y \in Y$ est tel que $0 < \nu(\{y\}) < 1$, alors comme pour $n = (\#Y)!$ on a $g^n = \mathrm{id}_Y$, on voit que $\pi^{-1}(\{y\})$ est f^n-invariant. Il se trouve que cette construction est la seule obstruction à la totale ergodicité.

Théorème. — Un système dynamique mesuré $(X, \mathcal{A}, \mu, f)$ est totalement ergodique si et seulement s'il n'admet pas de facteur fini non trivial.

Démonstration. — Le sens direct a déjà été démontré ci-dessus. Réciproquement supposons que $(X, \mathcal{A}, \mu, f)$ n'est pas totalement ergodique: il s'agit de construire un facteur fini non-trivial. Une première possibilité est que f ne soit pas ergodique. Il existe alors $A \in \mathcal{A}$ tel que $f^{-1}(A) = A$ et $0 < \mu(A) < 1$. On définit alors $\pi : X \to \{0, 1\}$ par $\pi|_A = 0$ et $\pi|_{A^\complement} = 1$, qui semiconjugue f à id_Y, et le résultat est démontré.

Supposons maintenant f ergodique et non totalement ergodique, et soit n l'entier minimal tel que f^n ne soit pas ergodique. Il existe alors $A \in \mathcal{A}$ tel que $f^{-n}(A) = A$ et $0 < \mu(A) < 1$. Remarquer que par ergodicité $\bigcup_{i=0}^{n-1}(f^{-i}A)$ est de mesure totale. Nous affirmons que si l'on peut trouver un tel A tel qu'en outre $\mu(A) = 1/n$, le résultat sera démontré: en effet dans ce cas les $f^{-i}A$ sont disjoints mod. 0 et on définit alors $\pi : X \to \{0, \ldots, n-1\}$ par $\pi|_{f^{-i}(A)} = i$, qui envoie μ sur la mesure équilibrée et semiconjugue f à $i \mapsto i + 1$ mod.n.

Reste à trouver un tel A. Posons $\varphi = \sum_{i=0}^{n-1} \mathbf{1}_A \circ f^i$. Comme $\mathbf{1}_A \circ f^i = \mathbf{1}_A$, φ est f-invariante, et par ergodicité elle est donc constante. En outre φ est à valeurs entières et $\int \varphi \, d\mu = n\mu(A)$, on en déduit qu'il existe $1 \leqslant k \leqslant n-1$ tel que $\mu(A) = k/n$. Si $k \geqslant 2$, les $f^i(A)$, $0 \leqslant i \leqslant n-1$ ne peuvent pas être tous disjoints mod. 0, il existe donc un tel i tel que $\mu(A \cap f^{-i}(A)) > 0$. Si $A = f^{-i}(A)$ mod. 0, alors f^i n'est pas ergodique, contredisant la minimalité de n. Donc $0 < \mu(A \cap f^{-i}(A)) < \mu(A)$. On remarque alors que $\tilde{A} := A \cap f^{-i}(A)$ est f^n-invariant, donc $\mu(\tilde{A})$ est à nouveau un multiple de $1/n$. Ainsi si l'on répète cet argument, le processus s'arrête et on aboutit à un $\tilde{A}$ tel que $\mu(\tilde{A}) = 1/n$, ce qui conclut la preuve. $\square$

4. Décomposition ergodique

4.1. — Nous allons ici donner (sans démonstration) une version plus précise du principe de décomposition ergodique qui a déjà été évoquée à la Remarque 3.9 du Chapitre II. L'idée est simple: toute mesure invariante μ s'écrit comme une moyenne de mesures ergodiques. Mieux: on peut construire une partition telle que les composantes de la décomposition de μ sont les "conditionnelles" de μ le long de cette partition, en particulier la décomposition ergodique est unique (ce qui n'apparaît pas naturellement quand on pense le problème uniquement en termes de géométrie convexe). La mise en place de ces idées soulève quelques points délicats de théorie de la mesure que nous allons brièvement discuter. Pour les détails sur le contenu de cette section, la lectrice est renvoyée aux livres de Coudène [**5**] (chapitres 13, 14 et 15) et Przytycki-Urbanski (chapitre 2).

4.2. *Définition-Proposition.* — Soit $(X, \mathcal{A}, \mu)$ un espace probabilisé complet (au sens où tout sous-ensemble d'un ensemble de mesure nulle est mesurable). On dit que $(X, \mathcal{A}, \mu)$ est un espace de Lebesgue, ou Lebesgue-Rokhlin, ou espace probabilisé standard si l'une des conditions équivalentes suivantes est satisfaite:

(1) $(X, \mathcal{A}, \mu)$ est isomorphe mod. 0 à $([0, 1], \mathcal{B}([0, 1]), \nu)$ où $\nu = c\text{Leb} + \sum_{n=0}^{\infty} \alpha_n \delta_{1/n}$.
(2) $(X, \mathcal{A}, \mu)$ est isomorphe mod. 0 à $(B, \mathcal{B}, \nu)$, où B un un borélien d'un espace métrique complet séparable, $\mathcal{B}$ est la trace de la tribu borélienne sur B est ν est une mesure borélienne.
(3) Il existe un plongement de $(X, \mathcal{A}, \mu)$ comme sous-espace mesuré de mesure totale dans un espace $(\overline{X}, \overline{\mathcal{A}}, \overline{\mu})$ qui a la propriété d'avoir une "base dénombrable" $(B_n) \in \overline{\mathcal{A}}^{\mathbb{N}}$ qui (i) sépare les points, (ii) engendre $\overline{\mathcal{A}}$ mod. 0, et (iii) pour toute suite $(\varepsilon_n) \in \{1, \complement\}^{\mathbb{N}}$, $\bigcap_{n=0}^{\infty} B_n^{\varepsilon_n}$ est non vide (et donc réduit à un point par (i); on a noté ici $B_n^1 = B_n$).

Le message qu'il faut retenir de cette définition quelque peu indigeste est: **les espaces "raisonnables" sont de Lebesgue**. On pourra à titre d'exercice construire une telle base (B_n) dans le cas de l'intervalle $[0, 1]$ muni de sa mesure de Lebesgue (indication: considérer un isomorphisme mod. 0 avec $\{0, 1\}^{\mathbb{N}}$ muni de la mesure équilibrée).

4.3. — L'intérêt des espaces de Lebesgue est qu'on peut y définir des partitions mesurables. Soit $\mathcal{P}$ une partition de X. Pour $x \in X$ nous noterons $\mathcal{P}(x)$ l'atome (la pièce) P de $\mathcal{P}$ contenant x. Désignons $\sigma_{\mathcal{A}}(\mathcal{P})$ la sous σ-algèbre de $\mathcal{A}$ formée des unions d'atomes de $\mathcal{P}$. Par définition on dit que $\mathcal{P}$ est une **partition mesurable** s'il existe un sous ensemble dénombrable (B_n) de $\sigma_{\mathcal{A}}(\mathcal{P})$ tel que tout atome de $\mathcal{P}$ peut s'écrire comme une intersection d'élements de B_n et de leurs complémentaires: $\mathcal{P}(x) = \bigcap_{n=0}^{\infty} B_n^{\varepsilon_n}$, où $(\varepsilon_n) \in \{1, \complement\}^{\mathbb{N}}$ (ici on note $B_n^1 = B_n$).

De manière plus élémentaire, une partition $\mathcal{P}$ est mesurable s'il existe une fonction mesurable $f : X \to [0, 1]$ telle que $\mathcal{P}$ soit l'image réciproque de la partition de $[0, 1]$ en points: $\mathcal{P}(x) = f^{-1}(\{f(x)\})$.

4.4. — Attention il ne suffit pas que tous les atomes de $\mathcal{P}$ soient mesurables pour que $\mathcal{P}$ le soit:

Exercice. — Montrer que la partition du cercle selon les orbites d'une rotation irrationnelle n'est pas une partition mesurable.

4.5. — Il faut bien comprendre que les partitions dont il est question ici ne sont pas dénombrables, et que les atomes sont en général de mesure nulle. Néanmoins on peut **désintégrer** la mesure μ le long d'une partition mesurable:

Théorème. — Soit $(X, \mathcal{A}, \mu)$ un espace de Lebesgue et $\mathcal{P}$ une partition mesurable de X. Alors il existe un ensemble d'atomes $\mathcal{P}'$ de mesure totale et une unique famille de mesures de probabilité μ_P telles que pour tout $P \in \mathcal{P}'$, $\mu_P(P) = 1$ et pour tout $A \in \mathcal{A}$, $A \cap P$ est μ_P-mesurable, la fonction $x \mapsto \mu_{\mathcal{P}(x)}(A)$ est mesurable et

$$\mu(A) = \int \mu_{\mathcal{P}(x)}(A)\mathrm{d}\mu(x). \tag{1}$$

Les $\mu_{\mathcal{P}(x)}$ sont appelées conditionnelles de μ selon $\mathcal{P}$. Nous avons volontairement laissé de côté la question délicate de définir exactement ce que sont les ensembles μ_P-mesurables.

Ce théorème est évidemment très similaire au théorème d'existence de l'espérance conditionnelle. Une différence, qui n'apparaît pas vraiment dans la présentation qui a été faite, est que la construction dépend peu de μ, au sens où si μ et ν sont des mesures boréliennes sur un espace (disons) métrique compact et $\mathcal{P}$ est une partition mesurable, alors on peut définir simultanément et comparer les conditionnelles relativement à $\mathcal{P}$.

Il y a une certaine multiplicité dans l'intégrale (1) puisque la même pièce $\mathcal{P}$ apparait pour tous les $x \in P$. On peut naturellement se demander s'il existe une structure mesurable sur espace quotient $X/\mathcal{P}$ qui permet d'écrire (1) comme une intégrale sur l' "espace des atomes": la théorie des espaces de Lebesgue permet bien de donner un sens à cela.

4.6. — Il est temps de revenir à l'objet initial de cette section: la décomposition ergodique. Nous allons faire apparaître les composantes ergodiques de μ comme les conditionnelles relatives à une certaine partition. Soit (X, f) un système dynamique défini sur un espace métrique séparable, et définissons une partition $\mathcal{P}$ en définissant $\mathcal{P}(x)$ comme l'ensemble des $y \in X$ tels que pour toute fonction continue bornée φ sur X, on a

$$\frac{1}{n}\sum_{k=0}^{n-1} \left(\varphi(f^k(y)) - \varphi(f^k(x))\right) \underset{n\to\infty}{\longrightarrow} 0.$$

Noter que la définition ne dépend pas de μ. On remarque que pour tout x, $\mathcal{P}(x)$ est invariant. Si ν est une mesure ergodique, d'après le théorème de Birkhoff elle donne masse totale à une unique pièce de $\mathcal{P}$.

Exercice. — Pour tout $x \in X$, montrer qu'il existe au plus une probabilité invariante ν telle que $\nu(\mathcal{P}(x)) = 1$ et que si elle existe, celle-ci est ergodique.

Lemme. — Si X est un borélien d'un espace métrique complet séparable, et f est une application borélienne préservant une mesure de probabilité μ, alors la partition $\mathcal{P}(x)$ construite ci-dessus est mesurable.

On a ainsi pour μ-presque tout x une probabilité conditionnelle $\mu_{\mathcal{P}}(x)$ sur $\mathcal{P}(x)$. Par invariance de μ et unicité de la mesure conditionnelle on a p.s. $f_*\mu_{\mathcal{P}}(x) = \mu_{\mathcal{P}}(x)$, et donc d'après l'exercice ci-dessus, $\mu_{\mathcal{P}}(x)$ est ergodique. La désintégration (1) permet bien de réaliser μ comme une intégrale de mesures ergodiques.

Pour un espace de Lebesgue général, en prenant un isomorphisme mod. 0 avec un borélien d'un espace métrique complet on obtient ainsi le théorème de décomposition ergodique suivant:

4.7. *Théorème. — Soit* $(X, \mathcal{A}, \mu, f)$ *un système dynamique mesuré sur un espace de Lebesgue. Il existe une unique décomposition de* μ *comme une intégrale de mesures ergodiques. Celle-ci est en outre la désintégration de* μ *relativement à une partition mesurable* f*-invariante.*

L'unicité n'est pas très facile à définir formellement, mais le contenu est clair: si $\mu = \int \mu_\alpha d\rho(\alpha)$ est n'importe quelle décomposition en mesures ergodiques, alors pour presque tout α, μ_α coincide avec l'une des conditionnelles $\mu_{\mathcal{P}(x)}$. En effet, si E_α est un borélien invariant de mesure totale pour μ_α, il doit coincider mod.0 avec une des pièces de la partition $\mathcal{P}$.

5. Exposant de Lyapunov et théorie de Pesin en dimension 1

Soit f un système dynamique de classe C^1 sur un intervalle I ou sur le cercle $\mathbb{R}/\mathbb{Z}$, et μ une mesure borélienne invariante.

5.1. *Lemme. — Si* $\log|f'| \in L^1(\mu)$ *alors pour* μ*-presque tout* x*, la limite*

$$\chi(x) = \lim_{n\to\infty} \frac{1}{n} \log|(f^n)'(x)|$$

existe et définit une fonction mesurable et invariante.

Par définition $\chi(x)$ est **l'exposant de Lyapunov** au point x, qui mesure le taux asymptotique de dilatation de la dynamique au point x. Un exposant strictement positif signifie que la dynamique est (infinitésimalement) dilatante en x, alors qu'un exposant strictement négatif signifie qu'elle est contractante. Remarquer que la définition de $\chi(x)$ fait intervenir une mesure invariante auxiliaire μ, qui sert à montrer l'existence presque sûre de la limite, mais $\chi(x)$ ne dépend que de x. On voit en particulier que $\chi(x)$ est bien défini pour presque tout point pour toute mesure invariante.

Démonstration. — C'est une application du théorème ergodique de Birkhoff. En effet la formule de dérivation des fonctions composés implique que

$$(f^n)'(x) = \prod_{k=0}^{n-1} f'(f^k(x)), \tag{2}$$

ainsi on voit que

$$\frac{1}{n}\log|(f^n)'(x)| = \frac{1}{n}\sum_{k=0}^{n-1} \log\left|f'(f^k(x))\right|$$

est une somme de Birkhoff, ce qui conclut la preuve. □

5.2. — On définit également **l'exposant de Lyapunov moyen** de μ par $\chi_\mu = \int \chi(x)\, d\mu(x)$ et on voit que

$$\chi_\mu = \int \log |f'|\ d\mu.$$

Si μ est ergodique alors pour μ-presque tout x on a $\chi(x) = \chi_\mu$.

5.3. — En dimension supérieure, l'équation (2) devient un produit de matrices, et la notion d'exposant de Lyapunov devient plus délicate. Ce sera l'objet d'étude principal du chapitre VI.

5.4. *Exercice.* — Pour $0 < a < 1$ on définit une application f_a sur l'intervalle $[0, 1]$ par

$$\begin{cases} f_a(x) = x/a \text{ si } 0 \leqslant x \leqslant a \\ f_a(x) = (1-x)/(1-a) \text{ si } a \leqslant x \leqslant 1. \end{cases}$$

(1) Montrer que pour tout $0 < a < 1$ f_a est topologiquement conjuguée à l'application tente $f_{1/2}$.
(2) Montrer que la mesure de Lebesgue est f_a-invariante et ergodique.
(3) Calculer son exposant de Lyapunov moyen.

5.5. *Exercice.* — Comme à l'exercice II.5.2.3 on considère l'application $f : x \mapsto 2x^2 - 1$ définie sur l'intervalle $[-1, 1]$.

(1) Montrer que f admet une unique mesure invariante absolument continue par rapport à la mesure de Lebesgue.
(2) Calculer son exposant de Lyapunov moyen.

5.6. — On doit s'attendre à ce que l'information infinitésimale apportée par l'exposant de Lyapunov $\chi(x)$ permette de tirer des conclusions sur la dynamique des points voisins: si $\chi(x) < 0$ on devrait voir une contraction, et si $\chi(x) > 0$ une dilatation (plus précisément une contraction pour l'itération en temps négatif). Cette problématique (en dimension arbitraire) est l'objet de la **théorie de Pesin**, qui vise à comprendre comment en dynamique différentiable l'information donnée par les exposants de Lyapunov se traduit localement. Voici probablement l'énoncé le plus simple de la théorie:

5.7. *Théorème. — Soit f un système dynamique de classe C^2 sur le cercle ou l'intervalle $[0, 1]$. Soit μ une mesure invariante ergodique telle que* $\log |f'| \in L^1(\mu)$. *On suppose que son exposant de Lyapunov χ_μ est strictement négatif. Alors μ est la mesure équidistribuée sur une orbite périodique attractive.*

D'un point de vue dynamique, le cas le plus intéressant est donc celui où l'exposant de Lyapunov est positif. Des méthodes analogues permettent de construire des **branches inverses** contractantes qui permettent d'analyser finement μ (voir par exemple [**12**, Chap. 11]).

Remarque. — L'hypothèse $\log |f'| \in L^1(\mu)$ n'est pas très naturelle, en effet dans ce cas si $\log |f'|$ n'est pas intégrable, c'est que $\int \log |f'|\, d\mu = -\infty$ et on devrait s'attendre à encore plus de contraction: en pratique c'est bien ce qui se produit, mais la démonstration est un peu plus délicate.

5.8. — Le corollaire suivant est immédiat en appliquant le théorème à f et f^{-1}. On peut l'interpréter en disant qu'il n'y a pas de dynamique chaotique inversible sur le cercle[(1)].

Corollaire. — Soit f un difféomorphisme de classe C^2 du cercle. Si μ est une mesure invariante ergodique qui n'est pas concentrée sur une orbite périodique, alors son exposant de Lyapunov est nul.

5.9. — Le cœur de la démonstration du théorème est la construction au voisinage d'un point générique x d'un petit intervalle où la dynamique est asymptotique à celle de x (c'est un cas simple de "variété stable de Pesin")

Proposition. — Soit f un système dynamique de classe C^2 sur le cercle ou l'intervalle $[0,1]$. Soit x tel que:

(1) *son exposant de Lyapunov $\chi(x) = \lim_{n\to\infty} \frac{1}{n} \log |(f^n)'(x)|$ est bien défini et strictement négatif;*
(2) *les itérés de x n'approchent pas trop vite l'ensemble critique:*

$$\liminf_{n\to\infty} \frac{1}{n} \log |f'(f^n(x))| \geqslant 0.$$

Alors il existe un voisinage I de x tel que pour tout $y \in I$ on a $d(f^n(y), f^n(x)) \to 0$ quand $n \to \infty$, et plus précisément

$$\lim_{n\to\infty} \frac{1}{n} \log d(f^n(y), f^n(x)) = \chi(x).$$

Remarquer que l'hypothèse (2) est automatiquement satisfaite si f est un difféomorphisme. Par ailleurs la preuve montre que dans (1) il suffirait de supposer que $\limsup_{n\to\infty} \frac{1}{n} \log |(f^n)'(x)| \leqslant \chi(x)$. La difficulté est que l'approche naïve pour la comparaison entre $(f^n)'(x)$ et $(f^n)'(y)$ pour y proche de x fait intervenir la dérivée seconde de (f^n) qui n'est pas directement contrôlable. L'idée consiste à estimer plutôt le produit:

$$\frac{(f^n)'(y)}{(f^n)'(x)} = \prod_{k=0}^{n-1} \frac{f'(y_k)}{f'(x_k)}, \text{ où } x_k = f^k(x) \text{ et } y_k = f^k(y).$$

Cette quantité s'appelle la **distortion** de f^n, dont la compréhension est un point clé de nombreux problèmes de dynamique différentiable en dimension 1.

Démonstration de la Proposition 5.9. — Posons $\chi = \chi(x) < 0$, et comme précédemment pour tout y on note $f^k(y) = y_k$. Fixons $\varepsilon > 0$ petit devant χ ($\varepsilon = |\chi|/4$ suffit pour montrer que $d(f^n(y), f^n(x)) \to 0$). L'hypothèse (1) assure l'existence d'une constante $C_1 > 0$ telle que pour tout $n \geqslant 0$ $|(f^n)'(x)| \leqslant C_1 e^{-n(|\chi|-\varepsilon)}$ et par (2), il existe $C_2 > 0$ telle que pour tout $n \geqslant n_0$, $|f'(f^n(x))| \geqslant C_2 e^{-n\varepsilon}$. Dans la suite sans perte de généralité on

---
[(1)] Il y a en fait une théorie classique de la dynamique des homéomorphismes du cercle (voir [**4**] ou [**6**]) qui montre ce résultat de façon plus directe et naturelle.

remplace x par $f^{n_0}(x)$ et on pose $n_0 = 0$. Par le caractère C^2 de f, il existe une constante $C_3 = \|f''\|$ telle que si y et z sont tels que[(2)] $d(y,z) \leqslant 1/10$, alors on a

$$|f'(x) - f'(y)| \leqslant C_3 d(x,y).$$

De même si $\rho \leqslant 1/10$ on a

$$\forall z \in B(y,\rho),\ d(f^k(y), f^k(x)) \leqslant \left(\sup_{B(y,\rho)} \left|(f^k)'\right|\right) d(x,y).$$

Nous allons montrer par récurrence l'existence d'une constante ρ telle que si $y \in B(x,\rho)$ alors pour tout $n \geqslant 0$ on a $d(x_n, y_n) \leqslant e^{-(|\chi|-\varepsilon)n}/10$, ce qui implique le résultat souhaité.

Raisonnons par analyse-synthèse, et supposons que cette estimation soit satisfaite pour tout $0 \leqslant k \leqslant n-1$. Pour tout $y \in B(x,\rho)$ on a alors

$$\frac{(f^n)'(y)}{(f^n)'(x)} = \prod_{k=0}^{n-1} \frac{f'(y_k)}{f'(x_k)} = \prod_{k=0}^{n-1} \left(1 + \frac{f'(x_k) - f'(y_k)}{f'(x_k)}\right)$$

d'où l'on tire

$$\frac{|(f^n)'(y)|}{|(f^n)'(x)|} \leqslant \prod_{k=0}^{n-1} \left(1 + \frac{C_3 e^{-(|\chi|-\varepsilon)k}/10}{C_2 e^{-k\varepsilon}}\right) \leqslant D := \prod_{k=0}^{\infty} \left(1 + \frac{C_3}{10 C_2} e^{-(|\chi|-2\varepsilon)k}\right).$$

Ainsi il vient

$$d(x_n, y_n) \leqslant \left(\sup_{B(y,\rho)} |(f^n)'|\right) d(x,y) \leqslant D\,|(f^n)'(x)|\,\rho \leqslant C_1 D \rho\, e^{-n(|\chi|-\varepsilon)}.$$

On voit donc qu'il suffit donc de choisir $\rho = C_1 D/20$ pour que la récurrence soit satisfaite, et on obtient que pour $y \in I$,

$$\limsup_{n\to\infty} \frac{1}{n} \log d(f^n(y), f^n(x)) \leqslant \chi(x) + \varepsilon.$$

Pour montrer que $\lim_{n\to\infty} \frac{1}{n} \log d(f^n(y), f^n(x)) = \chi(x)$, on reprend le même raisonnement une deuxième fois: maintenant que l'on sait que $\limsup \frac{1}{n} \log d(f^n(y), f^n(x)) < 0$, l'estimée de distortion dit que $|(f^n)'(x)| \approx |(f^n)'(y)|$ et on conclut (exercice: rédiger les détails de ce raisonnement). $\square$

Démonstration du Théorème 5.7. — Si x est un point μ-générique, alors les deux hypothèses de la Proposition 5.9 sont satisfaites: pour le (1) cela découle du théorème ergodique, et pour le (2) nous avons vu que si $\psi \in L^1(\mu)$ alors p.s. $\psi(f^n(x)) = o(n)$ (voir page 42). Soit $I(x)$ un intervalle stable fourni par la Proposition 5.9, et V un voisinage arbitraire de x. Par le théorème de récurrence de Poincaré, il existe une suite (n_j) telle que $f^{n_j}(x) \to x$, et donc par contraction $f^{n_j}(I(x)) \subset V$. Il vient alors $f^{-n_j}(V) \supset I(x)$, donc $\mu(f^{-n_j}(V)) = \mu(V) \geqslant \mu(I(x))$. Comme V est arbitraire, cela signifie que μ a un atome en x. On conclut par ergodicité que μ est concentrée sur une orbite périodique, qui doit nécessairement être attractive (exercice). $\square$

Remarques. —

[(2)] Cette démonstration est volontairement rédigée dans l'esprit d'une généralisation à une variété compacte quelconque, et il faut imaginer cette constante $1/10$ comme étant le diamètre des cartes.

1. Nous n'avons pas vraiment utilisé l'hypothèse d'ergodicité. En effet la preuve montre que si dans le Théorème 5.7 on suppose seulement que $\chi(x) < 0$ p.p., alors μ est supportée par une famille au plus dénombrable d'orbites périodiques attractives.
2. Au delà des applications C^2, le bon cadre pour la théorie de Pesin est celui des applications de classe $C^{1+\alpha}$, c'est à dire C^1 et dont la dérivée vérifie une condition de Hölder d'exposant α.

◇

V. UNIQUE ERGODICITÉ

1. Définition et premières propriétés

Dans toute cette section X désigne un espace métrique compact.

1.1. *Définition.* — Soit f une application continue sur X. On dit que le système dynamique (X, f) est uniquement ergodique s'il admet une unique mesure borélienne invariante.

1.2. *Proposition.* — *(X, f) est uniquement ergodique si et seulement s'il admet une unique mesure ergodique.*

Démonstration. — Ceci découle du Lemme II.3.6 (et de la remarque qui le suit). En effet si μ est l'unique mesure borélienne invariante, alors c'est nécessairement un point extrémal de l'espace des mesures invariantes, et donc μ est ergodique. Pour la réciproque, il suffit de se rappeler que le convexe des mesures invariantes est l'enveloppe convexe fermée de l'ensemble des mesures ergodiques. □

1.3. *Proposition.* — *(X, f) est uniquement ergodique si et seulement s'il existe une mesure μ telle que pour tout $x \in X$ on a $\frac{1}{n}\sum_{k=0}^{n-1}\delta_{f^k(x)} \rightharpoonup \mu$ quand $n \to \infty$.*

Démonstration. — Pour le sens direct on observe déjà que pour tout $x \in X$, les valeurs d'adhérence de la suite $\frac{1}{n}\sum_{k=0}^{n-1}\delta_{f^k(x)}$ sont des mesures invariantes. Donc il n'y a qu'une valeur d'adhérence possible et la suite converge donc vers l'unique mesure invariante μ. Réciproquement si ν est une mesure ergodique quelconque, par la Proposition 1.4 on a pour presque tout x, $\frac{1}{n}\sum_{k=0}^{n-1}\delta_{f^k(x)} \rightharpoonup \nu$. Ainsi il n'y a qu'une mesure ergodique et on conclut par la proposition précédente. □

1.4. *Proposition.* — *(X, f) est uniquement ergodique si et seulement si pour toute fonction continue φ sur X, la suite $\frac{1}{n}\sum_{k=0}^{n-1}\varphi\circ f^k$ converge uniformément vers une fonction constante.*

Remarque. — Par des arguments déjà vus plusieurs fois, on peut se contenter de vérifier l'hypothèse sur une famille dense de fonctions continues.

Démonstration. — Pour le sens direct on raisonne par l'absurde: supposons qu'il existe des suites (x_j) et (n_j) telles que $\frac{1}{n_j}\sum_{k=0}^{n_j-1}\varphi(f^k(x_j))$ converge vers $\ell \neq \int\varphi\,\mathrm{d}\mu$. On extrait alors à nouveau de sorte que $\frac{1}{n_j}\sum_{k=0}^{n_j-1}\delta_{f^k(x_j)}$ converge vers une mesure de probabilité ν, qui comme dans le cas où x est fixé, doit être invariante. Ainsi $\nu = \mu$ et $\int\varphi\,d\nu = \ell$, ce qui est contradictoire.

Réciproquement supposons que pour toute $\varphi \in \mathcal{C}(X)$ la suite $\frac{1}{n}\sum_{k=0}^{n-1}\varphi\circ f^k$ converge vers une fonction constante c_φ. Alors $\varphi \mapsto c_\varphi$ est une forme linéaire positive de norme

1 sur $\mathcal{C}(X)$ et donc par le théorème de représentation de Riesz, il existe une mesure de probabilité μ telle que $c_\varphi = \int \varphi \, d\mu$. En outre pour tout x on a pour toute $\varphi \in \mathcal{C}(X)$ $\frac{1}{n}\sum_{k=0}^{n-1}\varphi(f^k(x)) \to \int \varphi \, d\mu$, ainsi $\frac{1}{n}\sum_{k=0}^{n-1}\delta_{f^k(x)} \rightharpoonup \mu$ et par le théorème de Birkhoff, il n'existe qu'une seule mesure ergodique, et ainsi f est uniquement ergodique. □

2. Exemples de systèmes uniquement ergodiques

2.1. — Voici un exemple basique:

Proposition. — Les rotations irrationnelles sont uniquement ergodiques.

Démonstration. — Soit α un angle irrationnel et R_α la rotation d'angle $2\pi\alpha$ (ou translation de α dans $\mathbb{R}/\mathbb{Z}$). Pour tout polynôme trigonométrique P on vérifie simplement que pour tout x, $\frac{1}{n}\sum_{k=0}^{n-1} P(f^k(x))$ converge vers Leb(P). Comme les polynômes trigonométriques sont denses dans $\mathcal{C}(\mathbb{S}^1)$ on peut passer à l'adhérence et on obtient la même propriété pour toute $\varphi \in \mathcal{C}(\mathbb{S}^1)$, et on conclut par la proposition 1.3. □

2.2. — **Application: distribution des décimales dans les tables de logarithmes**. On appelle souvent **loi de Benford** l'observation empirique que dans beaucoup de jeux de données numériques où l'ordre de grandeur n'est pas imposé **a priori** (du type chiffre d'affaires, consommation d'électricité, etc.) la distribution du premier chiffre n'est pas uniforme: la probabilité d'apparition du chiffre k (en base 10) est égale à $\log_{10}\frac{k+1}{k}$. Elle a d'abord été observée empiriquement à l'examen de l'usure des pages des tables de logarithmes(1)! Le théorème suivant illustre ce phénomène.

Théorème. — Soit r_n le premier chiffre de 2^n en base 10. Alors pour tout $1 \leqslant k \leqslant 9$ on a

$$\frac{1}{N}\#\{1 \leqslant n \leqslant N,\ r_n = k\} \underset{N\to\infty}{\longrightarrow} \log_{10}\frac{k+1}{k}.$$

Démonstration. — On remarque que

$$\begin{aligned} r_n = k &\Leftrightarrow \exists r \in \mathbb{N},\ k\cdot 10^r < 2^n \leqslant (k+1)\cdot 10^r \\ &\Leftrightarrow \exists r \in \mathbb{N},\ r + \log_{10}(k) \leqslant n\log_{10}(2) < r + \log_{10}(k+1) \\ &\Leftrightarrow n\log_{10}(2) \mathrm{mod}.1 \in [\log_{10}(k), \log_{10}(k+1)[. \end{aligned}$$

En outre $\alpha = \log_{10}(2) = \log(2)/\log(10)$ est irrationnel. Par unique ergodicité des rotations irrationnelles on voit que la fréquence des retours de 0 par R_α^n dans l'intervalle

$$[\log_{10}(k), \log_{10}(k+1)[$$

converge vers sa longueur, d'où le résultat. □

(1)voir la page wikipedia `https://fr.wikipedia.org/wiki/Loi_de_Benford`

2.3. — L'exemple précédent se généralise au cas des translations sur des groupes plus généraux. Soit G un groupe topologique compact. Alors celui-ci est automatiquement métrisable (théorème de Birkhoff-Kakutani). Il est classique qu'un groupe topologique G localement compact admet une **mesure de Haar**, c'est à dire une (unique) mesure borélienne μ invariante par translation à gauche: pour tout $g \in G$, on a $(\ell_g)_*\mu = \mu$, où $\ell_g : x \mapsto g \cdot x$. En outre, si G est compact, cette mesure est finie, on la normalise alors par $\mu(G) = 1$. On voit alors que μ est également invariante à droite: en effet pour tout $h \in G$ si on note $r_h : x \mapsto x \cdot h$ la translation à droite, on a $\ell_g \circ r_h = r_h \circ \ell_g$ pour tous g, h. Ainsi $(\ell_g)_*((r_h)_*\mu) = (r_h)_*((\ell_g)_*\mu) = (r_h)_*\mu$ et on voit que $(r_h)_*\mu$ est invariante à gauche et par unicité on conclut $(r_h)_*\mu = \mu$.

Théorème. — Soit G un groupe topologique compact et τ une translation à gauche sur G. Alors si τ est topologiquement transitive, τ est uniquement ergodique.

Remarquer que l'hypothèse de transitivité est nécessaire: penser aux rotations *rationnelles.*

Démonstration. — Posons $\tau = \ell_g$. Par hypothèse il existe $x \in G$ tel que $(\tau^n(x))$ est dense, i.e. $(g^n x)$ est dense, et donc (g^n) est dense. Soit maintenant ν une mesure τ-invariante. Pour tout $h \in G$, prenons une suite n_j telle que $g^{n_j} \to h$. Alors

$$(\ell_h)_*\nu = \lim_{j\to\infty} (\tau^{n_j})_*\nu = \nu$$

et on voit que ν est la mesure de Haar. □

2.4. — En combinant le théorème précédent avec les résultats du § 5.4 du chapitre II, on obtient le:

Corollaire. — Si $\alpha \in \mathbb{R}^n$ est tel que la famille $(1, \alpha_1, \dots, \alpha_n)$ est $\mathbb{Q}$-libre, alors la translation τ_α sur $\mathbb{T}^n$ est uniquement ergodique.

(Pour montrer la transitivité topologique, on peut également utiliser la structure des sous-groupes de $\mathbb{R}^n$, ou alors montrer directement l'unique ergodicité comme à la Proposition 2.1.) Voici une conséquence amusante:

Exercice. — Il existe un entier n tel que les écritures décimales de 2^n et 3^n commencent par les quatre chiffres 2023.

2.5. — Un exemple intéressant qui relève de cette catégorie est la "**dyadic adding machine**" (ou odomètre): il s'agit de la translation de 1 sur le groupe topologique $\mathbb{Z}_2$. Rappelons que $\mathbb{Z}_2$ est construit de la manière suivante: on considère la distance sur $\mathbb{Z}$ définie par $d_2(x, y) = |x - y|_2 = 2^{-v_2(x-y)}$, où v_2 est la valuation 2-adique (i.e. si $x = 2^k y$ avec y impair alors $v_2(x) = k$). On vérifie simplement qu'il s'agit bien d'une distance qui a la propriété que $d_2(x, y)$ est petit lorsque $(x - y)$ est divisible par une grande puissance de 2. L'addition sur $\mathbb{Z}$ est continue pour cette distance. Par définition $\mathbb{Z}_2$ est la complétion de $(\mathbb{Z}, d_2)$. L'addition s'étend par prolongement uniformément continu et ainsi $(\mathbb{Z}_2, +)$ est un groupe.

En pratique, tout entier naturel admet un développement en base 2, $n = \sum_{i=0}^k \varepsilon_i 2^i$ avec $\varepsilon_i \in \{0, 1\}$. Comme $|2^i|_2 = 2^{-i}$, on voit que toute série de la forme $\sum_{i\geqslant 0} \varepsilon_i 2^i$ converge dans

$\mathbb{Z}_2$, et on montre que l'on obtient ainsi tous les éléments de $\mathbb{Z}_2$ (et l'écriture comme série est unique). On voit également que $\mathbb{Z}_2$ est homéomorphe à $\{0,1\}^{\mathbb{N}}$, c'est donc topologiquement un ensemble de Cantor. Pour faire des additions dans ce groupe, on utilise simplement l'écriture sous la forme $\sum_{i=0}^{\infty} \varepsilon_i 2^i$, et on additionne comme en base 2 dans $\mathbb{N}$, en propageant (peut être infiniment!) les retenues. On vérifie ainsi par exemple que $1 + \sum_{i=0}^{\infty} 2^i = 0$, soit $\sum_{i=0}^{\infty} 2^i = -1$.

2.6. *Proposition. — Soit τ la translation de 1 dans $\mathbb{Z}_2$. Alors τ est uniquement ergodique.*

Démonstration. — Grâce au Théorème 2.3 il suffit de vérifier que τ est topologiquement transitive. Pour cela on regarde comment τ agit sur les cylindres

$$C_{\alpha_0,\dots,\alpha_n} = \left\{ \sum_{i=0}^{\infty} \varepsilon_i 2^i,\ \forall 0 \leqslant i \leqslant n, \varepsilon_i = \alpha_i \right\}.$$

Partant d'un cylindre de la forme $C_{0,0,\dots,0,\alpha_k}$ et ajoutant successivement 1, on va visiter $C_{1,0,\dots,0,\alpha_k}$ puis $C_{0,1,0\dots,0,\alpha_k}$, puis $C_{1,1,0\dots,0,\alpha_k}$, puis $C_{0,0,1\dots,0,\alpha_k}$ et ainsi de suite jusqu'à arriver à $C_{1,1,\dots,1\alpha_k}$ et finalement $C_{0,0,\dots,0,\beta_k}$ avec $\beta_k \neq \alpha_k$. On voit que partant de 0 on va visiter tous les cylindres, et comme ceux-ci engendrent la topologie de $\mathbb{Z}_2$, la suite $\tau^n(0)$ est dense et τ est topologiquement transitive. $\square$

2.7. — Voici une autre manière de généraliser le résultat du §2.1

Théorème. — Si f est un homéomorphisme du cercle n'admettant aucun point périodique, alors f est uniquement ergodique.

Le théorème peut être vu comme conséquence d'une classification générale due à Poincaré des homéomorphismes du cercle. Furstenberg a donné une démonstration directe, que nous proposons sous forme d'exercice[(2)]. Ici par définition le cercle sera vu comme le cercle unité $\mathbb{S}^1 = \{z \in \mathbb{C}, |z| = 1\}$. Étant donnés $z_1, z_2 \in \mathbb{S}^1$ on note $[z_1, z_2]$ l'arc fermé allant de z_1 à z_2 dans le sens trigonométrique. On fixe f comme dans l'énoncé et on considère deux mesures de probabilité invariantes μ et μ'. On pose $\lambda = (\mu + \mu')/2$.

(a) Montrer que toute mesure f-invariante est diffuse.
(b) Pour $z \in \mathbb{S}^1$ on pose $\Phi(z) = e^{2i\pi\lambda([1,z])}$. Montrer que Φ définit une application continue $\mathbb{S}^1 \to \mathbb{S}^1$.
(c) Montrer qu'il existe un réel α que l'on précisera, tel que pour tout $z \in \mathbb{S}^1$, $\Phi(f(z)) = e^{2i\pi\alpha}\Phi(z)$.
(d) Montrer que α est nécessairement irrationnel.
(e) En déduire que si m est une mesure de probabilité f-invariante alors $\Phi_* m$ est la mesure de Lebesgue (normalisée) de $\mathbb{S}^1$.
(f) Montrer que si I_1 et I_2 sont des intervalles fermés de $\mathbb{S}^1$ tels que $\Phi(I_1) = \Phi(I_2)$ alors $\mu(I_1) = \mu(I_2)$ et $\mu'(I_1) = \mu'(I_2)$.
(g) Conclure.

[(2)] Voir l'article original de Furstenberg *Strict ergodicity and transformations of the torus* Amer. J. Math. 83 (1961) pp. 573–601 pour la solution.

3. Minimalité et unique ergodicité.

3.1. — Le résultat suivant est une conséquence directe de la Proposition 1.3.

Proposition. — Si (X, f) *est uniquement ergodique et si l'unique mesure invariante est de support total (i.e. égal à* X*), alors* (X, f) *est minimal.*

Bien sûr le résultat est faux si on retire l'hypothèse sur le support de μ: prendre par exemple $x \mapsto x/2$ sur $[0, 1]$.

3.2. — La réciproque de la proposition précédente est fausse: il existe des exemples dûs à Furstenberg (1961) d'homéomorphismes du tore $\mathbb{T}^2$ minimaux mais non uniquement ergodiques (voir par exemple le livre de Parry [**10**, §5.5]).

On a toutefois une réciproque partielle. On dira qu'un homéomorphisme f d'un espace métrique compact X est **presque périodique** s'il est topologiquement transitif et que la suite des itérés (f^n) est équicontinue. On a alors:

Proposition. — Si (X, f) *est presque périodique alors* f *est uniquement ergodique.*

Noter que cette proposition donne une nouvelle approche pour l'unique ergodicité des rotations irrationnelles et de la dyadic adding machine.

Démonstration. — Comme $(f^n)_{n \geqslant 0}$ est équicontinue, pour toute fonction continue $\varphi : X \to \mathbb{R}$ la suite de fonctions $\frac{1}{n} \sum_{k=0}^{n-1} \varphi \circ f^k$ est bornée et équicontinue. Par le théorème d'Ascoli elle admet donc une valeur d'adhérence ψ (pour la topologie de la convergence uniforme). Mais alors ψ est une fonction continue et f-invariante, et comme f est transitive et ψ est constante sur les orbites, en considérant une orbite dense on en déduit que ψ est constante. En outre, si μ est une mesure invariante quelconque, on a $\mu\left(\frac{1}{n} \sum_{k=0}^{n-1} \varphi \circ f^k\right) = \mu(\varphi)$ donc toute valeur d'adhérence de cette suite vaut $\psi = \mu(\varphi)$. On en déduit que la suite $\frac{1}{n} \sum_{k=0}^{n-1} \varphi \circ f^k$ converge uniformément vers une constante et le résultat découle de la Proposition 1.4. □

4. G-extensions de systèmes uniquement ergodiques

Une construction de Furstenberg permet de construire des systèmes uniquement ergodiques par extensions successives. Soit (X, f) un système dynamique et G un groupe. Par définition une **G-extension de f** (ou **G-cocycle au dessus de f**) est une application $F : X \times G \to X \times G$ de la forme $F(x, y) = (f(x), g(x) \cdot y)$.

4.1. *Théorème (Furstenberg). — Soit* (X, f) *une transformation uniquement ergodique et* F *une* G*-extension de* f*, où* G *est un groupe topologique compact. On note* μ *l'unique mesure invariante de* f *et* m *la probabilité de Haar de* G*. Alors si* F *est ergodique pour* $\mu \times m$*, elle est uniquement ergodique.*

Démonstration. — On montre déjà que $\mu \times m$ est invariante. En effet pour tout ensemble produit de la forme $A \times B$ on a $F^{-1}(A \times B) = \bigcup_{x \in f^{-1}(A)} (\{x\} \times g(f(x))^{-1}(B))$ et par l'invariance de μ et le théorème de Fubini on voit que $(\mu \times m)(F^{-1}(A \times B)) = \mu(A) m(B)$. Comme $\mu \times m$ est ergodique on a que pour $\mu \times m$ presque tout (x, y), $\frac{1}{n} \sum_{k=0}^{n-1} \delta_{F^k(x,y)} \rightharpoonup$

$\mu \times m$. On note maintenant que F commute avec les translations à droite sur le deuxième facteur, i.e. si on note $R_h(x,y) = (x, y \cdot h)$ on a $R_h \circ F = F \circ R_h$. En particulier on a $\delta_{F^k(x,yh)} = (R_h)_* \delta_{F^k(x,y)}$ et ainsi si (x,y) est $\mu \times m$-générique alors pour tout $h \in G$, (x, yh) est également générique. On en déduit la propriété suivante: pour μ-presque tout $x \in X$, pour tout $y \in G$, (x,y) est $\mu \times m$-générique.

Soit maintenant ν une autre mesure F-invariante. Alors $\pi_* \nu$ est f-invariante, où $\pi : X \times G \to X$ est la projection naturelle. En effet

$$\pi_*\nu(f^{-1}(A)) = \nu(\pi^{-1}(f^{-1}(A))) = \nu(F^{-1}(\pi^{-1}(A))) = \nu(\pi^{-1}(A)) = \pi_*\nu(A).$$

Par unique ergodicité on en déduit que $\pi_*\nu = \mu$. Soit N l'ensemble des points ν-génériques. On a $\nu(N) = 1$ donc par Fubini on a $\mu(\pi(N)) = 1$. Mais par l'argument précédent, on a que pour presque tout $x \in \pi(N)$ et tout $y \in \pi^{-1}(x)$, (x,y) est $\mu \times m$-générique. Donc presque tout $(x,y) \in N$ est $\mu \times m$-générique et on conclut que $\nu = \mu \times m$. □

Voici une application typique:

4.2. *Théorème (Furstenberg). — Si α est un nombre irrationnel, et si les $(a_{j,k})_{1 \leqslant k < j \leqslant n}$ sont des entiers tels que pour tout k on a $a_{k,k-1} \neq 0$, alors l'application définie sur le tore $\mathbb{T}^n$ par*

$$(x_1, \dots, x_n) \longmapsto (x_1 + \alpha, x_2 + a_{2,1}x_1, x_3 + a_{3,2}x_2 + a_{3,1}x_1, \dots, x_n + a_{n,n-1}x_{n-1} + \dots + a_{n,1}x_1)$$

est uniquement ergodique.

Démonstration. — La démonstration se fait par récurrence sur n. Nous traiterons le cas $n = 2$ (le cas général est similaire, les notations sont juste plus lourdes). Nous avons donc $f(x,y) = (x + \alpha, y + ax)$ sur $\mathbb{T}^2$ avec a entier non nul et α irrationnel. D'après le Théorème 4.1 pour démontrer l'unique ergodicité il suffit de vérifier que F est ergodique pour la mesure de Haar (Lebesgue). Comme précédemment nous allons pour cela utiliser les séries de Fourier. Soit donc $\varphi \in L^2(\mathbb{T}^2)$ une fonction L^2 F-invariante. On a alors dans L^2

$$\varphi(x,y) = \sum_{(j,k) \in \mathbb{Z}^2} c_{j,k} e^{2i\pi jx} e^{2i\pi ky} \text{ avec } (c_{j,k}) \in \ell^2(\mathbb{Z}^2)$$

et ainsi

$$\begin{aligned}
\varphi \circ f(x,y) &= \sum_{(j,k) \in \mathbb{Z}^2} c_{j,k} e^{2i\pi j(x+\alpha)} e^{2i\pi k(y+ax)} \\
&= \sum_{(j,k) \in \mathbb{Z}^2} c_{j,k} e^{2i\pi j\alpha} e^{2i\pi(j+ak)x} e^{2i\pi ky} \\
&= \sum_{(j,k) \in \mathbb{Z}^2} c_{j'-ak',k'} e^{2i\pi(j'-ak')\alpha} e^{2i\pi j'x} e^{2i\pi k'y} \text{ par changement d'indice.}
\end{aligned}$$

Donc par unicité du développement en série de Fourier on obtient que pour tous j,k, $c_{j,k} = c_{j-ak,k} e^{2i\pi(j-ak)\alpha}$. En faisant $k = 0$ on a comme pour le cas des rotations $c_{j,0} = 0$ pour $j \neq 0$. Pour $k \neq 0$ on écrit que pour tout q $c_{j,k} = c_{j-aqk,k} e^{2i\pi(j-aqk)\alpha}$. Comme $c_{j-aqk,k} \to 0$ quand $q \to \infty$ on en déduit que $c_{j,k} = 0$. Finalement $\varphi = c_{0,0}$, et le résultat est démontré. □

5. Théorème de Weyl

Nous concluons ce chapitre par une très jolie application de ce qui précède à un résultat d'équidistribution.

5.1. *Théorème (Weyl). — Soit $P \in \mathbb{R}[X]$ un polynôme dont au moins un coefficient est irrationnel. Alors la suite $(P(n))_{n\geqslant 0}$ est équidistribuée modulo 1.*

La densité de $(P(n))$ est déjà un résultat tout à fait non trivial. Par exemple on en déduit le résultat suivant, qui était initialement un théorème de Hardy et Littlewood:

5.2. *Corollaire. — Si α est irrationnel alors pour tout $\varepsilon > 0$ il existe n et m supérieurs ou égaux à 1 tels que $|\alpha n^2 - m| < \varepsilon$.*

Démonstration du théorème. — La démonstration qui va suivre est due à Furstenberg. Plaçons nous d'abord dans le cas où le terme dominant de P est irrationnel, c'est à dire que $P(X) = a_d X^d + \cdots + a_0$, avec $a_d \notin \mathbb{Q}$. L'observation est la suivante. Posons $P_d = P$ et pour $i \leqslant d-1$ définissons par récurrence $P_{i-1}(X) = P_i(X+1) - P_i(X)$. On vérifie simplement que pour $1 \leqslant i \leqslant d-1$ P_i est un polynôme de degré i de coefficient dominant $d(d-1)\cdots(d-i+1)a_d$ (en effet si $P_i(X) = cX^i + \cdots$ alors $P_i(X+1) - P_i(X) = icX^{i-1} + \cdots$). Ainsi P_1 est de la forme $P_1(X) = \alpha X + \beta$, avec $\alpha = d!a_d \notin \mathbb{Q}$. Définissons maintenant $f : \mathbb{T}^d \to \mathbb{T}^d$ par

$$f(y_1, \ldots, y_d) = (y_1 + \alpha, y_2 + y_1, \ldots y_d + y_{d-1}).$$

Alors comme $P_1(X+1) = P_1(X) + \alpha$ et pour tout $i \geqslant 1$, $P_i(X) + P_{i+1}(X) = P_{i+1}(X+1)$ on déduit que pour tout $n \geqslant 1$ et tout $x \in \mathbb{R}$ on a par récurrence immédiate

$$f^n(P_1(x), \ldots, P_d(x)) = (P_1(x+n), \ldots, P_d(x+n)).$$

D'après le Théorème 4.2 f est uniquement ergodique donc $(f^n(0))$ est équidistribuée, ainsi $(P_d(n))$ est équidistribuée, comme annoncé.

Dans le cas général notons e le plus grand indice tel que a_e est irrationnel. Si N est un multiple commun aux dénominateurs de $a_d, \ldots, a_{e+1}$, on a pour tous entiers $k \geqslant 0$ et $0 \leqslant r \leqslant N-1$

$$a_d(kN+r)^d + \cdots + a_{e+1}(kN+r)^{e+1} = A(k,r) + \sum_{j=q+1}^{d} a_j r^j = A(k,r) + c(r),$$

où $A(k,r)$ est un entier et $c(r)$ un rationnel ne dépendant que de r. Ainsi on a $P(kN+r) = A(k,r) + c(r) + Q_r(k)$, où $Q_r(X)$ est un polynôme de degré e et de coefficient dominant irrationnel. D'après la première étape de la preuve, pour tout $0 \leqslant r \leqslant N-1$ fixé, $(Q_r(k))_{k\geqslant 0}$ est équidistribuée modulo 1, ainsi il en est de même pour $(P(kN+r))_{k\geqslant 0}$, et finalement on conclut que $(P(n))_{n\geqslant 0}$ est équidistribuée. □

5.3. *Remarque.* — La démonstration classique de ce théorème consiste à utiliser le lemme suivant, combiné comme ci-dessus à une récurrence sur le degré de P:

Lemme de Van der Corput. — Soit (x_n) une suite réelle. Si pour tout $m \geqslant 1$, la suite $(x_{n+m} - x_n)$ est équidistribuée modulo 1, alors (x_n) est équidistribuée modulo 1.

Ce lemme peut par exemple se déduire du Lemme 4.3 du chapitre III, appliqué à $H = \mathbb{C}$ muni de sa structure hilbertienne standard, et à la suite $u_n = e^{2i\pi h x_n}$, où $h \in \mathbb{N}^*$. On montre ainsi que pour tout entier $h \geqslant 1$, la suite $\frac{1}{N}\sum_{n=1}^{N} e^{2i\pi h x_n}$ tend vers 0, ce qui implique que la suite (x_n) est équidistribuée (critère de Weyl).

◇

VI. COCYCLES LINÉAIRES ET THÉORÈME D'OSELEDETS

Ce chapitre inaugure une nouvelle partie du cours dans lequel le thème principal sera l'étude de produits de matrices $A_n \cdots A_1$, qui peuvent être des matrices aléatoires, ou plus généralement des matrices de la forme $A(f^n(x))$, où f est un certain système dynamique. Si les matrices sont de taille 1 (et disons de déterminant strictement positif), en passant au logarithme on retrouve respectivement les sommes de variables aléatoires et les sommes de Birkhoff. On sait que dans ce cas il y a des théorèmes limites pour ces sommes (loi des grands nombres, théorème de Birkhoff, et leurs raffinements éventuels). L'extension au cas de la dimension d arbitraire et l'obtention de "théorèmes limites" pour ces produits nécessite un certain nombre d'idées nouvelles qui font intervenir la géométrie de $\mathbb{R}^d$. En pratique nous nous limiterons essentiellement au cas de la dimension $d = 2$ où ces idées apparaissent déjà, mais qui est techniquement plus simple.

1. Définitions

1.1. — Soit $(X, \mathcal{A}, f, \mu)$ un système dynamique mesurable et $A : X \to \mathrm{GL}_d(\mathbb{R})$ une application mesurable. On peut alors faire une extension de f par A:

$$\begin{aligned} F : X \times \mathbb{R}^d &\longrightarrow X \times \mathbb{R}^d \\ (x, v) &\longmapsto (f(x), A(x)v) \end{aligned}$$

qui définit un système dynamique sur $X \times \mathbb{R}^d$. Les itérées de cette transformation s'expriment sous la forme:

$$F^n(x, v) = (f^n(x), A(f^{n-1}(x)) \cdots A(x)v) =: (f^n(x), A^{(n)}(x)v),$$

où la suite $A^{(n)}$ vérifie la relation

$$A^{(n+m)}(x) = A^{(n)}(f^m(x))A^{(m)}(x). \tag{1}$$

On dit qu'elle forme un **cocycle multiplicatif**.

1.2. — Si f est inversible alors F aussi et on a

$$F^{-1}(x, v) = (f^{-1}(x), (A(f^{-1}(x)))^{-1}v).$$

Ainsi pour $n \geqslant 1$ on peut écrire $F^{-n}(x, v) = \big(f^{-n}(x), A^{(-n)}(x)v\big)$, où

$$A^{(-n)}(x) = A(f^{-n}(x))^{-1} \cdots A(f^{-1}(x))^{-1} = \big(A^{(n)}(f^{-n}(x))\big)^{-1}.$$

On montre alors (exercice!) que la relation de cocycle (1) s'étend à $n \in \mathbb{Z}$.

1.3. — Posons $\mathrm{SL}_d^{\pm}(\mathbb{R})$ le groupe des matrices $d \times d$ de déterminant ± 1. Pour toute matrice $A \in \mathrm{GL}_d(\mathbb{R})$ on peut écrire $A = |\det(A)|^{1/d}\,\widetilde{A}$, avec $\widetilde{A} \in \mathrm{SL}_d^{\pm}(\mathbb{R})$, et donc

$$A^{(n)}(x) = \left(\prod_{i=0}^{d-1} \left|\det(A(f^i(x)))\right|^{1/d}\right) \widetilde{A}^{(n)}(x) = \exp\left(\sum_{i=0}^{d-1} \log\left|\det(A(f^i(x)))\right|^{1/d}\right) \widetilde{A}^{(n)}(x).$$

Le terme dans l'exponentielle est une somme de Birkhoff, donc on connait son comportement asymptotique. Ainsi dans l'étude des cocycles $A^{(n)}(\cdot)$ on peut se ramener au cas où $A(\cdot)$ est à valeurs dans $\mathrm{SL}_d^{\pm}(\mathbb{R})$. (On pourrait se ramener à $\mathrm{SL}_d(\mathbb{R})$, au prix d'utiliser des matrices à coefficients complexes, ce qui complique un peu la géométrie.)

1.4. *Définition. — Soit $(X, \mathcal{A}, f, \mu)$ un système dynamique mesurable et $A : X \to \mathrm{GL}_d(\mathbb{R})$ une application mesurable comme ci-dessus. Fixons une norme $\|\cdot\|$ sur $\mathrm{M}_d(\mathbb{R})$ et supposons que $\log^+ \|A\|(\cdot) \in L^1(X, \mu)$. L'**exposant de Lyapunov** dominant moyen est par définition*

$$\chi^+ = \lim_{n\to\infty} \frac{1}{n} \int \log\left\|A^{(n)}(x)\right\| \mathrm{d}\mu(x).$$

L'existence de cette limite nécessite quelques lignes de justification. D'abord, comme toutes les normes sont équivalentes on voit que l'existence de la limite et sa valeur ne dépendent pas de la norme choisie. Nous pouvons donc prendre pour $\|\cdot\|$ une norme d'opérateur, c'est à dire satisfaisant pour tous A, B $\|AB\| \leqslant \|A\| \cdot \|B\|$. Alors si on pose $\widetilde{\chi}_n = \int \log\left\|A^{(n)}(x)\right\| \mathrm{d}\mu(x)$, on voit que la suite $\widetilde{\chi}_n$ est sous-additive. En effet:

$$\begin{aligned}
\widetilde{\chi}_{n+m} &= \int \log\left\|A^{(n+m)}(x)\right\| \mathrm{d}\mu(x) = \int \log\left\|A^{(n)}(f^m(x))A^{(m)}(x)\right\| \mathrm{d}\mu(x)\\
&\leqslant \int \log\left\|A^{(n)}(f^m(x))\right\| \mathrm{d}\mu(x) + \int \log\left\|A^{(m)}(x)\right\| \mathrm{d}\mu(x)\\
&= \int \log\left\|A^{(n)}(x)\right\| \mathrm{d}\mu(x) + \int \log\left\|A^{(m)}(x)\right\| \mathrm{d}\mu(x) \text{ par invariance de } \mu\\
&= \widetilde{\chi}_n + \widetilde{\chi}_m.
\end{aligned}$$

Par sous-additivité, on conclut que la suite $\chi_n = \frac{1}{n}\widetilde{\chi}_n$ converge. □

1.5. — Plus généralement $(n, x) \mapsto \log\left\|A^{(n)}(x)\right\|$ définit un cocycle sous-additif, et donc le théorème de Kingman implique:

Proposition-Définition. — Si $(X, \mathcal{A}, f, \mu)$ est un système dynamique mesurable et $A : X \to \mathrm{GL}_d(\mathbb{R})$ une application borélienne telle que $\log^+ \|A\|(\cdot) \in L^1(X, \mu)$ alors pour presque tout x la suite $\frac{1}{n}\log\left\|A^{(n)}(x)\right\|$ converge vers un réel $\chi^+(x)$ qui est par définition l'exposant de Lyapunov dominant au point x.

1.6. — Si $\|\cdot\|$ est la norme d'opérateur associée à une certaine norme sur $\mathbb{R}^d$, la norme d'une matrice A, décrit le taux de dilatation maximal d'un vecteur unitaire. Le taux de contraction maximal est lui décrit par $\|A^{-1}\|$. Si $\log\|A(\cdot)^{-1}\| \in L^1$ on définira des exposants de Lyapunov "inférieurs"

$$\chi^- = \lim_{n\to\infty} \frac{1}{n} \int \log\left\|A^{(n)}(x)^{-1}\right\|^{-1} \mathrm{d}\mu(x) \text{ et } \chi^-(x) = \lim_{n\to\infty} \frac{1}{n} \log\left\|A^{(n)}(x)^{-1}\right\|^{-1}.$$

Exercice. — Montrer que si $\log\|A(\cdot)\| \in L^1$ et si A est à valeurs dans $\mathrm{SL}_d^{\pm}(\mathbb{R})$ on a $\chi^- = -\chi^+$. Donner une relation entre χ^+, χ^- et $\log|\det A|$ dans le cas général.

1.7. — Plus généralement, étant donné $(X, \mathcal{A}, f, \mu)$ un système dynamique mesuré, si G est un groupe (topologique) agissant sur un ensemble M et si l'on se donne $g : X \to G$ borélienne, alors comme précédemment on peut définir un système dynamique sur $X \times M$

$$F : (x, p) \longmapsto (f(x), g(x) \cdot p).$$

On voit que $F^n(x, p) = (f^n(x), g^{(n)} \cdot p)$ où $g^{(n)}(x) = g(f^{n-1}(x)) \cdots g(x)$ est un produit d'éléments dans le groupe. On dit que $(g^{(n)})$ est un cocycle à valeurs dans G.

Un cas qui sera important dans la suite est le suivant. On rappelle que l'**espace projectif** $\mathbb{P}^{d-1}(\mathbb{R})$ est l'espace des droites vectorielles de $\mathbb{R}^d$, plus précisément c'est le quotient de $\mathbb{R}^d \backslash \{0\}$ par la relation de colinéarité. Il est naturellement muni d'une topologie qui le rend compact (et c'est également une variété de dimension $d-1$). On notera $[v]$ la classe d'un vecteur v dans $\mathbb{P}^{d-1}(\mathbb{R})$. Si (f, A) est un cocycle linéaire à valeurs dans $\mathrm{GL}_d(\mathbb{R})$, alors on a une action induite sur $\mathbb{P}^{d-1}(\mathbb{R})$ donnée par

$$(x, p) \longmapsto (f(x), [A(x)v]) \text{ où } [v] = p.$$

Noter que A et $\widetilde{A}$ (où l'on a extrait le déterminant) ont la même action projective.

Exercice. — Montrer que $\mathbb{P}^1(\mathbb{R})$ est homéomorphe à un cercle. Quel est le lien entre $\mathbb{P}^1(\mathbb{R})$ et le cercle unité de $\mathbb{R}^2$?

Remarque pour les lecteurs mal à l'aise avec les espaces projectifs. — On pourrait travailler avec la sphère unité de $\mathbb{R}^d$ plutôt qu'avec le projectif mais c'est un peu moins intrinsèque. Surtout pour étudier les cocycles linéaires en dimension $d \geqslant 3$, il faut plus généralement considérer pour tout $1 \leqslant k \leqslant d-1$ l'action induite par (f, A) sur l'espace des sous-espaces vectoriels de dimension k, et l'analogue de la sphère est moins naturel dans ce cas...

2. Exemples

2.1. — **Exemple trivial.** Si (f, A) est un cocycle constant, alors χ^+ est le logarithme du rayon spectral de A et χ^- est l'opposé de celui de A^{-1}.

2.2. — **Produits aléatoires de matrices.** On se donne μ une mesure de probabilité sur $\mathrm{GL}_d(\mathbb{R})$ et on pose $\Omega = \mathrm{GL}_d(\mathbb{R})^{\mathbb{N}^*}$ et $\mathbb{P} = \mu^{\mathbb{N}^*}$. On considère le décalage $\sigma : (A_n)_{n \geqslant 1} \mapsto (A_{n+1})_{n \geqslant 1}$ et on définit $\pi : \Omega \to \mathrm{GL}_d(\mathbb{R})$ par $\pi(\omega) = A_1$, où $\omega = (A_n)_{n \geqslant 1}$. On a donc ainsi défini un cocycle (σ, π) et une transformation

$$\begin{aligned} F : \quad \Omega \times \mathbb{R}^d &\longrightarrow \Omega \times \mathbb{R}^d \\ ((A_n), v) &\longmapsto ((A_{n+1}), A_1 v). \end{aligned}$$

La loi de $\sigma^n(\omega)$ correspond à celle d'une suite i.i.d. de matrices de loi μ et $\pi^{(n)}(\omega) = A_n \cdots A_1$ est le produit à gauche de ces matrices aléatoires. Le théorème suivant est un analogue de la loi forte des grands nombres pour les produits $A_n \cdots A_1$. C'est en fait un cas particulier du §1.5 et a été établi avant celui de Kingman, par Furstenberg et Kesten:

Théorème. — Sous les hypothèses précédentes, si $\mathbb{E}(\log^+ \|A\|) < \infty$*, alors presque sûrement la suite* $\frac{1}{n} \log \|A_n \cdots A_1\|$ *converge vers un réel* $\chi^+(\mu)$.

Dans le chapitre suivant nous nous intéresserons aux propriétés de l'exposant de Lyapunov $\chi^+(\mu)$.

2.3. — Cocycle tangent. Si $f : M \to M$ est un difféomorphisme sur une variété M, sa différentielle agit sur le fibré tangent TM par

$$\begin{aligned} Df : \quad & TM \longrightarrow TM \\ & (x, v) \longmapsto (f(x), Df_x(v)). \end{aligned}$$

Le théorème de dérivation des fonctions composées dit que $D(f^n)_x v = D_{f^{n-1}(x)} \circ \cdots \circ Df_x(v)$, on est donc bien en présence d'un cocycle multiplicatif. Les propriétés de ce cocycle jouent un rôle important dans la compréhension de la dynamique de f.

Remarque. — Une différence avec le cadre présenté au §1 est que TM n'est en général pas un produit. Néanmoins dans la plupart des cas il est facile de construire un ensemble invariant de mesure totale $\Omega \subset M$ tel que $TM|_\Omega \simeq \Omega \times \mathbb{R}^d$: il suffit de faire des "coupures" sur M pour la rendre contractile. Par ailleurs le lecteur peu familier avec le formalisme des variétés pourra considérer le cas déjà riche où $M = \mathbb{R}^d$, et dans ce cas on a $TM \simeq \mathbb{R}^d \times \mathbb{R}^d$. Un autre cas intéressant est $M = \mathbb{T}^d$ que nous avons considéré précédemment et pour lequel TM est trivial, i.e. $TM \simeq M \times \mathbb{R}^d$.

Remarque bis. — Il sera utile dans la suite de voir $X \times \mathbb{R}^d$ comme un "fibré trivial de base X" (ce qui est effectivement le cas), et en particulier à $\{x\} \times \mathbb{R}^d$ comme à un **espace vectoriel dépendant de** x. Nous noterons $\mathbb{R}^d_x := \{x\} \times \mathbb{R}^d$, de sorte que F envoie $\mathbb{R}^d_x$ sur $\mathbb{R}^d_{f(x)}$.

3. Cocycles hyperboliques (en dimension 2)

Pour gagner un peu d'intuition sur la manière dont un produit de matrices agit sur l'espace il est intéressant d'étudier la notion d'**hyperbolicité**, introduite (pour le cocycle tangent) en dynamique par Smale(1), et qui joue un rôle considérable en dynamique différentiable.

Dans cette section on se donne $f : X \to X$ un homéomorphisme d'un espace métrique compact et $A : X \to \mathrm{GL}_2(\mathbb{R})$ continue: nous disons alors que (f, A) est un **cocycle topologique**. Remarquer que les exemples de la section précédente entrent dans cette catégorie. On se donne une norme quelconque $\|\cdot\|$ sur $\mathbb{R}^2$.

3.1. *Définition. — On dit que le cocycle* $(A^{(n)})$ *est hyperbolique s'il existe* $C > 0$ *et* $0 < \sigma < 1$ *et pour tout* $x \in X$ *deux droites* E^s_x *et* E^u_x *transverses dans* $\mathbb{R}^2_x$ *(i.e.* $E^s_x \oplus E^u_x = \mathbb{R}^2_x$*) telles que:*

(1)On pourra lire le classique *"Finding a horseshoe on the beaches of Rio"* (Math. Intelligencer 20 (1998) pp. 39-44), dans lequel Smale relate (avec une gourmandise certaine) les débuts de cette théorie.

1. *la famille des sous espaces $E_x^{s/u}$ est invariante au sens où pour tout $x \in X$ on a $A(x)E_x^s = E_{f(x)}^s$ et $A(x)E_x^u = E_{f(x)}^u$;*
2. *$\forall x \in X$, $\forall n \geqslant 1$, $\forall v^s \in E^s(x)$, $\left\|A^{(n)}(x)v^s\right\| \leqslant C\sigma^n \|v^s\|$;*
3. *$\forall x \in X$, $\forall n \geqslant 1$, $\forall v^u \in E^u(x)$, $\left\|A^{(-n)}(x)v^u\right\| \leqslant C\sigma^n \|v^u\|$.*

On appelle E_x^s l'**espace stable** et E_x^u l'**espace instable** en x. Noter que si $v^u \in E_x^u$ on a pour tout $n \geqslant 1$, $\left\|A^{(n)}(x)v^u\right\| \geqslant C^{-1}\sigma^{-n} \|v^u\|$: il suffit d'appliquer le 3. à $w = A^{(n)}(x)v^u \in E_{f^n(x)}^u$. On a donc dilatation dans une direction et contraction dans une autre direction, à vitesse exponentielle uniforme.

3.2. *Lemme. — Les droites E_x^s et E_x^u varient continûment avec x.*

Démonstration. — L'observation fondamentale est que si $v \in \mathbb{R}_x^2$ et $v \notin E_x^s$ alors $\left\|A^{(n)}(x)v\right\| \to \infty$: il suffit en effet pour cela de décomposer v selon $\mathbb{R}_x^2 = E_x^s \oplus E_x^u$. On a ainsi une caractérisation dynamique de E_x^s: c'est l'ensemble des vecteurs tels que $\left\|A^{(n)}(x)v\right\|$ tend vers 0 (ou encore ne tend pas vers l'infini).

Soit maintenant (x_i) une suite tendant vers x et pour chaque i un vecteur unitaire $e_{x_i}^s \in E_{x_i}^s$. Pour tout n on a $\left\|A^{(n)}(x_i)e_{x_i}^s\right\| \leqslant C\sigma^n$. Si e_x est une valeur d'adhérence de $(e_{x_i}^s)$ on aura par continuité de A que pour tout n, $\left\|A^{(n)}(x)e_x\right\| \leqslant C\sigma^n$, et donc par la remarque précédente e_x appartient à E_x^s, d'où la continuité de $x \mapsto E_x^s$. On procède de même pour E^u. □

3.3. *Exercice.* — Montrer que les applications $x \mapsto E_x^s$ et $x \mapsto E_x^u$ sont hölderiennes.

3.4. *Remarque importante.* — La définition 3.1 ne dépend pas du choix de la norme $\|\cdot\|$ sur $\mathbb{R}^2$. Beaucoup mieux: par analogie avec la géométrie riemannienne on peut en fait fixer une norme $\|\cdot\|_x$ dans chaque $\mathbb{R}_x^2$ et pour $v \in \mathbb{R}_x^2$ calculer la norme de v relativement à $\|\cdot\|_x$ et demander que les conditions 2 et 3 soient vraies dans ce cadre. Si $\|\cdot\|_x$ et $\|\cdot\|_x'$ sont des normes sur $\mathbb{R}_x^2$ telles que $x \mapsto \|\cdot\|_x$ et $x \mapsto \|\cdot\|_x'$ sont continues alors par compacité il existe une constante $c > 0$ telle que pour tout x, $c\|\cdot\|_x \leqslant \|\cdot\|_x' \leqslant c^{-1}\|\cdot\|_x$, et donc cette notion généralisée d'hyperbolicité est équivalente à la précédente.

3.5. — Nous appellerons **cône** (fermé) dans $\mathbb{R}^2$ une partie de la forme

$$\left\{v \in \mathbb{R}^2,\ v = \alpha e_1 + \beta e_2, |\alpha| \leqslant |\beta|\right\},$$

où (e_1, e_2) est une base de $\mathbb{R}^2$. Voici une condition nécessaire et suffisante utile pour l'hyperbolicité.

3.6. *Théorème. — Un cocycle topologique (f, A) est hyperbolique si et seulement s'il existe des familles continues de cônes $(K_x^s)_{x\in X}$ et $(K_x^u)_{x\in X}$ et de normes $(\|\cdot\|_x)_{x\in X}$ telles que:*

1. *pour tout $x \in X$, on a $A(x)K_x^u \subset \mathrm{Int}(K_{f(x)}^u) \cup \{0\}$ et $A(f(x))^{-1}K_{f(x)}^s \subset \mathrm{Int}(K_x^s) \cup \{0\}$;*
2. *pour tout $x \in X$, pour tout $v \in K_x^u \backslash \{0\}$, on a $\|A(x)v\|_{f(x)} > \|v\|_x$ et pour tout $v \in K_x^s \backslash \{0\}$ on a $\|A(x)v\|_{f(x)} < \|v\|_x$.*

Démonstration. — $\Leftarrow$. On doit trouver les espaces E_x^s et E_x^u. Par simplicité nous noterons simplement $\|\cdot\|$ les normes. On remarque d'abord que par compacité et continuité du champ de cônes il existe $\lambda > 1$ tel que pour tout $v \in K_x^u$, on a $\|A(x)v\| \geqslant \lambda \|v\|$ et pour tout $v \in K_x^s$ on a $\|A(x)v\| \leqslant \lambda^{-1} \|v\|$. Posons alors

$$E_x^s = \bigcap_{n \geqslant 0} \left(A^{(n)}(x)\right)^{-1} \left(K_{f^n(x)}^s\right) \text{ et } E_x^u = \bigcap_{n \geqslant 0} A^{(n)}(f^{-n}(x)) \left(K_{f^{-n}(x)}^u\right).$$

Chacun de E_x^s et E_x^u est l'intersection décroissante d'une famille de fermés de l'espace des droites de $\mathbb{R}^2$, donc non vide et réunion fermée de droites. Les relations d'invariance $A(x)E_x^s = E_{f(x)}^s$ et $A(x)E_x^u = E_{f(x)}^u$ sont immédiates. En outre, si $v \in E_x^s$ on a pour tout $n \geqslant 1$ $A^{(n)}(x)v \in K_{f^n(x)}^s$, donc par une récurrence immédiate il vient $\left\|A^{(n)}(x)v\right\| \leqslant \lambda^{-n} \|v\|$ et de même pour $v \in E_x^u$ $\left\|A^{(-n)}(x)v\right\| \leqslant \lambda^{-n} \|v\|$. Par invariance de E^u on en déduit que $\left\|A^{(n)}(x)v\right\| \geqslant \lambda^n \|v\|$. On voit maintenant que E_x^s est une droite car si E_x^s contenait deux droites distinctes, on aurait que pour tout $v \in \mathbb{R}_x^2$, $\left\|A^{(n)}(x)v\right\| \leqslant \lambda^{-n} \|v\|$ ce qui n'est pas le cas car E_x^u est non trivial. En inversant les rôles de E^u et E^s et en considérant f^{-1}, on en déduit le même résultat pour E_x^u. Finalement E_x^s et E_x^u sont transverses car un vecteur non nul commun à ces deux droites devrait à la fois être dilaté et contracté exponentiellement par itération.

$\Rightarrow$. L'idée est de jouer avec les normes sur $X \times \mathbb{R}^2$ pour faire apparaître naturellement la condition 2. Supposons dans un premier temps que, la norme $\|\cdot\|$ étant donnée sur $\mathbb{R}^2$, l'expansion/contraction dans la définition 3.1 arrive dès la première itération, i.e. $C = 1$. Écrivons la décomposition $\mathbb{R}_x^2 = E_x^s \oplus E_x^u$ et fixons des vecteurs unitaires e_x^s et e_x^u respectifs de E_x^s et E_x^u (pour la norme $\|\cdot\|$). Pour tout $v \in \mathbb{R}_x^2$ on peut écrire $v = \alpha e_x^s + \beta e_x^u$. Posons $\|v\|_x = \max(|\alpha|, |\beta|)$: c'est une norme sur $\mathbb{R}_x^2$. Posons également

$$K_x^s = \{v = \alpha e_x^s + \beta e_x^u,\ |\alpha| \geqslant 2|\beta|\} \text{ et } K_x^u = \{v = \alpha e_x^s + \beta e_x^u,\ |\beta| \geqslant 2|\alpha|\},$$

qui sont des cônes fermés. Par définition on a $\|A(x)e_x^s\|_{f(x)} \leqslant \sigma$ et $\left\|A(f(x))^{-1}e_{f(x)}^u\right\|_x \leqslant \sigma$ et ainsi si $v = \alpha e_x^s + \beta e_x^u$ on a $A(x)v = \alpha' e_{f(x)}^s + \beta' e_{f(x)}^u$ avec $|\alpha'| \leqslant \sigma |\alpha|$ et $|\beta'| \geqslant \sigma^{-1} |\beta|$. On en déduit alors immédiatement la propriété 1. ainsi que $\|A(x)v\|_{f(x)} > \|v\|_x$ pour tout $v \in K_x^u \backslash \{0\}$ ainsi que $\|A(x)v\|_{f(x)} < \|v\|_x$ pour tout $v \in K_x^s \backslash \{0\}$, d'où le résultat.

Dans le cas général où C est quelconque, nous allons modifier la norme de départ pour arriver à $C = 1$ et appliquer le raisonnement précédent. Fixons $\varepsilon > 0$ petit devant σ. Si (f, A) vérifie la définition 3.1 pour $v \in E^s(x)$ on pose

$$\|v\|_x = \sum_{n=0}^{\infty} (\sigma + \varepsilon)^{-n} \left\|A^{(n)}(x)v\right\| \tag{2}$$

(la série converge car $\left\|A^{(n)}(x)v\right\| \leqslant C\sigma^n$). Alors $A(x)v$ appartient à $E^s_{f(x)}$ et on a

$$\begin{aligned}
\|A(x)v\|_{f(x)} &= \sum_{n=0}^{\infty}(\sigma+\varepsilon)^{-n}\left\|A^{(n)}(f(x))A(x)v\right\| \\
&= \sum_{n=0}^{\infty}(\sigma+\varepsilon)^{-n}\left\|A^{(n+1)}(x)v\right\| \\
&= (\sigma+\varepsilon)\sum_{n=0}^{\infty}(\sigma+\varepsilon)^{-(n+1)}\left\|A^{(n+1)}(x)v\right\| \\
&= (\sigma+\varepsilon)\left(\|v\|_x - \|v\|\right) \leqslant (\sigma+\varepsilon)\|v\|_x .
\end{aligned}$$

De même on pose pour $v \in E^u_x$

$$\|v\|_x = \sum_{n=0}^{\infty}(\sigma+\varepsilon)^{-n}\left\|A^{(-n)}(x)v\right\|, \tag{3}$$

et on montre que $\|(A(f^{-1}(x)))^{-1}v\|_{f^{-1}(x)} \leqslant (\sigma+\varepsilon)\|v\|_x$. Si maintenant $v \in \mathbb{R}^2_x$ est quelconque on le décompose sous la forme $v = v^s + v^u$ et on pose $\|v\|_x = \max(\|v^s\|_x, \|v^u\|_x)$. Par convergence uniforme dans (2) et (3) on a ainsi défini une famille continue de normes sur les $\mathbb{R}^2_x$ et maintenant on vérifie immédiatement que pour ce choix de norme dans la définition 3.1 on a $C = 1$. On peut alors appliquer le raisonnement de la première partie de la preuve pour conclure. □

3.7. *Corollaire. — Si (f, A) est un cocycle hyperbolique, alors tout cocycle (f', A') suffisamment (uniformément) proche de (f, A) est hyperbolique.*

En effet la condition de cône dans le théorème est clairement stable par perturbation de A.

3.8. — Si $f : M \to M$ est un difféomorphisme C^1 d'une surface compacte, on dit que f a la **propriété d'Anosov** si son cocycle tangent est hyperbolique. L'application f_M induite par la matrice $M = \left(\begin{smallmatrix}2&1\\1&1\end{smallmatrix}\right)$ sur $\mathbb{T}^2$ (i.e. la transformation du chat d'Arnol'd) est d'Anosov (exercice). Ainsi le corollaire précédent montre que toute perturbation f (en topologie C^1) de f_M est également Anosov. C'est une étape importante dans la preuve du fait que la dynamique de f est qualitativement la même que celle de f_M (propriété dite de **stabilité structurelle**).

4. Théorème d'Oseledets en dimension 2

Le théorème d'Oseledets, dit "théorème ergodique multiplicatif" est un complément important au Théorème 1.5 puisqu'il explique le comportement de $A^{(n)}(x)v$ pour **tout** vecteur $v \in \mathbb{R}^2_x$.

4.1. — **Préliminaire d'algèbre linéaire: valeurs singulières.** Soit $A \in \mathrm{GL}_2(\mathbb{R})$ et examinons l'action métrique de A sur $\mathbb{R}^2$ muni de sa métrique euclidienne. Nous noterons par le même symbole $\|\cdot\|$ la norme euclidienne sur $\mathbb{R}^2$ et la norme matricielle associée, ce qui ne devrait pas créer de confusion. On rappelle que $\|A\| = \sqrt{\rho(A^*A)}$ où ρ est le rayon spectral et A^* l'adjoint (i.e. la transposée).

Proposition. — Il existe des vecteurs unitaires orthogonaux e^+ et e^- respectivement les plus dilatés et les plus contractés par A, i.e. $\|Ae^+\| = \|A\|$ et $\|Ae^-\| = \|A^{-1}\|^{-1}$.

La démonstration de cette proposition repose sur la **décomposition KAK de Cartan**: il existe des matrices orthogonales U et V telles que $A = U\Sigma V$, avec $\Sigma = \operatorname{diag}(\|A\|, \|A^{-1}\|^{-1})$. Pour cela, remarquons déjà que si l'on peut montrer que $A = U\Sigma V$ avec U et V orthogonales et Σ diagonale, $\Sigma = \operatorname{diag}(\sigma_1, \sigma_2)$ avec $\sigma_1 \geqslant \sigma_2$ alors nécessairement $\sigma_1 = \|A\|$ et $\sigma_2 = \|A^{-1}\|^{-1}$, et on a $e^+ = V^{-1}(e_1)$ et $e^- = V^{-1}(e_2)$. En effet le résultat est évident pour $A = \Sigma$ en analysant l'action métrique de Σ sur $\mathbb{R}^2$ et ensuite il est invariant par composition à gauche et à droite par une isométrie. Les nombres σ_1 et σ_2 sont les **valeurs singulières** de A.

Ensuite, pour démontrer l'existence d'une décomposition $A = U\Sigma V$, on commence par effectuer la **décomposition polaire** $A = OS$, avec O orthogonale et S symétrique définie positive: il suffit de poser $S = \sqrt{A^*A}$ et de remarquer que AS^{-1} est orthogonale. Ensuite on diagonalise S en base orthonormale: $S = V^*\Sigma V$ avec V orthogonale, qui donne $A = OV^*\Sigma V$ et il suffit de poser $U = OV^*$. □

4.2. *Remarque.* — Tout ceci vaut également en dimension arbitraire. On obtient alors une suite de valeurs singulières $\sigma_1 \geqslant \cdots \geqslant \sigma_d$ avec $\sigma_1 = \|A\|$ et $\sigma_d = \|A^{-1}\|^{-1}$. Géométriquement, on interprète les valeurs singulières comme les demi-axes de l'ellipsoïde image de la boule unité par A.

4.3. *Remarque-Exercice.* — Si $A \in \mathrm{SL}_2^{\pm}$, on voit simplement que $\|A\| = \|A^{-1}\| \geqslant 1$.

Si $(X, \mathcal{A}, f, \mu)$ est un système dynamique mesurable et $A : X \to \mathrm{GL}_2(\mathbb{R})$ borélienne, on a vu que si $\log^+\|A\|$ et $\log^+\|A^{-1}\|$ sont intégrables on peut définir pour presque tout x

$$\chi^+(x) = \lim_{n\to\infty} \frac{1}{n} \log \left\|A^{(n)}(x)\right\| \text{ et } \chi^-(x) = \lim_{n\to\infty} \frac{1}{n} \log \left\|A^{(n)}(x)^{-1}\right\|^{-1}.$$

Si μ est ergodique les exposants χ^+ et χ^- ne dépendent pas de x.

4.4. *Théorème ergodique multiplicatif d'Oseledets. — Soit $(X, \mathcal{A}, f, \mu)$ un système dynamique mesurable et $A : X \to \mathrm{GL}_2(\mathbb{R})$ borélienne. On suppose que les fonctions $\log^+\|A\|$ et $\log^+\|A^{-1}\|$ sont intégrables et on note $\chi^+(x)$ et $\chi^-(x)$ les exposants de Lyapunov supérieur et inférieur en x comme ci-dessus. Alors pour μ-presque tout x on a:*

1. *Soit $\chi^+(x) = \chi^-(x)$ et alors*

$$\text{pour tout } v \in \mathbb{R}^2_x,\ \lim_{n\to\infty} \frac{1}{n} \log \left\|A^{(n)}(x)v\right\| = \chi^{\pm}(x).$$

2. *Soit $\chi^+(x) > \chi^-(x)$ et il existe une droite $E^s_x \in \mathbb{R}^2_x$ telle que*

$$\text{pour tout } v \in E^s_x \backslash \{0\} \text{ on a } \lim_{n\to\infty} \frac{1}{n} \log \left\|A^{(n)}(x)v\right\| = \chi^-(x)$$

et

$$\text{pour tout } v \in \mathbb{R}^2_x \backslash E^s_x \text{ on a } \lim_{n\to\infty} \frac{1}{n} \log \left\|A^{(n)}(x)v\right\| = \chi^+(x).$$

En outre $x \mapsto E^s_x$ est mesurable et invariante: $A(x)E^s_x = E^s_{f(x)}$ p.s.

Démonstration. — Remarquer que les hypothèses du théorème impliquent que $\log|\det A(\cdot)|$ est intégrable. Ainsi en extrayant le déterminant il suffit de traiter le cas où A est à valeurs dans $\mathrm{SL}_2^{\pm}(\mathbb{R})$ (l'adaptation au cas général est laissée en exercice pour le lecteur). Fixons un point μ-générique x (on se permettra d'extraire à nouveau dans la preuve un sous-ensemble de mesure totale). Dans ce cas comme $\|A^{(n)}(x)\| = \|A^{(n)}(x)^{-1}\|$ on voit que $\chi^-(x)$ existe si et seulement si $\chi^+(x)$ existe et $\chi^-(x) = -\chi^+(x)$. En outre comme $\|A^{(n)}(x)\| \geqslant 1$ on a $\chi^+(x) \geqslant 0$.

Premier cas: $\chi^+(x) = \chi^-(x) = 0$. Alors dans ce cas par définition de la norme d'opérateur on a pour tout $v \in \mathbb{R}^2$

$$\left\|A^{(n)}(x)^{-1}\right\|^{-1} \cdot \|v\| \leqslant \left\|A^{(n)}(x)v\right\| \leqslant \left\|A^{(n)}(x)\right\| \cdot \|v\|$$

et donc clairement

$$\frac{1}{n}\log\left\|A^{(n)}(x)v\right\| \underset{n\to\infty}{\longrightarrow} 0.$$

Deuxième cas: $\chi^-(x) < 0 < \chi^+(x)$. Le point est de trouver l'espace E^s_x. Remarquons que $\|A^{(n)}(x)\|$ est de l'ordre de $\exp(n\chi^+(x))$. On pose $u_n(x)$ le vecteur unitaire le plus dilaté et $s_n(x)$ le vecteur unitaire le plus contracté, de sorte que

$$\left\|A^{(n)}(x)u_n\right\| = \left\|A^{(n)}(x)\right\|,\ \left\|A^{(n)}(x)s_n\right\| = \left\|A^{(n)}(x)^{-1}\right\|^{-1} = \left\|A^{(n)}(x)\right\|^{-1}, \text{ et } u_n \perp s_n.$$

Nous allons démontrer que la suite $(s_n(x))$ converge vers un vecteur unitaire $s(x)$ et on posera alors $E^s_x = \mathbb{R}s(s)$.

4.5. *Lemme. — L'angle $\angle(s_n(x), s_{n+1}(x))$ décroit exponentiellement vers 0. Plus précisément*

$$\limsup_{n\to\infty} \frac{1}{n}\log|\angle(s_n(x), s_{n+1}(x))| \leqslant -2\chi^+.$$

Démonstration. — Posons α_n cet angle. On a alors (au signe près car on ne sait pas si la base (s_{n+1}, u_{n+1}) est directe)

$$s_n(x) = (\sin\alpha_n)u_{n+1}(x) + (\cos\alpha_n)s_{n+1}(x)$$

et donc

$$\begin{aligned}
\left\|A^{(n+1)}(x)s_n(x)\right\| &= \left\|(\sin\alpha_n)A^{(n+1)}(x)u_{n+1}(x) + (\cos\alpha_n)A^{(n+1)}(x)s_{n+1}(x)\right\| \\
&\geqslant |\sin\alpha_n|\left\|A^{(n+1)}(x)u_{n+1}(x)\right\| - \left\|A^{(n+1)}(x)s_{n+1}(x)\right\| \\
&= |\sin\alpha_n|\left\|A^{(n+1)}(x)\right\| - \left\|A^{(n+1)}(x)^{-1}\right\|^{-1},
\end{aligned}$$

et par ailleurs

$$\left\|A^{(n+1)}(x)s_n(x)\right\| = \left\|A(f^n(x))A^{(n)}(x)s_n(x)\right\| \leqslant \|A(f^n(x))\| \cdot \left\|A^{(n)}(x)^{-1}\right\|^{-1}$$

d'où

$$|\sin\alpha_n| \leqslant \frac{\|A(f^n(x))\| \cdot \left\|A^{(n)}(x)^{-1}\right\|^{-1} + \left\|A^{(n+1)}(x)^{-1}\right\|^{-1}}{\left\|A^{(n+1)}(x)\right\|}.$$

Comme $\left\|A^{(n)}(x)^{-1}\right\|^{-1} \approx e^{-n\chi^+}$, de même que $\left\|A^{(n+1)}(x)^{-1}\right\|^{-1}$, et comme $\left\|A^{(n+1)}(x)\right\| \approx e^{n\chi^+}$, le lemme suivant appliqué à $\varphi = \log\|A(\cdot)\|$ conclut la preuve. $\square$

4.6. *Lemme. — Soit $(X, \mathcal{A}, f, \mu)$ un système dynamique mesurable et φ une fonction intégrable. Alors*

$$\lim_{n\to\infty} \frac{1}{n}\varphi(f^n(x)) = 0 \ p.p.$$

Démonstration. — Appliquer le théorème de Birkhoff à $\psi = \varphi \circ f - \varphi$. $\square$

Le lemme 4.5 implique que la suite de vecteurs unitaires $(s_n(x))$ est p.s. de Cauchy, elle converge donc vers une fonction mesurable $s(x)$. En outre, x étant fixé on a

$$\limsup_{n\to\infty} \frac{1}{n}\log|\angle(s(x), s_n(x))| \leqslant -2\chi^+.$$

4.7. *Lemme. — Le vecteur $(s(x))$ est exponentiellement contracté par $A^{(n)}(x)$ au taux $\chi^- = -\chi^+$, i.e.*

$$\lim_{n\to\infty} \frac{1}{n}\log\left\|A^{(n)}(x)s(x)\right\| = -\chi^+(x).$$

Démonstration. — Soit $\beta_n = \angle(s(x), s_n(x))$. On a $\limsup_{n\to\infty}\frac{1}{n}\log|\beta_n| \leqslant -2\chi^+$. On décompose comme précédemment

$$s(x) = (\cos\beta_n)s_n(x) + (\sin(\beta_n))u_n(x)$$

et donc

$$A^{(n)}(x)s(x) = \underbrace{(\cos\beta_n)}_{\approx 1}\underbrace{A^{(n)}(x)s_n(x)}_{\approx \exp(-n\chi^+)} + \underbrace{(\sin\beta_n)}_{\lessapprox \exp(-2n\chi^+)}\underbrace{A^{(n)}(x)u_n(x)}_{\approx \exp(n\chi^+)}$$

d'où l'on déduit que

$$\limsup_{n\to\infty} \frac{1}{n}\log\left\|A^{(n)}(x)s(x)\right\| = -\chi^+(x).$$

Comme la norme minimale de $A^{(n)}(x)$ est du même ordre de grandeur $e^{-n\chi^+(x)}$, on conclut. $\square$

4.8. *Lemme. — Si v n'est pas colinéaire à $s(x)$ on a*

$$\lim_{n\to\infty} \frac{1}{n}\log\left\|A^{(n)}(x)v\right\| = \chi^+(x).$$

Démonstration. — On applique toujours la même idée. Soit $\gamma_n = \angle(v, s_n(x))$, de sorte que $\gamma_n \to \gamma = \angle(v, s(x)) \neq 0$. On a alors

$$A^{(n)}(x)v = \underbrace{(\cos\gamma_n)A^{(n)}(x)s_n(x)}_{\lessapprox \exp(-n\chi^+)} + \underbrace{(\sin\gamma_n)}_{\sim \sin\gamma}\underbrace{A^{(n)}(x)u_n(x)}_{\approx \exp(n\chi^+)}$$

d'où le résultat. $\square$

On voit donc que $E^s_x = \mathbb{R}s(x)$ vérifie les conclusions du théorème. Il ne reste qu'à vérifier l'invariance $A(x)E^s_x = E^s_{f(x)}$: en effet on a $A^{(n)}(f(x)) \cdot A(x)s(x) = A^{(n+1)}(x)s(x)$ donc la norme de ce vecteur tend vers 0 et donc nécessairement $A(x)s(x)$ est colinéaire à $s(f(x))$ par le Lemme 4.8. □

Dans le cas inversible on a un énoncé qui rappelle la définition des cocycles hyperboliques.

4.9. *Théorème d'Oseledets dans le cas inversible.* — *Soit* $(X, \mathcal{A}, f, \mu)$ *un système dynamique mesurable inversible et* $A : X \to \mathrm{GL}_2(\mathbb{R})$ *une application borélienne. On suppose que les fonctions* $\log^+ \|A\|$ *et* $\log^+ \|A^{-1}\|$ *sont intégrables et on note* $\chi^+(x)$ *et* $\chi^-(x)$ *les exposants de Lyapunov supérieur et inférieur en* x *comme précédemment. Alors pour presque tout* x *on a:*

1. *Soit* $\chi^+(x) = \chi^-(x)$ *et alors*
$$\textit{pour tout } v \in \mathbb{R}^2_x,\ \lim_{n\to\pm\infty} \frac{1}{n} \log \left\| A^{(n)}(x)v \right\| = \chi^{\pm}(x).$$
2. *Soit* $\chi^+(x) > \chi^-(x)$ *et il existe une décomposition en somme directe* $\mathbb{R}^2_x = E^s_x \oplus E^u_x$ *telle que*

$$\lim_{n\to+\infty} \frac{1}{n} \log \left\| A^{(n)}(x)v \right\| = \begin{cases} \chi^-(x) & \textit{si } v \in E^s_x \backslash \{0\} \\ \chi^+(x) & \textit{si } x \notin E^s_x \end{cases}, \tag{4}$$

et

$$\lim_{n\to-\infty} \frac{1}{n} \log \left\| A^{(n)}(x)v \right\| = \begin{cases} \chi^+(x) & \textit{si } v \in E^u_x \backslash \{0\} \\ \chi^-(x) & \textit{si } x \notin E^u_x \end{cases}. \tag{5}$$

En outre les droites E^s_x *et* E^u_x *dépendent mesurablement de* x, *vérifient la relation d'invariance* $A(x)E^{u/s}_x = E^{u/s}_{f(x)}$ *et ne s'approchent pas trop l'une de l'autre au sens où p.s. on a*
$$\lim_{n\to\pm\infty} \frac{1}{n} \log \left| \angle \left(E^s_{f^n(x)}, E^u_{f^n(x)} \right) \right| = 0.$$

Remarquer que lorsque $n \to -\infty$, $\frac{1}{n} \log \left\| A^{(n)}(x)v \right\| \to \chi^-$ signifie que $\left\| A^{(n)}(x)v \right\| \approx \exp(-|n|\,\chi^-)$, donc si $\chi^- < 0$ (par exemple dans le cas SL_2) cela induit bien une croissance exponentielle.

Démonstration. — Comme précédemment, en extrayant le déterminant on peut supposer que A est à valeurs dans $\mathrm{SL}^{\pm}_2(\mathbb{R})$, ainsi on a pour presque tout x l'égalité $\chi^-(x) = -\chi^+(x)$. En utilisant la décomposition ergodique de μ, nous pouvons et allons supposer que μ est ergodique. Ainsi les exposants de Lyapunov ne dépendent pas de x.

On commence par appliquer le théorème d'Oseledets non inversible 4.4 séparément à f et f^{-1} qui fournit respectivement des familles mesurables de sous-espaces E^s_x et E^u_x. La propriété (4) est alors satisfaite, de même que (5), sauf que dans ce deuxième cas on ne connaît pas la valeur des exposants. Nous noterons donc $\widetilde{\chi}^{\pm}$ tels que p.s.

$$\lim_{n\to-\infty}\frac{1}{n}\log\left\|A^{(n)}(x)v\right\| = \begin{cases}\widetilde{\chi}^{-} \text{ si } v \in E_x^u\backslash\{0\} \\ \widetilde{\chi}^{+} \text{ si } x \notin E_x^u\end{cases}.$$

Il nous faut alors démontrer que:

- $\widetilde{\chi}^{+} = \chi^{+}$, qui donne automatiquement $\widetilde{\chi}^{-} = \chi^{-}$,
- p.s. on a $\mathbb{R}_x^2 = E_x^s \oplus E_x^u$,
- et p.s. on a $\displaystyle\lim_{n\to\pm\infty}\frac{1}{n}\log\left|\angle\left(E^s_{f^n(x)}, E^u_{f^n(x)}\right)\right| = 0$.

Montrons d'abord que $\chi^{+} = \widetilde{\chi}^{+}$. On voit que

$$\frac{1}{n}\int\log\left\|A^{(n)}(x)\right\|\mathrm{d}\mu(x) = \frac{1}{n}\int\log\left\|A^{(-n)}(f^n(x))\right\|\mathrm{d}\mu(x) = \frac{1}{n}\int\log\left\|A^{(-n)}(y)\right\|\mathrm{d}\mu(y).$$

En effet dans la première égalité on utilise $(A^n(x))^{-1} = A^{(-n)}(f^n(x))$ et le fait que dans $\mathrm{SL}_2^{\pm}(\mathbb{R})$ on a $\|A\| = \|A^{-1}\|$ et dans la deuxième on utilise l'invariance de μ. Ensuite le membre de gauche converge vers χ^{+} lorsque $n \to \infty$, alors que le membre de droite converge vers $\widetilde{\chi}^{+}$, d'où le résultat.

Pour presque tout x notons $s(x)$ (resp. $u(x)$) un vecteur directeur unitaire de E_x^s (resp. E_x^u). Le deuxième point découle du lemme suivant:

4.10. *Lemme. — Pour presque tout x les vecteurs $s(x)$ et $u(x)$ ne sont pas colinéaires.*

Démonstration. — D'après le théorème d'Oseledets non-inversible, on sait que $v \notin E_x^s$ si et seulement si $\lim_{n\to\infty}\frac{1}{n}\log\left\|A^{(n)}(x)v\right\| = \chi^{+}$. Il suffit donc de démontrer cette propriété pour $u(x)$. En outre, on sait que p.s. la suite $\frac{1}{n}\log\left\|A^{(n)}(x)u(x)\right\|$ converge, soit vers χ^{+}, soit vers χ^{-}. Notons $\psi(x)$ sa limite. Notons également

$$\psi_n(x) = \frac{1}{n}\log\left\|A^{(n)}(x)u(x)\right\| \text{ et } \phi_n(y) = \frac{1}{n}\log\left\|A^{(-n)}(y)u(y)\right\|$$

de sorte que $\psi_n(x) = -\phi_n(f^n(x))$. Par le théorème d'Oseledets appliqué à f^{-1} on a que pour tout $\delta > 0$,

$$\mu\left(\left\{y,\ \left|\phi_n(y) - \chi^{-}\right| > \delta\right\}\right) \underset{n\to\infty}{\longrightarrow} 0.$$

Donc par invariance de μ (et constance de χ^{-}) on déduit que

$$\mu\left(\left\{x,\ \left|\phi_n(f^{-n}(x)) - \chi^{-}\right| > \delta\right\}\right) \underset{n\to\infty}{\longrightarrow} 0.$$

Or

$$\left\{x,\ \left|\phi_n(f^{-n}(x)) - \chi^{-}\right| > \delta\right\} = \left\{x,\ \left|\psi_n((x)) - \chi^{+}\right| > \delta\right\}$$

donc $\psi_n - \chi^{+}$ converge en probabilité vers 0. Comme ψ_n converge p.s. vers ψ on en conclut que $\psi = \chi^{+}$ p.p., ce qu'il fallait démontrer. □

La dernière propriété $\lim_{n\to\pm\infty}\frac{1}{n}\log\left|\angle\left(E^s_{f^n(x)}, E^u_{f^n(x)}\right)\right| = 0$ repose sur le résultat élémentaire suivant:

Exercice. — Il existe une constante universelle c telle que pour toute matrice $A \in \mathrm{SL}_2^{\pm}(\mathbb{R})$ et tout couple de vecteurs $(u,v) \in (\mathbb{R}^2 \backslash \{0\})^2$ on a

$$c^{-1} \|A\|^{-2} \leqslant \frac{|\angle(Au, Av)|}{|\angle(u,v)|} \leqslant c\,\|A\|^2 .$$

Ceci étant admis, si l'on pose $\theta(x) = |\angle(s(x), u(x))|$, par la relation d'invariance des directions E^s et E^u on voit que

$$|\log\theta(f(x)) - \log\theta(x)| \leqslant 2\log\|A(x)\| + \log c$$

et est donc une fonction intégrable. En appliquant le Lemme 4.6 à $\log\theta \circ f - \log\theta$ on obtient alors que $\lim \frac{1}{n}\log\theta(f^n(x)) = 0$ p.s. ce qui termine la preuve. □

5. Théorème d'Oseledets en dimension arbitraire

5.1. — Commençons par un peu de terminologie.

Définition. — Un **drapeau** de $\mathbb{R}^d$ est une famille strictement (dé)croissante de sous espaces vectoriels de $\mathbb{R}^d$, donc de la forme

$$\{0\} < V_k < V_{k-1} < \cdots < V_1 = \mathbb{R}^d.$$

Il est dit complet si $k = d$ (de sorte que la dimension de V_i est égale à $d+1-i$.

Nous avons alors comme précédemment deux versions du théorème ergodique multiplicatif dans les cas inversible et non-inversible.

5.2. *Théorème.* — *Soit $(X, \mathcal{A}, \mu, f)$ un système dynamique mesurable et $A : X \to \mathrm{GL}_d(\mathbb{R})$ un cocycle mesurable tel que $\log^+\|A^{\pm 1}\| \in L^1$. Alors pour μ presque tout x il existe $k = k(x)$, des réels $\lambda_1(x) > \cdots > \lambda_k(x)$ et un drapeau $V_1(x) = \mathbb{R}^d > \cdots V_k(x) > \{0\} = V_{k+1}$ tels que*

1. *les fonctions k, λ_i et le drapeau sont mesurables et invariants,*
2. *pour tout $1 \leqslant i \leqslant k$, pour tout $v \in V_i(x) \backslash V_{i+1}(x)$ on a*

$$\frac{1}{n}\log A^{(n)}(x)v \underset{n\to\infty}{\longrightarrow} \lambda_i(x).$$

Les réels $\lambda_i(x)$ s'appellent les **exposants de Lyapunov** du cocycle A au point x et $\lambda_1(x) = \chi^+(x)$ est l'exposant dominant. De même $\lambda_k(x) = \chi^-(x)$. L'ensemble des exposants forme le **spectre de Lyapunov**. On dit que le spectre est **simple** si $k = d$ (i.e. le drapeau associé est complet). Si μ est ergodique (ou plus généralement le long d'une composante ergodique de μ), le nombre d'éléments du spectre ainsi que les dimensions des sous espaces formant le drapeau sont constantes p.s. On voit que le spectre de Lyapunov décrit l'ensemble des taux de croissance possibles pour les normes des vecteurs itérés le long du cocycle.

5.3. *Théorème.* — *En conservant les hypothèses et notations du théorème précédent, si l'on suppose de plus f inversible, on a une décomposition $\mathbb{R}^d = E_x^1 \oplus \cdots \oplus E_x^{k(x)}$ telle que pour tout $1 \leqslant i \leqslant k$ on a*

1. *$x \mapsto E_x^i$ est mesurable et invariante;*
2. *$V_x^i = \bigoplus_{j=i}^{k} E_x^j$;*
3. *pour tout $v \in E_x^i \backslash \{0\}$ on a*

$$\lim_{n \to \pm\infty} \frac{1}{n} \log \left\| A^{(n)}(x) v \right\| = \lambda_i(x).$$

La démonstration de ces résultats est significativement plus délicate qu'en dimension 2, même si elle s'appuie sur des idées analogues. On pourra consulter pour cela le livre de Viana [**13**, Chap. 4].

5.4. — Le cas le plus important en dynamique différentiable est celui du cocycle tangent, puisque dans ce cas les exposants de Lyapunov décrivent la dynamique infinitésimale de f. Comme expliqué au chapitre IV, le passage de information infinitésimale donnée par la dynamique tangente à une information locale sur M est l'objet de la **théorie de Pesin**. Un exemple de résultat est la généralisation du Théorème IV.5.7: *pour une dynamique C^2 si tous les exposants de Lyapunov d'une mesure ergodique sont strictement négatifs, alors cette mesure est concentrée sur une orbite périodique attractive.*

◇

VII. PRODUITS ALÉATOIRES DE MATRICES 2×2

1. Présentation et principaux résultats

1.1. — Nous allons dans ce chapitre nous intéresser aux produits aléatoires de matrices dans $\mathrm{GL}_2(\mathbb{R})$ [(1)]. Pour toutes les questions relatives aux exposants de Lyapunov, on peut se ramener par factorisation du déterminant au cas de cocycles à valeurs dans $\mathrm{SL}_2^{\pm}(\mathbb{R})$ (l'ensemble des matrices 2×2 de déterminant ± 1). Nous allons également nous intéresser à l'action induite sur l'espace projectif $\mathbb{P}^1(\mathbb{R})$. Pour ce problème on peut travailler indistinctement avec des matrices de GL_2 ou $\mathrm{SL}_2^{\pm}$.

Comme précédemment on note $\|\cdot\|$ la norme euclidienne sur $\mathbb{R}^2$ ainsi que la norme associée sur $\mathrm{M}_2(\mathbb{R})$. Pour $A \in \mathrm{SL}_2^{\pm}$ on a $\|A\| = \|A^{-1}\|$.

1.2. — On a vu au chapitre précédent que l'étude d'un produit aléatoire de matrices i.i.d. revient à celle d'un cocycle au dessus du décalage unilatéral. Rappelons quelques notations: ν est une mesure de probabilité sur $\mathrm{SL}_2^{\pm}(\mathbb{R})$, $\Omega = (\mathrm{SL}_2^{\pm}(\mathbb{R}))^{\mathbb{N}^*}$ et $\mathbb{P} = \nu^{\mathbb{N}^*}$. On supposera toujours satisfaite la condition d'intégrabilité (moment d'ordre 1)

$$\text{(M)} \qquad \int \log \|A\| \, \mathrm{d}\nu(A) < \infty$$

qui permet de définir les exposants de Lyapunov. Un exemple typique, et qui contient déjà toute la complexité du cas général est celui d'une mesure supportée par un ensemble **fini** de matrices, par exemple $\nu = \frac{1}{4}\left(\delta_A + \delta_B + \delta_{A^{-1}} + \delta_{B^{-1}}\right)$.

On considère le décalage $\sigma : (A_n)_{n\geqslant 1} \mapsto (A_{n+1})_{n\geqslant 1}$ et on définit $\pi : \Omega \to \mathrm{GL}_2(\mathbb{R})$ par $\pi(\omega) = A_1$, où $\omega = (A_n)_{n\geqslant 1}$. On a donc ainsi défini un cocycle (σ, π) et une transformation

$$\begin{aligned} F : \quad \Omega \times \mathbb{R}^2 &\longrightarrow \Omega \times \mathbb{R}^2 \\ ((A_n), v) &\longmapsto ((A_{n+1}), A_1 v). \end{aligned}$$

On a également une dynamique induite sur l'espace projectif $\mathbb{P}^1(\mathbb{R})$

$$\begin{aligned} \Theta : \quad \Omega \times \mathbb{P}^1(\mathbb{R}) &\longrightarrow \Omega \times \mathbb{P}^1(\mathbb{R}) \\ ((A_n), [v]) &\longmapsto ((A_{n+1}), [A_1 v]) \end{aligned}$$

qui jouera un rôle fondamental. Nous noterons

$$\ell_n(\omega) := A_n \cdots A_1 \text{ et } r_n(\omega) := A_1 \cdots A_n$$

les produits à gauche (**left**) et à droite (**right**) des n premiers éléments de ω. Le produit à gauche est plus naturel du point de vue des systèmes dynamiques car il correspond à l'ordre naturel de composition dans F^n, mais le produit à droite a de meilleures propriétés

---

(1) On pourra consulter l'ouvrage de Bougerol et Lacroix [**3**] pour le cas des matrices $d \times d$.

de convergence. Ceci étant posé, on a des exposants de Lyapunov $\chi^- \leqslant 0 \leqslant \chi^+$, avec $\chi^- = -\chi^+$ et

$$\chi^+ = \chi^+(\nu) = \text{p.s.} \lim_{n\to\infty} \frac{1}{n} \log \|\ell_n(\omega)\| = \lim_{n\to\infty} \frac{1}{n} \int \log \|\ell_n(\omega)\| \, \mathrm{d}\mathbb{P}(\omega).$$

Noter que comme

$$\int \log \|\ell_n(\omega)\| \, \mathrm{d}\mathbb{P}(\omega) = \int \log \|A_n \cdots A_1\| \, \mathrm{d}\nu(A_n) \cdots \mathrm{d}\nu(A_n),$$

on a également

$$\chi^+ = \lim_{n\to\infty} \frac{1}{n} \int \log \|r_n(\omega)\| \, \mathrm{d}\mathbb{P}(\omega)$$

Exercice. — Montrer que $\frac{1}{n} \log \|r_n(\omega)\|$ converge p.s. vers χ^+.

1.3. *Définition.* — Soit Γ un sous-groupe de $\mathrm{GL}_2(\mathbb{R})$. On dit que Γ est **élémentaire** si l'on a:

- ou bien Γ est relativement compact
- ou bien il existe un sous-ensemble fini de $\mathbb{P}^1(\mathbb{R})$ invariant par Γ.

Nous donnerons une classification des sous-groupes élémentaires à la section suivante. Pour le moment, il suffit d'admettre l'idée que les sous-groupes élémentaires sont "spéciaux" et qu'un sous-groupe "générique" est non-élémentaire (voir l'exercice 2.7 pour une formalisation possible).

1.4. *Théorème (Furstenberg-Kesten). — Soit ν une mesure de probabilité sur $\mathrm{SL}_2^{\pm}(\mathbb{R})$ satisfaisant la condition de moment (M). On suppose que le sous-groupe engendré par* $\mathrm{Supp}(\nu)$ *est non-élémentaire. Alors $\chi^+(\nu) > 0$.*

Ici $\mathrm{Supp}(\nu)$ désigne le support fermé, c'est à dire le complémentaire du plus grand ouvert U tel que $\nu(U) = 0$. L'objet principal de ce chapitre est de démontrer ce résultat, qu'il faut interpréter comme: "sauf dans les cas exceptionnels, les exposants de Lyapunov sont non-nuls", i.e. il y a bien croissance exponentielle des normes des produits $A_1 \cdots A_n$. Cette non-nullité est une différence fondamentale avec la loi des grands nombres! En effet elle est valable même lorsque la mesure ν est "centrée", c'est à dire symétrique par $A \mapsto A^{-1}$. On peut comparer ce phénomène à celui de l'**irréversibilité** en mécanique statistique, où les lois de la mécanique appliquées aux particules individuelles, qui sont invariantes par changement de sens du temps, donnent lieu à la limite d'échelle à des phénomènes qui ne le sont pas, comme la croissance de l'entropie.

1.5. *Exercice.* — Montrer que si ν est une mesure de probabilité sur $\mathrm{GL}_2(\mathbb{R})$ telle que $\log \|A\|$ et $\log \|A^{-1}\|$ sont intégrables et $\mathrm{Supp}(\nu)$ engendre un sous-groupe non élémentaire, alors le spectre de Lyapunov du cocycle associé est simple.

1.6. — Un élément essentiel de la démonstration est l'étude de l'action induite par ν sur l'ensemble des mesures de probabilité sur $\mathbb{P}^1(\mathbb{R})$. En effet si $A \in \mathrm{GL}_2(\mathbb{R})$ on a une transformation induite (toujours notée A) sur $\mathbb{P}^1(\mathbb{R})$ définie par $[v] \mapsto [Av]$, et donc une action induite sur les mesures de probabilité:

$$\mathcal{P}(\mathbb{P}^1) \ni \eta \longmapsto A_* \eta \in \mathcal{P}(\mathbb{P}^1).$$

Remarquer également que si $A \in \mathrm{GL}_2(\mathbb{R})$ et $A = |\det(A)|^{1/2} \widetilde{A}$ (donc $\widetilde{A} \in \mathrm{SL}_2^{\pm}(\mathbb{R})$), alors A et $\widetilde{A}$ ont la même action projective.

Exercice. — Montrer que $(A, \eta) \mapsto A_* \eta$ est continue: $\mathrm{GL}_2(\mathbb{R}) \times \mathcal{P}(\mathbb{P}^1) \to \mathcal{P}(\mathbb{P}^1)$.

Si ν est une mesure de probabilité sur $\mathrm{GL}_2(\mathbb{R})$ On peut alors définir une opération de **convolution** sur les mesures de probabilité de $\mathbb{P}^1(\mathbb{R})$:

$$\eta \longmapsto \nu * \eta = \int A_* \eta \, \mathrm{d}\nu(A),$$

c'est à dire que pour φ une fonction continue sur $\mathbb{P}^1(\mathbb{R})$ on a

$$\langle \nu * \eta, \varphi \rangle = \int \varphi([Av]) \, \mathrm{d}\eta([v]) \mathrm{d}\nu(A).$$

1.7. — La donnée d'une mesure de probabilité ν sur un groupe G détermine une **marche aléatoire** (à gauche) comme suit: on se donne X_0 une variable aléatoire à valeurs dans G et on définit par récurrence une suite de variables aléatoires (X_n) par $X_{n+1} = g_{n+1} X_n$, où (g_n) est une suite de variables aléatoires i.i.d. de loi ν. Si on suppose que $X_0 = \{e\}$ p.s. alors la loi de X_n est la mesure ν^n définie par

$$\int_G \varphi(g) \, \mathrm{d}\nu^n(g) = \int \varphi(g_n \cdots g_1) \, \mathrm{d}\nu(g_1) \cdots \mathrm{d}\nu(g_n).$$

Exercice. — Montrer que si $\eta \in \mathcal{P}(\mathbb{P}^1)$ pour tout n on a $\nu * (\nu^n * \eta) = \nu^{n+1} * \eta$.

1.8. *Définition.* — On dit que $\mu \in \mathcal{P}(\mathbb{P}^1(\mathbb{R}))$ est **ν-stationnaire** si $\nu * \mu = \mu$.

1.9. *Proposition.* — *Si ν est une mesure de probabilité quelconque sur* $\mathrm{GL}_2(\mathbb{R})$ *alors il existe une mesure de probabilité ν-stationnaire sur* $\mathbb{P}^1(\mathbb{R})$.

Démonstration. — Exercice (penser à Krylov-Bogolyubov). □

1.10. — La donnée de la probabilité ν sur $\mathrm{GL}_2(\mathbb{R})$ détermine également une **chaîne de Markov** à temps discret sur $\mathbb{P}^1(\mathbb{R})$, c'est à dire une chaîne de Markov à temps discret sur un espace d'états continu. La définition est similaire à celle du §1.7: étant donnée une variable aléatoire Y_0 à valeurs dans $\mathbb{P}^1(\mathbb{R})$ de loi μ_0, on définit par récurrence (Y_n) par $Y_{n+1} = (A_{n+1})_* Y_n$, où les (A_n) sont i.i.d. de loi ν. Concrètement, à chaque étape un point $[v]$ de $\mathbb{P}^1(\mathbb{R})$ est envoyé sur $[Av]$, où A est de loi ν. La loi de Y_n est donnée par $\nu^n * \mu_0$. On voit que si la loi de Y_0 est une mesure stationnaire, alors les Y_n ont la même loi, d'où la terminologie.

1.11. — Le résultat suivant est une étape essentielle du Théorème 1.4. Il est également le reflet d'un principe un peu général: alors que l'unique ergodicité est un phénomène plutôt rare pour un système dynamique “classique” (i.e. lorsqu'on itère une seule transformation), cela devient la norme lorsque plusieurs transformations sont en jeu (phénomène dit de **rigidité**).

Théorème. — Soit ν *une mesure de probabilité sur* $\mathrm{GL}_2(\mathbb{R})$ *telle que le sous-groupe engendré par* $\mathrm{Supp}(\nu)$ *est non-élémentaire. Alors il existe une unique mesure stationnaire* μ *sous l'action de* ν *sur* $\mathbb{P}^1(\mathbb{R})$*. En outre pour toute mesure de probabilité* η *sur* $\mathbb{P}^1(\mathbb{R})$*,* $\nu^n * \eta$ *converge vers* μ*.*

2. Sous-groupes élémentaires

Dans cette section nous allons donner une classification des sous-groupes élémentaires de $\mathrm{GL}_2(\mathbb{R})$. Rappelons qu'un sous groupe est élémentaire s'il est relativement compact ou bien préserve une famille finie de droites vectorielles.

2.1. — Le cas compact est réglé par le résultat classique suivant, qui vaut en dimension arbitraire d:

Proposition. — Soit Γ *un sous-groupe relativement compact de* $\mathrm{GL}_d(\mathbb{R})$*. Alors* Γ *est conjugué à un sous-groupe de* $\mathrm{O}_d(\mathbb{R})$*, i.e. il existe* $P \in \mathrm{GL}_d(\mathbb{R})$ *tel que* $P^{-1}\Gamma P \leqslant \mathrm{O}_d(\mathbb{R})$*.*

Démonstration. — On rappeller que $A \in \mathrm{O}_d(\mathbb{R})$ si et seulement si A préserve le produit scalaire canonique de $\mathbb{R}^d$:

$$\forall x, y \in \mathbb{R}^d,\ \langle Ax, Ay\rangle = \langle x, y\rangle.$$

L'observation fondamentale est la suivante:

2.2. *Lemme. — Si* Γ *est un sous groupe compact de* $\mathrm{GL}_d(\mathbb{R})$ *alors* Γ *admet un produit scalaire invariant.*

Démonstration. — Soit m la probabilité de Haar de Γ, qui est invariante à gauche et à droite. Pour $x, y \in \mathbb{R}^d$, on pose

$$\langle x, y\rangle_\Gamma = \int_\Gamma \langle gx, gy\rangle dm(g).$$

Il est clair que $\langle\cdot,\cdot\rangle_\Gamma$ est un produit scalaire sur $\mathbb{R}^d$ et par invariance à droite de m on a pour tout $h \in \Gamma$,

$$\langle hx, hy\rangle_\Gamma = \int_\Gamma \langle ghx, ghy\rangle dm(g) = \int_\Gamma \langle g'x, g'y\rangle dm(g') = \langle x, y\rangle_\Gamma$$

d'où le résultat. □

On conclut alors la preuve de la proposition. Déjà, quitte à remplacer Γ par $\overline{\Gamma}$, on peut le supposer compact. Ensuite, le produit scalaire invariant s'exprime sous la forme

$$\langle x, y\rangle_\Gamma = {}^t x H y,$$

où H est une certaine matrice symétrique définie positive. Comme pour toute matrice $A \in \Gamma$ on a pour tous x, y, ${}^t x {}^t A H A y = {}^t x H y$, on déduit que pour toute $A \in \Gamma$, ${}^t A H A = H$. Ainsi

$$(\sqrt{H})^{-1}\, {}^t A \sqrt{H} \cdot \sqrt{H} A (\sqrt{H})^{-1} = I_d$$

et donc $(\sqrt{H})^{-1}\, {}^t A \sqrt{H} \in \mathrm{O}_d(\mathbb{R})$, ce qu'il fallait démontrer. □

2.3. — Dans le cas $d = 2$, la structure du groupe $\mathrm{O}_2(\mathbb{R})$ est particulièrement simple: ses éléments sont des rotations ou des symétries axiales, et il est le produit semi-direct du groupe $\mathrm{SO}_2(\mathbb{R})$ des rotations par le groupe engendré par une symétrie, donc isomorphe à $\mathbb{T}^1 \rtimes (\mathbb{Z}/2\mathbb{Z})$.

2.4. — Il faut maintenant comprendre les cas où Γ préserve un ensemble fini $\mathcal{D}$ de droites.

1er cas: $\#\mathcal{D} = 1$. Soit $\mathcal{D} = D$. Alors pour tout $A \in \Gamma$, $AD = D$, D est une droite propre et tous les éléments de Γ ont un vecteur propre commun. Par une conjugaison de Γ on peut supposer que $D = \mathrm{Vect}(1, 0)$ et on conclut que **quitte à conjuguer toutes les matrices de Γ sont triangulaires supérieures.**

2e cas: $\#\mathcal{D} = 2$. Quitte à conjuguer Γ on peut supposer que $\mathcal{D} = \{D_1, D_2\}$, avec $D_1 = \mathrm{Vect}(1, 0)$ et $D_2 = \mathrm{Vect}(0, 1)$. On montre alors aisément qu'après ce changement de base toute matrice de Γ est diagonale ou anti-diagonale, autrement dit Γ est **conjugué à un sous groupe du groupe suivant:**

$$\left\{ \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix},\ (a, b) \in (\mathbb{R}^*)^2 \right\} \cup \left\{ \begin{pmatrix} 0 & a \\ b & 0 \end{pmatrix},\ (a, b) \in (\mathbb{R}^*)^2 \right\}.$$

3e cas: $\#\mathcal{D} \geqslant 3$. Toute matrice $A \in \Gamma$ permute les éléments de $\mathcal{D}$ donc il existe N divisant $(\#\mathcal{D})!$ tel que pour toute droite $D \in \mathcal{D}$, $A^N D = D$. Choisissons N minimal tel que cette relation soit vraie pour tout $A \in \Gamma$. Une matrice 2×2 fixant 3 directions différentes est un multiple de l'identité (exercice) donc $A^N = \lambda(A) I_2$. Soit $\Gamma' \leqslant \Gamma$ le sous groupe distingué (pourquoi?) des matrices fixant chaque élément de $\mathcal{D}$. Alors Γ' est un groupe d'homothéties donc **isomorphe à un sous groupe de** $\mathbb{R}$. En outre, Γ/Γ' est **fini**.

Remarquer que si Γ est un sous groupe de $\mathrm{SL}_2^{\pm}(\mathbb{R})$ la conclusion s'exprime un peu plus simplement: Γ est fini. En effet les transformations de Γ' sont de la forme $\pm \mathrm{I}_2$.

2.5. *Remarque-Exercice.* — Les résultats qui précèdent montrent que la structure algébrique des sous-groupes élémentaires de $\mathrm{SL}_2^{\pm}(\mathbb{R})$ est relativement simple. Si Γ est relativement compact ou si Γ préserve un ensemble $\mathcal{D}$ de droites de cardinal $\geqslant 2$ alors Γ admet un sous-groupe abélien d'indice fini. Si Γ préserve une droite alors Γ est isomorphe à un sous groupe du groupe affine de $\mathbb{R}$ (pourquoi?).

2.6. *Remarque.* — On peut montrer que si Γ est non-élémentaire, alors il existe des éléments a et b de Γ qui engendrent un sous-groupe libre. Ainsi la structure algébrique de Γ est très différente du cas élémentaire.

2.7. *Exercice.* — Montrer que les sous-groupes typiques sont non-élémentaires au sens suivant: pour tout $n \geqslant 2$, pour presque tous $(a_1, \dots, a_n) \in \mathrm{GL}_2(\mathbb{R})^n$ (resp. $\mathrm{SL}_2(\mathbb{R})^n$), le sous-groupe $\langle a_1, \dots, a_n \rangle$ est non-élémentaire.

3. Argument de contraction et unicité de la mesure stationnaire

Notre objectif dans cette section est de démontrer le théorème 1.11. On se donne donc une mesure de probabilité ν sur $\mathrm{GL}(\mathbb{R})$ telle que le sous-groupe Γ engendré par $\mathrm{Supp}(\nu)$ est non-élémentaire. Si comme ci-dessus on note $\pi : \mathrm{GL}_2(\mathbb{R}) \to \mathrm{SL}_2^{\pm}(\mathbb{R})$ la projection naturelle, alors il revient au même d'étudier l'action de ν ou celle de $\pi_*\nu$ donc on se ramène au cas où $\mathrm{Supp}(\nu) \subset \mathrm{SL}_2^{\pm}$. On va étudier l'action de ν par convolution sur $\mathcal{P}(\mathbb{P}^1(\mathbb{R}))$.

3.1. — Nous travaillerons avec la distance suivante sur $\mathbb{P}^1(\mathbb{R})$:

$$d([U],[V]) = |\sin(\angle(u,v))|\,,\ (u,v) \text{ représentants unitaires de } (U,V).$$

Ainsi $d([U],[V]) = (1-\langle u,v\rangle^2)^{1/2}$ et on vérifie qu'en coordonnées si $U = (u_1,u_2)$ et $V = (v_1,v_2)$ on obtient la formule

$$d([U],[V]) = \left(1 - \frac{\langle U,V\rangle^2}{\|U\|^2\,\|V\|^2}\right)^{1/2} = \frac{|u_1v_2 - u_2v_1|}{\|U\|\cdot\|V\|}.$$

Noter que le diamètre de $\mathbb{P}^1$ pour cette distance est égal à 1. Nous noterons $B(a,r)$ la boule fermée de centre a et de rayon r.

3.2. — On rappelle la décomposition KAK: $A = U\Sigma V$, où U et V sont orthogonales et $\Sigma = \mathrm{diag}(\|A\|, \|A\|^{-1})$. Ainsi à des rotations à la source et au but près, l'action métrique de A sur $\mathbb{P}^1(\mathbb{R})$ est la même que celle de Σ. Si $\|A\|$ est grande, alors Σ envoie toute direction pas trop proche de celle de $e_1 := (0,1)$ à proximité de celle de $e_2 := (1,0)$. Plus précisément, si $u = (u_1,u_2)$ alors

$$d([\Sigma u],[e_1]) = \frac{\|A\|^{-1}|u_2|}{\left(\|A\|^2 u_1^2 + \|A\|^{-2} u_2^2\right)^{1/2}},$$

et donc si $|u_1| \geqslant \varepsilon\,|u_2|$, ceci est majoré par

$$\frac{\|A\|^{-1}|u_2|}{\left(\|A\|^2\varepsilon^2 + \|A\|^{-2}\right)^{1/2}|u_2|} \leqslant \frac{\|A\|^{-1}}{\varepsilon\,\|A\|} = \frac{1}{\varepsilon\,\|A\|^2}.$$

En prenant $\varepsilon = \|A\|^{-1}$ on trouve ainsi que $\Sigma(\mathbb{P}^1\backslash B(e_2,\varepsilon)) \subset B(e_1,\varepsilon)$. Pour A, cela donne:

Proposition. — Pour toute matrice $A \in \mathrm{SL}_2^{\pm}$ de norme assez grande (disons $\|A\| \geqslant 10$) il existe des points $\alpha(A)$ et $\omega(A)$ dans $\mathbb{P}^1(\mathbb{R})$ tels que

$$A\big(\mathbb{P}^1\backslash B(\alpha, 2\,\|A\|^{-1})\big) \subset B(\omega, 2\,\|A\|^{-1}).$$

3.3. *Lemme. — Soit η une mesure de probabilité diffuse sur $\mathbb{P}^1(\mathbb{R})$ et $(A_n) \in (\mathrm{SL}_2^{\pm})^{\mathbb{N}^*}$. Alors on a l'équivalence*

$$\|A_n\| \underset{n\to\infty}{\longrightarrow} \infty \Leftrightarrow \textit{ toutes les valeurs d'adhérence de } (A_n)_*\eta \textit{ sont des masses de Dirac.}$$

Démonstration. — $\Leftarrow$ Supposons que $\|A_n\|$ ne tende pas vers l'infini. Alors on peut extraire $A_{n_j} \to B \in \mathrm{SL}_2^{\pm}$ et ainsi $(A_{n_j})_\star \eta \to B_\star \eta$ qui n'est pas atomique.

$\Rightarrow$ Le fait bien connu suivant découle de la compacité de $\mathbb{P}^1(\mathbb{R})$ (exercice!)

$$\forall \varepsilon > 0,\ \exists \delta > 0,\ \forall x \in \mathbb{P}^1(\mathbb{R}),\ \eta(B(x,\delta)) < \varepsilon.$$

D'après la proposition précédente, pour tout n il existe α_n et ω_n tels que

$$A_n \left(\mathbb{P}^1 \backslash B\left(\alpha_n, 2\|A_n\|^{-1}\right)\right) \subset B\left(\omega_n, 2\|A_n\|^{-1}\right).$$

Donc si $\|A_n\| > 2/\delta$ on obtient que

$$(A_n)_\star \eta \left(B\left(\omega_n, 2\|A_n\|^{-1}\right)\right) \geqslant 1 - \varepsilon.$$

Si maintenant (n_j) est une sous-suite telle que $(A_{n_j})_\star \eta$ converge vers η_∞, quitte à extraire à nouveau on peut supposer que $\omega_{n_j} \to \omega$ et on voit que pour tout $\varepsilon > 0$ on a pour n assez grand $\eta_\infty(B(\omega,\varepsilon)) \geqslant 1 - \varepsilon$, c'est à dire que $\eta_\infty = \delta_\omega$, ce qu'il fallait démontrer. □

3.4. *Corollaire. — Si η est une probabilité diffuse sur $\mathbb{P}^1(\mathbb{R})$, alors*

$$\mathrm{Stab}(\eta) = \left\{A \in \mathrm{SL}_2^{\pm}(\mathbb{R}),\ A_\star \eta = \eta\right\}$$

est un sous-groupe compact de $\mathrm{SL}_2^{\pm}(\mathbb{R})$.

3.5. *Proposition. — Si $\langle \mathrm{Supp}(\nu) \rangle$ est non élémentaire alors toute mesure de probabilité ν-stationnaire sur $\mathbb{P}^1(\mathbb{R})$ est sans atome.*

Démonstration. — Soit μ une proba stationnaire et écrivons sa décomposition $\mu = \mu^a + \mu^d$ en parties atomique et diffuse. Si $\mu^a \neq 0$, il existe un ensemble fini F d'atomes de masse maximale, i.e.

$$\mu^a = \sum_{x \in F} c_{\max} \delta_x + \sum_y c_y \delta_y, \text{ avec } c_y < c_{\max}.$$

On a donc une décomposition $\mu = \mu_{\max} + \mu'$, où μ' n'a aucun atome de masse supérieure ou égale à $c_{\max}$. Considérons maintenant la relation d'invariance

$$\mu = \nu * \mu = \nu * \mu_{\max} + \nu * \mu'.$$

Le fait est que $\nu * \mu'$ n'a aucun atome de masse supérieure ou égale à $c_{\max}$. En effet pour tout $x \in \mathbb{P}^1$ on a

$$\nu * \mu'(\{x\}) = \int A_\star \mu'(\{x\})\, \mathrm{d}\nu(A) = \int \mu'\left(\left\{A^{-1}x\right\}\right) \mathrm{d}\nu(A) < c_{\max}$$

car la dernière intégrale est une moyenne de nombres inférieurs à $c_{\max}$. En utilisant $\nu * \mu = \mu$ en déduit que $\nu * \mu' \leqslant \mu'$. Mais comme la convolution préserve la masse, ces deux mesures ont la même masse et donc $\nu * \mu' = \mu'$, et donc également $\nu * \mu_{\max} = \mu_{\max}$. Cette dernière relation se traduit en

$$\sum_{x \in F} \int A_\star \delta_x \mathrm{d}\nu(A) = \sum_{x \in F} \delta_x,$$

d'où il vient que pour ν-presque tout A on a $AF \subset F$ et donc $AF = F$ car A est bijective et F fini. Finalement, l'ensemble des A tels que $AF = F$ est fermé et on conclut que tous les éléments de $\mathrm{Supp}(\nu)$ préservent F. Comme cette propriété passe au groupe engendré, on conclut que $\langle \mathrm{Supp}(\nu)\rangle$ préserve l'ensemble fini F, et est donc élémentaire. □

3.6. — On rappelle que la mesure ν^n est définie comme l'image de la mesure $\nu^{\times n}$ par le produit

$$\big(\mathrm{SL}_2^{\pm}(\mathbb{R})\big)^n \ni (A_1, \ldots, A_n) \longmapsto A_1 \cdots A_n \in \mathrm{SL}_2^{\pm}(\mathbb{R}),$$

ou autrement dit

$$\int_G \varphi(A)\,\mathrm{d}\nu^n(A) = \int \varphi(A_n \cdots A_1)\,\mathrm{d}\nu(A_1)\cdots \mathrm{d}\nu(A_n).$$

On pose alors

$$\nu^\infty = \sum_{k=0}^{\infty} \frac{1}{2^{k+1}} \nu^k.$$

Exercice. — Montrer que $\mathrm{Supp}(\nu^\infty)$ est le sous-semigroupe engendré par $\mathrm{Supp}(\nu)$.

3.7. — On rappelle que pour $\omega \in \Omega = \big(\mathrm{SL}_2^{\pm}(\mathbb{R})\big)^{\mathbb{N}^*}$, $r_n(\omega)$ désigne le produit à droite des n premiers éléments de ω. On rappelle également la notation $\mathbb{P} = \nu^{\mathbb{N}^*}$.

Lemme. — Soit ν une mesure de probabilité arbitraire sur $\mathrm{SL}_2^{\pm}(\mathbb{R})$ et μ une mesure ν-stationnaire sur $\mathbb{P}^1(\mathbb{R})$. Alors pour $\mathbb{P}$-presque tout $\omega \in \Omega$ il existe une mesure μ_ω telle que pour ν^∞ presque tout B on a

$$(r_n(\omega))_* B_* \mu \underset{n\to\infty}{\longrightarrow} \mu_\omega.$$

Démonstration du cas $B = I_2$. — Considérons la suite de mesures de probabilité aléatoires $r_n(\omega)_*\mu$. L'observation fondamentale est que celle-ci forme une **martingale**: en effet en raisonnant informellement on voit que

$$\mathbb{E}\left((A_1)_* \cdots (A_{n+1})_*\mu | A_1, \ldots, A_n\right) = (A_1)_* \cdots (A_n)_*\mu.$$

Ainsi cette suite de mesures converge presque sûrement et on a le résultat.

Plus formellement, pour $n \geqslant 1$ posons $\mathcal{F}_n$ la tribu sur Ω engendrée par les cylindres de profondeur n, i.e. $\mathcal{F}_n = \sigma(\{A_1, \ldots, A_n\})$, ce qui définit une filtration croissante: $\mathcal{F}_n \subset \mathcal{F}_{n+1}$. On fixe une fonction continue φ sur $\mathbb{P}^1$ et pour $\omega \in \Omega$ on pose

$$\phi_n(\omega) = \langle r_n(\omega)_*\mu, \varphi\rangle = \int \varphi(A_1 \cdots A_n x)\,\mathrm{d}\mu(x), \tag{1}$$

qui est $\mathcal{F}_n$ mesurable par définition. On a ainsi défini ainsi une suite de variables aléatoires réelles uniformément bornées. Par l'indépendance des A_i, $\mathbb{E}\left(\varphi_{n+1}|\mathcal{F}_n\right)$ s'obtient simplement par intégration par rapport à la $(n+1)^{\mathrm{e}}$ variable, et donc

$$\mathbb{E}\left(\varphi_{n+1}|\mathcal{F}_n\right) = \int \varphi(A_1 \cdots A_{n+1}x)\,\mathrm{d}\mu(x)\mathrm{d}\nu(A_{n+1}) = \int \varphi(A_1 \cdots A_n x)\,\mathrm{d}\mu(x)$$

par stationnarité. Ainsi (ϕ_n) est une martingale bornée, et donc converge p.s. vers une limite $\phi_\infty(\omega)$. On répète ensuite cet argument pour une suite dense (φ_j) de fonctions continues sur $\mathbb{P}^1$ et on obtient un ensemble de mesure totale $\Omega' \subset \Omega$ tel que si $\omega \in \Omega'$, pour tout j, $\langle r_n(\omega)_*\mu, \varphi_j\rangle$ converge vers une limite $\phi_{\infty,j}(\omega)$. Finalement, pour $\omega \in \Omega'$,

$\varphi_j \mapsto \phi_{\infty,j}(\omega)$ s'étend par densité en une fonctionnelle continue sur l'espace des fonctions continues sur $\mathbb{P}^1$, qui est la mesure limite cherchée. □

Démonstration du cas général. — En raisonnant comme ci-dessus, il suffit de montrer la convergence presque sûre de $\langle r_n(\omega)_* B_* \mu, \varphi\rangle$ pour une fonction continue φ fixée. Pour ϕ_n comme en (1) on écrit

$$\begin{aligned}\mathbb{E}\left(|\phi_{n+k}-\phi_n|^2\right) &= \mathbb{E}\left(\phi_{n+k}^2\right)+\mathbb{E}\left(\phi_n^2\right)-2\mathbb{E}\left(\phi_{n+k}\phi_n\right)\\ &= \mathbb{E}\left(\phi_{n+k}^2\right)+\mathbb{E}\left(\phi_n^2\right)-2\mathbb{E}\left(\phi_n^2\right)=\mathbb{E}\left(\phi_{n+k}^2\right)-\mathbb{E}\left(\phi_n^2\right),\end{aligned}$$

et on en déduit par téléscopage que pour tout $p \geqslant 1$,

$$\sum_{n=1}^{p}\mathbb{E}\left(|\phi_{n+k}-\phi_n|^2\right)=\sum_{n=1}^{k}\mathbb{E}\left(\phi_{p+n}^2\right)-\sum_{n=1}^{k}\mathbb{E}\left(\phi_n^2\right)\leqslant\sum_{n=1}^{k}\mathbb{E}\left(\phi_{p+n}^2\right)\leqslant k\,\|\varphi\|_\infty^2.$$

Posons maintenant $\Psi(h) = \langle h_*\mu, \varphi\rangle$ de sorte que $\phi(\omega) = \Psi(r_n(\omega))$. On estime alors

$$\begin{aligned}&\sum_{n=1}^{\infty}\mathbb{E}\left(\int|\Psi(r_n(\omega)B)-\Psi(r_n(\omega))|^2\,\mathrm{d}\nu^\infty(B)\right)\\ &\quad=\sum_{n=1}^{\infty}\sum_{k=0}^{\infty}\frac{1}{2^{k+1}}\underbrace{\mathbb{E}\left(\int|\Psi(r_n(\omega)B)-\Psi(r_n(\omega))|^2\,\mathrm{d}\nu^k(B)\right)}_{\mathbb{E}\left(|\phi_{n+k}-\phi_n|^2\right)}\leqslant\sum_{k=0}^{\infty}\frac{k}{2^{k+1}}\,\|\varphi\|_\infty^2<\infty.\end{aligned}$$

Donc p.s. la série de terme général $|\Psi(r_n(\omega)B) - \Psi(r_n(\omega))|^2$ converge et on conclut que pour $(\mathbb{P}\times\nu^\infty)$-presque tout (ω, B),

$$\langle r_n(\omega)_*B_*\mu,\varphi\rangle-\langle r_n(\omega)_*\mu,\varphi\rangle=\Psi(r_n(\omega)B)-\Psi(r_n(\omega))\underset{n\to\infty}{\longrightarrow}0,$$

ce qui, grâce à la première partie de la démonstration, implique le résultat souhaité. □

3.8. — En rassemblant tout ce qui précède on obtient un résultat de contraction presque sûre sur $\mathbb{P}^1$, soit de divergence presque sûre des produits $r_n(\omega)$.

Proposition. — Si ν une mesure de probabilité sur $\mathrm{SL}_2^\pm(\mathbb{R})$ telle que $\langle\mathrm{Supp}(\nu)\rangle$ est non-élémentaire, alors pour presque tout ω, $\|r_n(\omega)\|$ tend vers l'infini.

Remarquer que $r_n(\omega)$ et $\ell_n(\omega)$ ont la même loi, ainsi $\|\ell_n(\omega)\|$ tend vers l'infini en loi.

Démonstration. — Soit ω satisfaisant les conclusions du Lemme 3.7, et montrons que $\|r_n(\omega)\| \to \infty$. Supposons par l'absurde que ce ne soit pas le cas. Alors $r_n(\omega)$ admet une sous-suite bornée, et donc une sous-suite convergeant vers $A \in \mathrm{SL}_2^\pm$. En passant à la limite dans le Lemme 3.7 on obtient que pour ν^∞ presque tout B, on a $A_*B_*\mu = \mu_\omega$, et donc $B_*\mu = A^*\mu_\omega$. Donc pour $\nu^\infty\times\nu^\infty$-presque tout (B_1, B_2) on a $(B_1)_*\mu = (B_2)_*\mu$, soit $(B_2^{-1}B_1)_*\mu = \mu$. Soit $\widetilde{\nu}^\infty$ la mesure image de $\nu^\infty\times\nu^\infty$ par l'application $(B_1, B_2) \mapsto B_2^{-1}B_1$. On vérifie simplement que $\mathrm{Supp}(\widetilde{\nu}^\infty)$ est précisément le sous-groupe engendré par ν. En outre pour $\widetilde{\nu}^\infty$-presque tout C, on a $C_*\mu = \mu$, ainsi $C \in \mathrm{Stab}(\mu)$ qui d'après le Corollaire 3.4 est un sous-groupe compact de $\mathrm{SL}_2^\pm(\mathbb{R})$. Ceci contredit donc l'hypothèse de non-élémentarité, et la proposition est démontrée. □

3.9. *Corollaire. — Si μ est une mesure stationnaire, alors p.s. $\mu_\omega = \delta_{\mathrm{e}(\omega)}$ est une masse de Dirac.*

Démonstration. — En effet μ est diffuse par la Proposition 3.5 et donc toutes les valeurs d'adhérence de $r_n(\omega)_*\mu$ sont des masses de Dirac par le Lemme 3.3. On conclut par la convergence $r_n(\omega)_*\mu \to \mu_\omega$. □

3.10. — Nous sommes maintenant en mesure de démontrer l'unicité de la mesure stationnaire (Théorème 1.11) La preuve découle de deux observations: la première est que si η est n'importe quelle mesure diffuse sur $\mathbb{P}^1$ et ω est comme dans la proposition précédente, alors $r_n(\omega)_*\eta \to \delta_{\mathrm{e}(\omega)}$. Cela découle de la même analyse qu'au Lemme 3.3: η charge peu la boule répulsive de $r_n(\omega)$, donc presque toute sa masse est envoyée par $r_n(\omega)$ dans sa boule attractive, qui est proche de $\mathrm{e}(\omega)$. En particulier, si μ' est n'importe quelle autre proba stationnaire, on a pour presque tout ω que $\mu'_\omega = \mu_\omega = \delta_{\mathrm{e}(\omega)}$.

La deuxième observation est que n'importe quelle mesure stationnaire se reconstruit à partir des mesures limites μ_ω selon la formule

$$\mu = \int \mu_\omega \, \mathrm{d}\mathbb{P}(\omega). \tag{2}$$

En effet par stationnarité on a

$$\mu = \int A_*\mu \, \mathrm{d}\nu(A) = \int (A_1)_* \cdots (A_n)_*\mu \, \mathrm{d}\nu(A_1)\cdots \mathrm{d}\nu(A_n) = \int r_n(\omega)_*\mu \, \mathrm{d}\mathbb{P}(\omega),$$

(la deuxième égalité découle de l'indépendance des A_i). Comme p.s. $r_n(\omega)_*\mu \to \mu_\omega$ d'où (2) par le théorème de convergence dominée. Or on a vu plus haut que les μ_ω ne dépendent pas de la mesure stationnaire, et on conclut ainsi que celle ci est unique.

3.11. — Le dernier point à établir est le fait que si η est une mesure quelconque sur $\mathbb{P}^1$, alors $\nu^n * \eta$ converge vers μ. Nous ne traiterons que le cas où η est diffuse, le cas général est plus délicat, voir [**3**] (voir également le théorème de Breiman au chapitre X pour un résultat apparenté). D'après ce qui précède, presque sûrement on a $r_n(\omega)_*\eta \rightharpoonup \mu_\omega$, donc par le théorème de convergence dominée et la décomposition (2) $\mathbb{E}(r_n(\omega)_*\eta) \rightharpoonup \mu$. Mais dans cette espérance l'ordre des compositions ne compte pas, donc on a aussi $\mathbb{E}(\ell_n(\omega)_*\eta) \rightharpoonup \mu$, qui est exactement le résultat souhaité. □

4. Positivité de l'exposant de Lyapunov

Nous sommes maintenant en mesure de démontrer le théorème 1.4 de positivité de l'exposant. L'argument repose en grande partie sur l'étude du système dynamique induit sur $\mathbb{P}^1(\mathbb{R})$:

$$\begin{aligned} \Theta : \quad \Omega \times \mathbb{P}^1(\mathbb{R}) &\longrightarrow \Omega \times \mathbb{P}^1(\mathbb{R}) \\ ((A_n), [v]) &\longmapsto ((A_{n+1}), [A_1 v]). \end{aligned}$$

4.1. Formule de Furstenberg pour l'exposant. —

4.1.1. — Remarquons tout d'abord que pour $A \in \mathrm{SL}_2^{\pm}$, l'application

$$\mathbb{R}^2 \ni V \longmapsto \frac{\|AV\|}{\|V\|}$$

est invariante par $V \mapsto \lambda V$, donc elle descend en une fonction sur $\mathbb{P}^1$. Ainsi pour $v \in \mathbb{P}^1(\mathbb{R})$ on peut considérer

$$\sigma(A, v) := \log \frac{\|AV\|}{\|V\|},$$

où V est n'importe quel relevé de v.

4.1.2. *Proposition. — Sous les hypothèses du Théorème 1.4 on a*

$$\chi^+ = \int_{\mathrm{SL}_2^{\pm} \times \mathbb{P}^1} \sigma(A, v)\, \mathrm{d}\nu(A) \mathrm{d}\mu(v), \tag{3}$$

où μ est l'unique probabilité ν-stationnaire sur $\mathbb{P}^1$.

Démonstration. — Pour $(\omega, v) \in \Omega \times \mathbb{P}^1$ posons $L(\omega, v) = \sigma(\ell_1(\omega), v)$. C'est une fonction mesurable bornée sur $\Omega \times \mathbb{P}^1$. On a

$$\begin{aligned} \sigma(A_n \cdots A_1, v) = \log \frac{\|A_n \cdots A_1 V\|}{\|V\|} &= \sum_{k=1}^{n} \log \frac{\|A_k \cdots A_1 V\|}{\|A_{k-1} \cdots A_1 V\|} \\ &= \sum_{k=1}^{n} \log \frac{\|A_k W\|}{\|W\|}, \text{ où } W = A_{k-1} \cdots A_1 V \\ &= \sum_{k=1}^{n} L\left(\sigma^k \omega, \ell_{k-1}(\omega) v\right) = \sum_{k=1}^{n} L \circ \Theta^k(\omega, v) \text{ où } \omega = (A_n)_{n \geqslant 1}. \end{aligned}$$

On voit donc que $\frac{1}{n}\sigma(A_n \cdots A_1, v)$ est une somme de Birkhoff associée à Θ.

D'après le théorème d'Oseledets, pour $\mathbb{P}$-presque tout $\omega \in \Omega$ et pour tout V in $\mathbb{R}^2 \backslash \{0\}$ à l'exception d'au plus une droite on a

$$\frac{1}{n} \log \|\ell_n(\omega) V\| \underset{n \to \infty}{\longrightarrow} \chi^+.$$

En particulier comme μ ne charge pas les points, on a que pour μ-presque tout v

$$\frac{1}{n} \log \sigma(\ell_n(\omega), v) \underset{n \to \infty}{\longrightarrow} \chi^+.$$

Par ailleurs d'après le théorème de Birkhoff on a

$$\frac{1}{n} \sum_{k=1}^{n} L \circ \Theta^k(\omega, v) \underset{n \to \infty}{\longrightarrow} \mathbb{E}(L | \mathcal{I})$$

p.s. et dans L^1. Ainsi $\chi^+ = \mathbb{E}(L|\mathcal{I})$ p.s. Finalement comme $\mathbb{E}(\mathbb{E}(L|\mathcal{I})) = \mathbb{E}(L)$ on conclut que

$$\chi^+ = \mathbb{E}(L) = \int \sigma(\ell_1(\omega), v)\, \mathrm{d}\mathbb{P}(\omega) \mathrm{d}\mu(v) = \int \sigma(A, v)\, \mathrm{d}\nu(A) \mathrm{d}\mu(v)$$

où dans la dernière égalité on a utilisé la structure de produit de $\mathbb{P} = \nu^{\mathbb{N}^*}$ et le fait que $\sigma(\ell_1(\omega), v)$ ne dépend que de A_1. $\square$

Remarque. — On montre assez facilement que $\mathbb{P}\times\mu$ est Θ-ergodique, de sorte qu'on aurait pu directement remplacer les espérances conditionnelles $\mathbb{E}(L|\mathcal{I})$ par des $\mathbb{E}(L)$.

4.2. Démonstration du théorème. — Nous voulons démontrer que $\chi^+ > 0$, c'est à dire que p.s.

$$\lim_{n\to\infty} \frac{1}{n} \log \|\ell_n(\omega)\| > 0.$$

À ce stade on sait que p.s on a $\|r_n(\omega)\| \to \infty$.

4.2.1. *Première étape.* — Pour $\mathbb{P} \times \mu$-presque tout (ω, v) on a $\sigma(\ell_n(\omega), v) \to \infty$.

Pour démontrer ce résultat on introduit le **transposé** de notre produit aléatoire, c'est à dire qu'on considère la mesure ${}^t\nu$ sur $\mathrm{SL}_2^{\pm}$ image de ν par $A \mapsto {}^tA$. Pour toute matrice de $\mathrm{SL}_2^{\pm}$ on a $\|A\| = \|{}^tA\|$ (exercice) donc ${}^t\nu$ vérifie la condition de moment (M) et on vérifie également simplement que $\langle\mathrm{Supp}({}^t\nu)\rangle$ est non-élémentaire. Notons ℓ_n^t, r_n^t, etc. les objets associés à ce nouveau produit de matrices [(2)]. Alors d'après les résultats du §3 on a p.s. $\|r_n^t(\omega)\| \to \infty$ et $r_n^t(\omega)_\star m \to \delta_{z(\omega)}$ pour toute mesure diffuse m sur $\mathbb{P}^1$.

De cela on déduit que pour le produit initial on a p.s. $\|\ell_n(\omega)\| \to \infty$. En effet on a $\ell_n(\omega) = {}^t(r_n^t(\omega))$ (i.e. $A_n \cdots A_1 = {}^t({}^tA_1 \cdots {}^tA_n)$) et ce dernier tend vers l'infini.

4.2.2. *Lemme.* — *Soit (M_n) une suite de matrices dans $\mathrm{SL}_2^{\pm}$ telle que $(M_n)_\star m \to \delta_z$ pour une certaine mesure diffuse m sur $\mathbb{P}^1$. Alors $\|M_n\| = \|{}^tM_n\| \to \infty$ et pour tout vecteur unitaire X on a*

$$\frac{\|{}^tM_nX\|}{\|{}^tM_n\|} \underset{n\to\infty}{\longrightarrow} |\langle X, Z\rangle| \,,$$

où Z est un relevé unitaire de z.

Grâce au lemme on voit que si $V \notin z(\omega)^\perp$ on a p.s.

$$\frac{\|\ell_n(\omega)V\|}{\|\ell_n(\omega)\|} = \frac{\|{}^t(r_n^t(\omega))\, V\|}{\|r_n^t(\omega)\|} \underset{n\to\infty}{\longrightarrow} |\langle V, Z(\omega)\rangle| \neq 0.$$

Ainsi comme μ est diffuse on conclut bien que pour presque tout v on a $v \notin z(\omega)^\perp$) et donc

$$\sigma\left(\ell_n(\omega), v\right) \asymp \|\ell_n(\omega)\| \to \infty,$$

ce qui était le résultat souhaité.

Démonstration du lemme. — On sait déjà que $\|M_n\| = \|{}^tM_n\| \to \infty$. Quitte à extraire une sous suite on peut supposer que $\frac{M_n}{\|M_n\|}$ tend vers B, qui est donc une matrice non-nulle et de déterminant 0, donc de rang 1. Posons $\operatorname{im} B = \mathbb{R}Z$, où Z est unitaire. On vérifie simplement que Z est nécessairement un relevé de z. Soit alors W tel que (Z, W) est une base orthonormale de $\mathbb{R}^2$. Pour tout $X \in \mathbb{R}^2$ on a

$$X = \langle X, Z\rangle Z + \langle X, W\rangle W.$$

[(2)] On peut supposer que ces objets sont définis sur Ω, i.e. par exemple $r_n^t(\omega) = r_n^t(A_n) = {}^tA_1 \cdots {}^tA_n$.

Comme $W \in (\operatorname{im} B)^\perp = \ker {}^tB$ on a ${}^tBW = 0$ et donc

$$ {}^tBX = \langle X, Z\rangle {}^tBZ. $$

On a par ailleurs $\|{}^tBZ\| = 1$: en effet il existe X_0 unitaire tel que $\|{}^tBX_0\| = 1$, et en l'entrant dans l'équation précédente on voit que

$$ \|{}^tBX_0\| = \underbrace{|\langle X_0, Z\rangle|}_{\leqslant 1} \underbrace{\|{}^tBZ\|}_{\leqslant 1} = 1 $$

donc les deux termes sont égaux à 1. Finalement, revenant à la définition de B, on voit que

$$ \frac{\|{}^tM_nX\|}{\|{}^tM_n\|} \underset{n\to\infty}{\longrightarrow} \|{}^tBX\| = |\langle X, Z\rangle|\,, $$

ce qui achève la démonstration. □

4.2.3. *Deuxième étape.* — Conclusion

D'après la première étape et les calculs faits au §4.1, on a que pour presque tout (ω, v),

$$ \sigma(\ell_n(\omega), v) = \sum_{k=1}^{n} L \circ \Theta^k(\omega, v) \underset{n\to\infty}{\longrightarrow} \infty. \tag{4} $$

On a le lemme général suivant:

4.2.4. *Lemme.* — *Soit $(X, \mathcal{A}, f, m)$ un système dynamique mesuré et $\varphi \in L^1(X, m)$. Si pour presque tout $x \in X$ on a*

$$ \sum_{k=0}^{n-1} \varphi\left(f^k(x)\right) \underset{n\to\infty}{\longrightarrow} \infty $$

alors $\int \varphi\, dm > 0$.

Admettant ce lemme, la positivité de l'exposant découle alors de (3) et (4), ce qui conclut la démonstration du Théorème 1.4.

Pour comprendre l'intuition qui sous-tend ce lemme, penser à une suite (X_n) de variables aléatoires i.i.d.: par le théorème central limite, si p.s. on a $\sum_{i=0}^n X_i \to \infty$ alors $\mathbb{E}(X_0) > 0$.

Démonstration du lemme. — Remarquons d'abord que par le théorème de Birkhoff, l'inégalité $\int \varphi\, dm \geqslant 0$ est évidente: le point est de démontrer que cette intégrale est non-nulle. Posons $S_n = \sum_{j=1}^n \varphi \circ f^j$ et pour tout $\varepsilon > 0$ on définit

$$ A_\varepsilon = \{x \in X,\ \forall n \geqslant 1,\ S_n(x) > \varepsilon\} \text{ et } B_\varepsilon = \bigcup_{k\geqslant 0} f^{-k}(A_\varepsilon). $$

Affirmation. — Si $S_n(x) \to \infty$ alors il existe $\varepsilon > 0$ tel que $x \in B_\varepsilon$.

En effet supposons par l'absurde qu'il existe x tel que $S_n(x) \to \infty$ et pour tout $\ell \geqslant 1$, $x \notin B_{1/\ell^2}$. Ainsi

$$ \forall k \geqslant 0,\ \forall \ell \geqslant 1,\ f^k(x) \notin A_{1/\ell^2} $$

ou ce qui revient au même

$$\forall k \geqslant 0,\ \forall \ell \geqslant 1,\ \exists n \geqslant 1,\ S_n(f^k(x)) \leqslant \frac{1}{\ell^2}.$$

On définit alors par récurrence une suite (n_ℓ) par $n_0 = 0$ et

$$n_\ell = \min\left\{m > n_{\ell-1},\ S_m\left(f^{n_{\ell-1}+\cdots+n_0}(x)\right) \leqslant \frac{1}{\ell^2}\right\}.$$

Ainsi par construction on a pour tout ℓ,

$$S_{n_\ell+\cdots+n_0}(x) \leqslant \sum_{j=1}^{\infty} \frac{1}{j^2} < \infty,$$

d'où la contradiction, et notre affirmation est démontrée.

L'affirmation ci-dessus implique que $m(B_\varepsilon) > 0$ pour ε assez petit, et donc également $m(A_\varepsilon) > 0$. Fixons ε satisfaisant cette propriété. Soit τ_ε le temps de séjour moyen dans A_ε, i.e.

$$\tau_\varepsilon(x) = \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \mathbf{1}_{A_\varepsilon}(f^k(x)),$$

qui est bien défini p.p. et vérifie $\int \tau_\varepsilon\, dm = m(A_\varepsilon) > 0$. Posons maintenant

$$\psi(x) = \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \varphi\left(f^k(x)\right)$$

qui existe p.p. par le théorème de Birkhoff. On a $\psi(x) \geqslant 0$ p.s. par hypothèse et $\int \psi\, dm = \int \varphi\, dm$.

Pour $x \in B_\varepsilon$, soit k_1 le plus petit entier tel que $f^k(x) \in A_\varepsilon$. On a

$$\sum_{j=0}^{n-1} \varphi \circ f^j(x) = \sum_{j=0}^{k_1-1} \varphi \circ f^j(x) + \sum_{j=k_1}^{n-1} \varphi \circ f^j(x),$$

où par définition de A_ε le deuxième terme est supérieur ou égal à ε. Soit alors k_2 le plus petit entier strictement supérieur à $k > k_1$ tel que $f^k(x) \in A_\varepsilon$. Il est fini p.s. car $S_n(f^{k_1}(x)) \to \infty$. On a alors pour $n > k_2$,

$$\sum_{j=0}^{n-1} \varphi \circ f^j(x) = \sum_{j=0}^{k_1-1} \varphi \circ f^j(x) + \underbrace{\sum_{j=k_1}^{k_2-1} \varphi \circ f^j(x)}_{\geqslant \varepsilon} + \underbrace{\sum_{j=k_2}^{n-1} \varphi \circ f^j(x)}_{\geqslant \varepsilon},$$

et ainsi de suite on découpe la somme de Birkhoff en morceaux et on trouve que

$$\sum_{j=0}^{n-1} \varphi \circ f^j(x) \geqslant \sum_{j=0}^{k_1-1} \varphi \circ f^j(x) + \varepsilon \sum_{j=k_1}^{n-1} \mathbf{1}_{A_\varepsilon}(f^j(x)).$$

En divisant par n et faisant tendre n vers l'infini il vient $\psi \geqslant \varepsilon\tau_\varepsilon$, et donc

$$\int \psi\, dm \geqslant \varepsilon \int \tau_\varepsilon\, dm > 0,$$

ce qui était le résultat souhaité. □

◇

VIII. ENTROPIE D'UNE MESURE INVARIANTE

Il y a plusieurs notions d'entropie en systèmes dynamiques, qui sont toutes inspirées de la notion d'entropie de Shannon en théorie de l'information [(1)], et qui visent à mesurer la "complexité" d'un système dynamique.

Dans ce chapitre nous allons aborder la notion d'**entropie mesurée** (ou **métrique**), introduite par Kolmogorov dans les années 50 (on la désigne également sous le nom d'**entropie de Kolmogorov-Sinai**) notamment dans le but de répondre à la question suivante: le décalage sur 2 symboles, muni de sa mesure équilibrée, est-il mesurablement conjugué au décalage sur 3 symboles? Nous verrons que la réponse est, comme attendu, négative et que l'entropie métrique formalise effectivement l'idée intuitive que le décalage sur 3 symboles est plus complexe que le décalage sur 2 symboles. (Exercice: montrer qu'il ne peut pas y avoir de conjugaison topologique.)

1. Information et entropie d'une partition

Considérons un espace probabilisé $(X, \mathcal{A}, \mu)$ (qui sera muni d'un système dynamique un peu plus loin).

1.1. — Une **partition mesurable finie** est une collection finie $\mathcal{P}$ de sous-ensembles mesurables tels que:

- pour tous P_1, P_2 dans $\mathcal{P}$, $\mu(P_1 \cap P_2) = 0$);
- $\mu\left(\bigcup_{P \in \mathcal{P}} P\right) = 1.$

Les éléments de $\mathcal{P}$ s'appellent les **atomes** (ou pièces) de la partition et pour tout $x \in X$ on note $\mathcal{P}(x)$ l'atome contenant x (il y a une petite ambiguïté car les atomes ne sont pas disjoints, mais celle-ci est négligeable dans les arguments qui vont suivre). La **partition triviale** est celle constituée de l'unique pièce X. Dans toute la suite, les partitions seront considérées "mod. 0", c'est à dire que deux partitions $\mathcal{P}$ et $\mathcal{Q}$ seront identifiées s'il existe un sous-ensemble A de mesure nulle tel que $\mathcal{P}$ et $\mathcal{Q}$ coïncident sur $X \setminus A$.

Si $\mathcal{P}$ et $\mathcal{Q}$ sont deux partitions mesurables finies, on dit que $\mathcal{Q}$ est plus fine que $\mathcal{P}$ (noté $\mathcal{P} \leqslant \mathcal{Q}$) si tout atome de $\mathcal{Q}$ est contenu dans un atome de $\mathcal{P}$.

La **partition jointe** $\mathcal{P} \vee \mathcal{Q}$ est la partition dont les pièces sont les $P \cap Q$, $(P, Q) \in \mathcal{P} \times \mathcal{Q}$. C'est la partition la moins fine qui est à la fois plus fine que $\mathcal{P}$ et $\mathcal{Q}$.

[(1)] La légende raconte que vers 1940, Shannon aurait demandé à Von Neumann au cours d'une conversation comment nommer le nouveau concept qu'il venait de découvrir, ce à quoi Shannon répondit: "You should call it 'entropy' for two reasons: first, the function is already used in thermodynamics under that name; second, and more importantly, most people don't know what entropy really is (...)."

Deux partitions sont **indépendantes** (noté $\mathcal{P} \perp \mathcal{Q}$) si pour tout $(P, Q) \in \mathcal{P} \times \mathcal{Q}$ on a $\mu(P \cap Q) = \mu(P)\mu(Q)$.

1.2. *Remarque.* — Dans tout ce chapitre, les partitions mesurables considérées seront systématiquement finies, et nous n'utiliserons pas le formalisme des partitions mesurables générales du § IV.4. Ce formalisme a été introduit par Rokhlin justement en lien avec des questions d'entropie. On pourra consulter le chapitre 2 du livre de Przytycki et Urbanski [**12**] pour quelques résultats dans cet esprit, notamment la belle formule de Rokhlin liant entropie et jacobien.

1.3. — Du point de vue de la théorie de l'information, on se donne un espace X représentant l'ensemble des états d'un certain système, et on fait une "expérience" qui le subdivise en plusieurs sous-ensembles d'états possibles $C_1, \ldots, C_n$. L'**information** de l'évènement C est par définition $I(C) = -\log \mu(C)$. Le choix de cette définition est justifié par les axiomes suivants:

- l'information est une quantité positive;
- l'information donnée par l'évènement trivial X est 0;
- plus un évènement est improbable, plus l'information donnée par cet évènement est grande[(2)];
- si C et D sont indépendants alors $I(C \cap D) = I(C) + I(D)$.

Étant donnée une partition mesurable finie $\mathcal{P}$, on peut définir une fonction information par

$$I(\mathcal{P}) : x \mapsto I(\mathcal{P}(x)) = -\sum_{P \in \mathcal{P}} \mathbf{1}_P \log \mu(P).$$

1.4. *Définition.* — L'**entropie** d'une partition mesurable finie $\mathcal{P}$ est la moyenne de $I(\mathcal{P})$, ou encore l'information moyenne des atomes de $\mathcal{P}$:

$$H(\mathcal{P}) = \int I(\mathcal{P}(x))\, \mathrm{d}\mu(x) = \int -\log(\mathcal{P}(x))\, \mathrm{d}\mu(x) = -\sum_{P \in \mathcal{P}} \mu(P) \log \mu(P)$$

(on a adopté ici la convention $0 \log 0 = 0$).

1.5. *Proposition.* — *L'entropie vérifie les propriétés suivantes:*

1. *$H(\mathcal{P}) \geqslant 0$ avec égalité ssi $\mathcal{P}$ est égale mod. 0 à la partition triviale.*
2. *Plus généralement si $\mathcal{P} \leqslant \mathcal{Q}$ alors $H(\mathcal{P}) \leqslant H(\mathcal{Q})$ avec égalité ssi $\mathcal{P} = \mathcal{Q}$ mod. 0.*
3. *Si $\mathcal{P}$ a n atomes alors $H(\mathcal{P}) \leqslant \log n$ avec égalité ssi tous les atomes sont de masse $1/n$;*
4. *$H(\mathcal{P} \vee \mathcal{Q}) \leqslant H(\mathcal{P}) + H(\mathcal{Q})$ avec égalité ssi $\mathcal{P} \perp \mathcal{Q}$.*

Démonstration. — Les deux premières assertions découlent directement de la formule $H(\mathcal{P}) = \int I(\mathcal{P}(x))\, \mathrm{d}\mu(x)$. Pour le 3. il faut montrer que

$$-\sum_{P \in \mathcal{P}} \mu(P) \log \mu(P) \leqslant -\log \frac{1}{n} \Leftrightarrow -\frac{1}{n} \sum_{P \in \mathcal{P}} \mu(P) \log \mu(P) \leqslant -\frac{1}{n} \log \frac{1}{n}.$$

(2) Dire: "le dé est tombé sur 6" apporte plus d'information que "le dé est tombé sur un nombre pair".

Ceci découle du lemme suivant:

1.6. *Lemme. — La fonction définie sur* $[0,1]$ *par* $\phi(x) = -x\log x$ *est positive, nulle en 0 et en 1, et strictement concave.* L'inégalité précédente s'écrit en effet

$$\frac{1}{n}\sum_{P\in\mathcal{P}}\phi(\mu(P)) \leqslant \phi\left(\frac{1}{n}\sum_{P\in\mathcal{P}}\mu(P)\right) = \phi\left(\frac{1}{n}\right),$$

ce qui est bien l'inégalité donnée par la concavité. La concavité stricte dit qu'il y a égalité ssi tous les $\mu(P)$ sont égaux, ainsi qu'affirmé.

Pour le 4. on va encore appliquer la concavité de ϕ. En effet fixons une pièce $P \in \mathcal{P}$. Cette pièce est partitionnée par celles de $\mathcal{P}\vee\mathcal{Q}$. Posons $\mathcal{P} = \{P_i\}$ avec $\mu(P_i) = a_i$, $\mathcal{Q} = \{Q_j\}$ avec $\mu(Q_j) = b_j$ et $\mathcal{P}\vee\mathcal{Q} = \{P_i\cap Q_j\}$ avec $\mu(P_i\cap Q_j) = c_{ij}$ (sans tenir compte des pièces de mesure nulle). On a $a_i = \sum_j c_{ij}$ et $b_j = \sum_i c_{ij}$. Écrivons b_j sous la forme $b_j = \sum_i a_i\frac{c_{ij}}{a_i}$. Comme $\sum a_i = 1$, par la concavité de ϕ on a

$$\phi(b_j) \geqslant \sum_i a_i\phi\left(\frac{c_{ij}}{a_i}\right),$$

soit

$$\begin{aligned}-b_j\log b_j &\geqslant -\sum_i a_i\frac{c_{ij}}{a_i}\log\frac{c_{ij}}{a_i} = -\sum_i c_{ij}\log\frac{c_{ij}}{a_i}\\ &= -\sum_i c_{ij}\log c_{ij} + \sum_i c_{ij}\log a_i.\end{aligned}$$

On somme alors sur j et l'on obtient

$$-\sum_j b_j\log b_j = H(\mathcal{Q}) \geqslant -\sum_{i,j} c_{ij}\log c_{ij} + \sum_i a_i\log a_i = H(\mathcal{P}\vee\mathcal{Q}) - H(\mathcal{P}),$$

qui est l'inégalité souhaitée. Par stricte concavité, si l'égalité a lieu, c'est que les $\frac{c_{ij}}{a_i}$ sont égaux lorsque i varie, et si l'on pose $\lambda = \frac{c_{ij}}{a_i}$, on a

$$b_j = \sum_i c_{ij} = \lambda\sum_i a_i = \lambda = \frac{c_{ij}}{a_i},$$

et donc $c_{ij} = a_ib_j$: c'est l'indépendance. □

2. Entropie de Kolmogorov-Sinai

2.1. — On introduit maintenant une transformation f sur $(X,\mathcal{A},\mu)$ préservant la mesure. Si $\mathcal{P}$ est une partition mesurable finie, on pose $f^{-1}\mathcal{P} = \{f^{-1}(P), P\in\mathcal{P}\}$ et on a clairement $H(f^{-1}\mathcal{P}) = H(P)$. On introduit alors une suite de **partitions itérées**

$$\mathcal{P}^{(n)} = \mathcal{P}\vee f^{-1}(\mathcal{P})\vee\cdots\vee f^{-(n-1)}\mathcal{P}.$$

Lemme. — La suite $n\mapsto H\left(\mathcal{P}^{(n)}\right)$ *est sous additive*

Démonstration. — En effet $\mathcal{P}^{(n+m)} = \mathcal{P}^{(n)} \vee f^{-n}\mathcal{P}^{(m)}$ et donc

$$H\left(\mathcal{P}^{(n+m)}\right) \leqslant H\left(\mathcal{P}^{(n)}\right) + H\left(f^{-n}\mathcal{P}^{(m)}\right) = H\left(\mathcal{P}^{(n)}\right) + H\left(\mathcal{P}^{(m)}\right),$$

ce qu'il fallait démontrer. □

2.2. — Ainsi par sous-additivité, la suite $\frac{1}{n}H\left(\mathcal{P}^{(n)}\right)$ converge, et sa limite est par définition **l'entropie de f par rapport à $\mathcal{P}$**, notée $h_\mu(f,\mathcal{P})$ (ou plus simplement $h(f,\mathcal{P})$). Noter que cette entropie mesure le taux d'accroissement de l'information moyenne de $\mathcal{P}^{(n)}$ comme fonction de n. Finalement on pose

$$h_\mu(f) = \sup\left\{h_\mu(f,\mathcal{P}),\ \mathcal{P} \text{ partition mesurable finie}\right\}:$$

c'est par définition **l'entropie** du système dynamique mesuré $(X,\mathcal{A},\mu,f)$.

2.3. — Il découle immédiatement de ces définitions que l'entropie est un **invariant de conjugaison mesurable**. Elle est en revanche assez difficile à calculer explicitement, nous verrons quelques exemples de tels calculs au §5. Avant cela, il nous faudra développer quelques outils.

2.4. *Exercice.* — Montrer que si μ est une mesure atomique invariante alors $h_\mu(f) = 0$. (Indication: commencer par le cas d'une mesure ergodique et utiliser la Proposition 5.2 pour le cas général.)

2.5. *Exercice.* — Soit σ le décalage unilatéral sur d symboles, muni de sa mesure équilibrée μ_{eq} et $\mathcal{P}$ la partition en cylindres de profondeur 0. Montrer que $h_\mu(f,\mathcal{P}) = \log d$.

3. Information et entropie relatives

3.1. *Définition.* — Si $\mathcal{P}$ et $\mathcal{Q}$ sont deux partitions mesurables finies, on définit l'**information de $\mathcal{Q}$ relativement à $\mathcal{P}$** par

$$I(\mathcal{P}|\mathcal{Q}) = -\sum_{P\in\mathcal{P}} \mathbf{1}_P \log\mu(P|\mathcal{Q}) = -\sum_{P\in\mathcal{P},Q\in\mathcal{Q}} \mathbf{1}_{P\cap Q}\log\frac{\mu(P\cap Q)}{\mu(Q)}.$$

C'est une fonction sur X qui quantifie l'information **supplémentaire** qu'apporte la connaissance de $\mathcal{P}$ si on connaît déjà $\mathcal{Q}$. En particulier on a la propriété suivante:

3.2. *Lemme.* — *Si $\mathcal{P}$ et $\mathcal{Q}$ sont indépendantes, alors $I(\mathcal{P}|\mathcal{Q}) = I(\mathcal{P})$.*

La démonstration est immédiate.

3.3. *Lemme.* — *Si $\mathcal{P}$, $\mathcal{Q}$, $\mathcal{R}$ sont trois partitions mesurables finies, on a*

$$I(\mathcal{P}\vee\mathcal{Q}|\mathcal{R}) = I(\mathcal{Q}|\mathcal{R}) + I(\mathcal{P}|\mathcal{Q}\vee\mathcal{R}).$$

Démonstration. — Le membre de gauche est par définition égal à

$$-\sum_{(P,Q,R)\in\mathcal{P}\times\mathcal{Q}\times\mathcal{R}} \mathbf{1}_{P\cap Q\cap R} \log \frac{\mu(P\cap Q\cap R)}{\mu(R)},$$

il suffit d'écrire

$$\log \frac{\mu(P\cap Q\cap R)}{\mu(R)} = \log \frac{\mu(Q\cap R)}{\mu(R)} + \log \frac{\mu(P\cap Q\cap R)}{\mu(Q\cap R)}$$

et de réarranger la somme pour obtenir le résultat. □

3.4. — On définit alors l'**entropie relative** $H(\mathcal{P}|\mathcal{Q})$ par $H(\mathcal{P}|\mathcal{Q}) = \int I(\mathcal{P}|\mathcal{Q})\mathrm{d}\mu$, expression qui se développe en

$$\begin{aligned}
H(\mathcal{P}|\mathcal{Q}) - \sum_{P\in\mathcal{P},\ Q\in\mathcal{Q}} \mu(P\cap Q) \log \frac{\mu(P\cap Q)}{\mu(Q)} \\
= -\sum_{Q\in\mathcal{Q}} \mu(Q) \sum_{P\in\mathcal{P}} \frac{\mu(P\cap Q)}{\mu(Q)} \log \frac{\mu(P\cap Q)}{\mu(Q)} \\
= -\sum_{Q\in\mathcal{Q}} \mu(Q) \sum_{P\in\mathcal{P}} \mu(P|Q) \log \mu(P|Q).
\end{aligned}$$

3.5. *Proposition. — On a les propriétés suivantes:*

1. *$H(\mathcal{P})$ est l'entropie relative à la partition triviale;*
2. *$H(\mathcal{P}|\mathcal{Q}) \geqslant 0$ avec égalité ssi $\mathcal{P} \leqslant \mathcal{Q}$;*
3. *$H(\mathcal{P} \vee \mathcal{Q}) = H(\mathcal{Q}) + H(\mathcal{P}|\mathcal{Q})$;*
4. *Si $\mathcal{Q} \leqslant \mathcal{R}$, $H(\mathcal{P}|\mathcal{Q}) \geqslant H(\mathcal{P}|\mathcal{R})$;*
5. *$H(\mathcal{P} \vee \mathcal{Q}|\mathcal{R}) = H(\mathcal{P}|\mathcal{Q} \vee \mathcal{R}) + H(\mathcal{Q}|\mathcal{R})$;*
6. *$H(\mathcal{P}|\mathcal{Q}) \leqslant H(\mathcal{P})$ avec égalité ssi $\mathcal{P} \perp \mathcal{Q}$.*

La plupart de ces propriétés sont faciles à établir. Noter que la troisième correspond bien à l'interprétation de l'information relative donnée plus haut.

Démonstration. — Les propriétés 1. et 2. sont immédiates et la 5. s'obtient directement par intégration du Lemme 3.2. La propriété 3. découle d'un calcul direct sur les fonctions information:

$$\begin{aligned}
I(\mathcal{P} \vee \mathcal{Q}) &= -\sum_{P\in\mathcal{P},Q\in\mathcal{Q}} \mathbf{1}_{P\cap Q} \log \mu(P\cap Q) \\
&= -\sum_{P\in\mathcal{P},Q\in\mathcal{Q}} \mathbf{1}_{P\cap Q} \log \mu(Q) \ - \sum_{P\in\mathcal{P},Q\in\mathcal{Q}} \mathbf{1}_{P\cap Q} \log \frac{\mu(P\cap Q)}{\mu(Q)} \\
&= -\sum_{Q\in\mathcal{Q}} \mathbf{1}_{P\cap Q} \log \mu(Q) \ - \sum_{P\in\mathcal{P},Q\in\mathcal{Q}} \mathbf{1}_{P\cap Q} \log \frac{\mu(P\cap Q)}{\mu(Q)} \\
&= I(\mathcal{Q}) + I(\mathcal{P}|\mathcal{Q}).
\end{aligned}$$

Pour la propriété 4. on va utiliser la concavité de ϕ comme précédemment. Il suffit de travailler dans une pièce $Q \in \mathcal{Q}$, et on doit démontrer que

$$-\mu(Q) \sum_{P\in\mathcal{P}} \mu(P|Q) \log \mu(P|Q) \geqslant -\sum_k \mu(R_k) \sum_{P\in\mathcal{P}} \mu(P|R_k) \log \mu(P|R_k),$$

où les R_k sont les pièces de $\mathcal{R}$ décomposant Q (qui on le rappelle, est plus fine que $\mathcal{P}$). On fixe alors une pièce $P \in \mathcal{P}$ et on veut démontrer que

$$\mu(Q)\phi(\mu(P|Q)) \geqslant \sum_k \mu(R_k)\phi(\mu(P|R_k)) \text{ soit } \phi(\mu(P|Q)) \geqslant \sum_k \frac{\mu(R_k)}{\mu(Q)}\phi(\mu(P|R_k)).$$

Comme $\sum_k \frac{\mu(R_k)}{\mu(Q)} = 1$ et

$$\mu(P|Q) = \frac{\mu(P \cap Q)}{\mu(Q)} = \sum_k \frac{\mu(R_k)}{\mu(Q)} \frac{\mu(P \cap R_k)}{\mu(R_k)} = \sum_k \frac{\mu(R_k)}{\mu(Q)}\mu(P|R_k),$$

cela correspond bien à l'inégalité de concavité pour ϕ. Finalement l'inégalité du 6. découle de 1. et 4., et pour le cas d'égalité, on utilise la concavité stricte dans le raisonnement du 4: ici $Q = X$ et on a égalité si les $\mu(P|R_k)$ ne dépendent pas de k. Ainsi $\mu(P \cap R_k) = \alpha\mu(R_k)$ et en sommant sur k on trouve $\alpha = \mu(P)$, ce qu'il fallait démontrer. □

3.6. *Exercice.* — Montrer que pour un système dynamique mesuré (X, μ, f), si $\mathcal{P}$ est une partition finie on a[3] $h(f, \mathcal{P}) = \lim_{n\to\infty} H\left(\mathcal{P}|f^{-1}\mathcal{P}^{(n)}\right)$.

4. Partitions génératrices

Nous allons dans cette section introduire un formalisme qui va permettre de calculer l'entropie à l'aide de partitions particulières. On se donne comme précédemment un système dynamique mesuré $(X, \mathcal{A}, \mu, f)$.

4.1. *Définition-Proposition.* — La fonction définie par

$$\rho(\mathcal{P}, \mathcal{Q}) = H(\mathcal{P}|\mathcal{Q}) + H(\mathcal{Q}|\mathcal{P})$$

est une distance sur l'ensemble des partitions définies mod. 0. C'est par définition la **distance de Rokhlin**.

Intuitivement il est clair que $\rho(\mathcal{P}, \mathcal{Q})$ mesure une "proximité" entre $\mathcal{P}$ et $\mathcal{Q}$: si $\rho(\mathcal{P}, \mathcal{Q})$ est petit, c'est que chacune apporte peu d'information par rapport à l'autre.

Démonstration. — Déjà, il est clair que la définition est symétrique. Ensuite si $\rho(\mathcal{P}, \mathcal{Q}) = 0$ alors par la propriété 2. de la Proposition 3.5 on a $\mathcal{P} \leqslant \mathcal{Q}$ et $\mathcal{Q} \leqslant \mathcal{P}$, c'est à dire que $\mathcal{P}$ et $\mathcal{Q}$ coincident mod. 0.

---

[3] Voir [**4**, Prop. 9.3.1] pour la solution.

Pour démontrer l'inégalité triangulaire, nous allons établir l'inégalité $H(\mathcal{P}|\mathcal{R}) \leqslant H(\mathcal{P}|\mathcal{Q}) + H(\mathcal{Q}|\mathcal{R})$. La preuve repose sur la formule $H(\mathcal{P} \vee \mathcal{Q}) = H(\mathcal{Q}) + H(\mathcal{P}|\mathcal{Q})$:

$$\begin{aligned} H(\mathcal{P}|\mathcal{Q}) + H(\mathcal{Q}|\mathcal{R}) &= H(\mathcal{P} \vee \mathcal{Q}) + H(\mathcal{Q} \vee \mathcal{R}) - H(\mathcal{Q}) - H(\mathcal{R}) \\ &= H(\mathcal{P} \vee \mathcal{Q}) + H(\mathcal{R}|\mathcal{Q}) - H(\mathcal{R}) \\ &= H(\mathcal{P} \vee \mathcal{Q} \vee \mathcal{R}) \underbrace{-H(\mathcal{R}|\mathcal{P} \vee \mathcal{Q}) + H(\mathcal{R}|\mathcal{Q})}_{\geqslant 0 \text{ car } \mathcal{P}\vee\mathcal{Q} \geqslant \mathcal{Q}} - H(\mathcal{R}) \\ &\geqslant H(\mathcal{P} \vee \mathcal{Q} \vee \mathcal{R}) - H(\mathcal{R}) \\ &\geqslant H(\mathcal{P} \vee \mathcal{R}) - H(\mathcal{R}) = H(\mathcal{P}|\mathcal{R}). \end{aligned}$$

□

4.2. — On peut construire une distance entre deux partitions mesurables finies $\mathcal{P}$ et $\mathcal{Q}$ de façon plus élémentaire. Quitte à les compléter par des pièces vides (ou de mesure nulle), on peut supposer qu'elles ont le même nombre de pièces, et on pose alors

$$d(\mathcal{P}, \mathcal{Q}) = \inf_{\tau \in \mathfrak{S}(\mathcal{P},\mathcal{Q})} \sum_{P \in \mathcal{P}} \mu(P \triangle \tau(P)),$$

où la somme est indexée sur les bijections de $\mathcal{P}$ sur $\mathcal{Q}$. Une autre présentation consiste à écrire $\mathcal{P} = \{P_1, \dots, P_k\}$ et $\mathcal{Q} = \{Q_1, \dots, Q_k\}$ et alors

$$d(\mathcal{P}, \mathcal{Q}) = \inf_{\sigma \in \mathfrak{S}_k} \sum_{i=1}^{k} \mu(P_i \triangle Q_{\sigma_i}).$$

On vérifie aisément que ceci définit une distance sur l'ensemble des partitions mod. 0. Nous admettrons le résultat suivant (voir [**6**, Prop. 4.3.5]):

Proposition. — La distance d est plus fine que la distance ρ au sens suivant: pour tout $\varepsilon > 0$ et tout k il existe $\delta = \delta(\varepsilon, k) > 0$ tel que si $\mathcal{P}$ et $\mathcal{Q}$ sont des partitions mesurables à (au plus) k éléments telles que $d(\mathcal{P}, \mathcal{Q}) < \delta$, alors $\rho(\mathcal{P}, \mathcal{Q}) < \varepsilon$.

4.3. — La distance de Rokhlin est particulièrement bien adaptée à l'estimation de l'entropie grâce au résultat suivant:

Proposition (Inégalité de Rokhlin). — Si $(X, \mathcal{A}, \mu, f)$ est un système dynamique mesuré et $\mathcal{P}$ et $\mathcal{Q}$ sont deux partitions mesurables finies, on a

$$|h_\mu(f, \mathcal{P}) - h_\mu(f, \mathcal{Q})| \leqslant \rho(\mathcal{P}, \mathcal{Q}).$$

En d'autres termes, l'entropie est une fonction 1-Lipschitzienne pour la distance de Rokhlin. Cette proposition découle immédiatement du lemme suivant:

4.4. *Lemme. — Sous les hypothèses de la proposition précédente on a*

$$h_\mu(f, \mathcal{P}) \leqslant h_\mu(f, \mathcal{Q}) + H(\mathcal{P}|\mathcal{Q}).$$

En particulier si $\mathcal{P} \leqslant \mathcal{Q}$, on a $h_\mu(f, \mathcal{P}) \leqslant h_\mu(f, \mathcal{Q})$.

Noter cette dernière assertion peut être obtenue directement à partir de la définition de $h_\mu(f, \mathcal{P})$.

Démonstration du lemme. — On rappelle que $h(f,\mathcal{P}) = \lim_{n\to\infty} H(\mathcal{P}^{(n)})$, où $\mathcal{P}^{(n)} = \mathcal{P} \vee \cdots \vee f^{-(n-1)}\mathcal{P}$. Pour démontrer l'inégalité, nous écrivons tout d'abord

$$H\left(\mathcal{P}^{(n)}\right) \leqslant H\left(\mathcal{P}^{(n)} \vee \mathcal{Q}^{(n)}\right) = H\left(\mathcal{Q}^{(n)}\right) + H\left(\mathcal{P}^{(n)}|\mathcal{Q}^{(n)}\right),$$

puis estimons le terme $H\left(\mathcal{P}^{(n)}|\mathcal{Q}^{(n)}\right)$ en appliquant successivement la relation $H(\mathcal{P} \vee \mathcal{Q}|\mathcal{R}) = H(\mathcal{P}|\mathcal{Q} \vee \mathcal{R}) + H(\mathcal{Q}|\mathcal{R})$:

$$\begin{aligned}
H\left(\mathcal{P}^{(n)}|\mathcal{Q}^{(n)}\right) &= H\left(\mathcal{P} \vee f^{-1}\mathcal{P}^{(n-1)}|\mathcal{Q}^{(n)}\right) \\
&= H\left(\mathcal{P}|\mathcal{Q}^{(n)}\right) + H\left(f^{-1}\mathcal{P}^{(n-1)}|\mathcal{P} \vee \mathcal{Q}^{(n)}\right) \\
&\leqslant H(\mathcal{P}|\mathcal{Q}) + H\left(f^{-1}\mathcal{P}^{(n-1)}|f^{-1}\mathcal{Q}^{(n-1)}\right) \\
&\qquad \text{car } \mathcal{Q}^{(n)} \geqslant \mathcal{Q} \text{ et } \mathcal{P} \vee \mathcal{Q}^{(n)} \geqslant \mathcal{Q}^{(n)} \geqslant f^{-1}\mathcal{Q}^{(n-1)} \\
&= H(\mathcal{P}|\mathcal{Q}) + H\left(\mathcal{P}^{(n-1)}|\mathcal{Q}^{(n-1)}\right) \\
&\leqslant nH(\mathcal{P}|\mathcal{Q}) \text{ par récurrence.}
\end{aligned}$$

Ainsi

$$\frac{1}{n}H\left(\mathcal{P}^{(n)}\right) \leqslant \frac{1}{n}H\left(\mathcal{Q}^{(n)}\right) + H(\mathcal{P}|\mathcal{Q}),$$

et on termine la preuve en faisant tendre n vers l'infini. □

4.5. — On dit qu'une suite $(\mathcal{P}_k)$ de partitions est **génératrice** si elle est asymptotiquement plus fine que toute partition finie, c'est à dire que pour toute partition finie $\mathcal{Q}$ et tout $\delta > 0$, il existe k_0 tel que pour $k \geqslant k_0$, il existe $\mathcal{P}' \leqslant \mathcal{P}_k$ tel que $d(\mathcal{P}', \mathcal{Q}) < \delta$.

Un exemple simple de suite génératrice est la suite de partitions

$$\mathcal{P}_k = \{[j/k, (j+1)/k], 0 \leqslant j \leqslant k-1\}$$

sur $[0,1]$ muni de sa mesure de Lebesgue. Cet exemple montre qu'*un espace probabilisé standard admet une suite génératrice.* Plus généralement on a le résultat suivant qui découle essentiellement de la régularité des mesures finies dans un espace localement compact:

Proposition. — Soit μ une mesure de probabilité borélienne sur un espace métrique localement compact. Alors une suite $(\mathcal{P}_k)$ de partitions finies vérifiant

$$\sup_{P\in\mathcal{P}_k} \operatorname{diam}(P) \underset{k\to\infty}{\longrightarrow} 0$$

est génératrice.

La démonstration est laissée en exercice.

4.6. *Proposition. — Soit $(X, \mathcal{A}, \mu, f)$ est un système dynamique mesuré et $(\mathcal{P}_k)$ une suite croissante de partitions finies, qui est génératrice au sens précédent, alors*

$$h_\mu(f) = \lim_{k\to\infty} h_\mu(f, \mathcal{P}_k).$$

Démonstration. — Fixons $\varepsilon > 0$ et considérons une partition finie $\mathcal{Q}$ telle que $h_\mu(f,\mathcal{Q}) \geqslant h_\mu(f) - \varepsilon$. Soit $\delta = \delta(\varepsilon, \#\mathcal{Q})$ comme à la Proposition 4.2. Alors pour k_0 assez grand il existe $\mathcal{P}' \leqslant \mathcal{P}_{k_0}$ tel que $d(\mathcal{P}',\mathcal{Q}) \leqslant \delta$ et ainsi $\rho(\mathcal{P}',\mathcal{Q}) \leqslant \varepsilon$. En effet $\mathcal{Q}$ étant fixée, et $\mathcal{P}'$ étant formée d'unions de pièces de $\mathcal{P}_k$, quitte à fusionner à nouveau des pièces de $\mathcal{P}'$ on peut toujours supposer que $\mathcal{P}'$ a un nombre d'éléments inférieur ou égal à celui de $\mathcal{Q}$, de sorte que la Proposition 4.2 s'applique. On déduit alors de la Proposition 4.3 que $h_\mu(f,\mathcal{P}') \geqslant h_\mu(f) - 2\varepsilon$. Finalement pour tout $k \geqslant k_0$ on a $\mathcal{P}_k \geqslant \mathcal{P}_{k_0} \geqslant \mathcal{P}'$ et donc par le Lemme 4.4, $h_\mu(f,\mathcal{P}_k) \geqslant h_\mu(f) - 2\varepsilon$, ce qui termine la preuve. □

4.7. — Nous pouvons finalement mettre en évidence une classe de partitions $\mathcal{P}$ telles que $h_\mu(f,\mathcal{P}) = h_\mu(f)$.

Définition. — On dit que $\mathcal{P}$ est un **générateur** (unidirectionnel)[4] si la suite de partitions $\mathcal{P}^{(k)} = \bigvee_{i=0}^{k-1} f^{-i}\mathcal{P}$ est génératrice. Si f est inversible un générateur bidirectionnel est une partition $\mathcal{P}$ telle que la suite $\bigvee_{i=-k}^{k} f^{-i}\mathcal{P}$ est génératrice.

On peut alors résumer la discussion précédente dans l'énoncé suivant:

4.8. *Théorème (Kolmogorov-Sinai). — Soit $(X,\mathcal{A},\mu,f)$ un système dynamique mesuré. Si $\mathcal{P}$ est un générateur (resp. un générateur bidirectionnel dans le cas où f est inversible) alors*

$$h_\mu(f) = h_\mu(f,\mathcal{P}).$$

Démonstration. — Dans le cas unidirectionnel le théorème découle de la proposition précédente, modulo l'observation suivante: pour tout n on a $h_\mu(f,\mathcal{P}) = h_\mu\left(f,\mathcal{P}^{(k)}\right)$. En effet

$$h_\mu\left(f,\mathcal{P}^{(k)}\right) = \lim_{n\to\infty}\frac{1}{n}H\left(\mathcal{P}^{(k)} \vee \ldots \vee f^{-n}\mathcal{P}^{(k)}\right) = \lim_{n\to\infty}\frac{1}{n}H\left(\mathcal{P}^{(n+k)}\right) = h_\mu\left(f,\mathcal{P}\right).$$

Dans le cas bidirectionnel l'argument est analogue. Si on pose $\widetilde{\mathcal{P}}^{(k)} = \bigvee_{i=-(k-1)}^{k-1} f^{-i}\mathcal{P}$, on obtient

$$H\left(\widetilde{\mathcal{P}}^{(k)} \vee f^{-1}\widetilde{\mathcal{P}}^{(k)} \vee \ldots f^{-(n-1)}\widetilde{\mathcal{P}}^{(k)}\right) = H\left(f^{-(k-1)}\left(\mathcal{P}^{(n+2k)}\right)\right) = H\left(\mathcal{P}^{(n+2k)}\right)$$

et on conclut comme dans le cas précédent. □

Remarque. — On peut se demander pourquoi les cas inversible et non-inversible ont été traités séparément. La raison est qu'en général un système dynamique inversible n'admet **pas** de générateur unidirectionnel. Plus précisément: *un système inversible admettant un générateur unidirectionnel est d'entropie nulle* (voir [**6**, Prop. 4.3.15]).

[4] On parle aussi parfois simplement de partition génératrice, attention donc au conflit avec la terminologie précédente.

5. Calculs d'entropie

5.1. *Proposition. — Soient* $(X, \mathcal{A}, \mu, f)$ *et* $(Y, \mathcal{B}, \nu, g)$ *des systèmes dynamiques mesurés.*

1. *Pour* $q \in \mathbb{N}$, $h_\mu(f^q) = qh_\mu(f)$ *et si* f *est inversible,* $h_\mu(f^{-1}) = h_\mu(f)$.
2. *Si* (f, μ) *est un facteur de* (g, ν) *alors* $h_\mu(f) \leqslant h_\nu(g)$.
3. $h_{\mu\times\nu}(f \otimes g) = h_\mu(f) + h_\nu(g)$.

Démonstration. — On a vu dans la démonstration du théorème de Kolmogorov-Sinai que pour toute partition finie on a $h_\mu(f, \mathcal{P}) = h_\mu\left(f, \mathcal{P}^{(k)}\right)$. En outre

$$h_\mu\left(f^q, \mathcal{P}^{(q)}\right) = \lim_{n\to\infty} \frac{1}{n} H_\mu\left(\bigvee_{i=0}^{n-1} (f^q)^{-i}\mathcal{P}^{(q)}\right) = \lim_{n\to\infty} \frac{1}{n} H_\mu\left(f, \bigvee_{i=0}^{qn-1} \mathcal{P}\right) = qh_\mu(f, \mathcal{P}).$$

En passant au sup sur $\mathcal{P}$ on obtient que $h_\mu(f^q) \geqslant qh_\mu(f)$. Pour l'inégalité inverse, on note que

$$h_\mu\left(f^q, \mathcal{P}\right) \leqslant h_\mu\left(f^q, \bigvee_{i=0}^{q-1} \mathcal{P}\right) = qh_\mu(f, \mathcal{P}),$$

où la première inégalité vient du fait qu'on a raffiné la partition et la deuxième égalité est un calcul similaire au précédent. On conclut ainsi que $h_\mu(f^q) = qh_\mu(f)$.

Pour l'entropie de l'inverse, on remarque que

$$\bigvee_{i=0}^{n-1} f^{-i}\mathcal{P} = f^{-(n-1)} \bigvee_{i=0}^{n-1} f^i\mathcal{P}.$$

Ainsi

$$H\left(\bigvee_{i=0}^{n-1} f^{-i}\mathcal{P}\right) = H\left(\bigvee_{i=0}^{n-1} f^i\mathcal{P}\right)$$

et le résultat suit.

Pour la relation 2. soit $\pi : Y \to X$ l'application mesurable semi-conjuguant les dynamiques de f et g. Si $\mathcal{P}$ est une partition mesurable finie sur X on a une partition induite $\pi^{-1}\mathcal{P}$ sur Y, et il est clair que $h_\mu(f, \mathcal{P}) = h_\nu(g, \pi^{-1}\mathcal{P})$. En passant au sup sur $\mathcal{P}$ on obtient $h_\mu(f) = \sup_{\mathcal{P}} h_\nu(g, \pi^{-1}\mathcal{P}) \leqslant h_\nu(g)$.

Pour démontrer la relation 3. on note π_X et π_Y les projections naturelles de $X \times Y$ sur X et Y et on remarque que si $\mathcal{P}$ et $\mathcal{Q}$ sont des partitions respectives sur X et Y, les partitions $\pi_X^{-1}\mathcal{P}$ et $\pi_Y^{-1}\mathcal{Q}$ sur $(X \times Y, \mu \times \nu)$ sont indépendantes. En outre $\pi_X^{-1}\mathcal{P} \vee \pi_Y^{-1}\mathcal{Q}$ est la partition en pièces de la forme $\pi_X^{-1}P \times \pi_X^{-1}Q$, que l'on peut noter $\mathcal{P} \times \mathcal{Q}$. Par indépendance, on voit que

$$H_{\mu\times\nu}(\mathcal{P} \times \mathcal{Q}) = H_{\mu\times\nu}(\pi_X^{-1}\mathcal{P}) + H_{\mu\times\nu}(\pi_Y^{-1}\mathcal{Q}) = H_\mu(\mathcal{P}) + H_\nu(\mathcal{Q})$$

et de même pour les partitions itérées. Pour conclure, supposons pour simplifier que $(X, \mathcal{A}, \mu)$ et $(Y, \mathcal{B}, \nu)$ sont des espaces de Lebesgue, de sorte qu'on dispose de suites génératrices de partitions $(\mathcal{P}_k)$ et $(\mathcal{Q}_k)$, et on vérifie alors que $(\mathcal{P}_k \times \mathcal{Q}_k)$ est une suite génératrice dans $(X \times Y, \mathcal{A} \otimes \mathcal{B}, \mu \times \nu)$, et le résultat suit. □

5.2. *Proposition. — L'application entropie est affine sur le convexe des probabilités invariantes. En clair, si μ et ν sont deux mesures de probabilité f-invariantes*

$$h_{\alpha\mu+(1-\alpha)\nu}(f) = \alpha h_\mu(f) + (1-\alpha)h_\nu(f).$$

Nous allons démontrer cette proposition dans le cas particulier où les mesures μ et ν sont mutuellement singulières –ce qui suffit donc pour la décomposition ergodique–, le cas général est plus délicat (cf. [**14**]). La démonstration repose sur le lemme suivant:

Lemme. — Si A est un ensemble invariant non-trivial, alors

$$h_\mu(f) = \mu(A)h_{\mu_A}(f) + \mu(X\backslash A)h_{\mu_{X\backslash A}}(f)$$

(où l'on a noté $\mu_A = \mu(\cdot|A)$ et $\mu_{X\backslash A} = \mu(\cdot|X\backslash A)$).

Démonstration. — Soit $\mathcal{P}$ une partition finie et $\mathcal{Q}$ la partition $\{A, X\backslash A\}$. Remarquer que comme A est invariant on a $\mathcal{P}^{(n)} \vee \mathcal{Q} = (\mathcal{P} \vee \mathcal{Q})^{(n)}$. Ainsi

$$\begin{aligned}
H_\mu\left((\mathcal{P}\vee\mathcal{Q})^{(n)}\right) &= -\sum_{P\in\mathcal{P}^{(n)}} \mu(P\cap A)\log\mu(P\cap A) - \sum_{P\in\mathcal{P}^{(n)}} \mu(P\cap(X\backslash A))\log\mu(P\cap(X\backslash A)) \\
&= -\mu(A)\sum_{P\in\mathcal{P}^{(n)}} \mu_A(P)\log\mu_A(P) - \mu(X\backslash A)\sum_{P\in\mathcal{P}^{(n)}} \mu_{X\backslash A}(P)\log\mu_{X\backslash A}(P) \\
&\qquad - \mu(A)\log\mu(A) - \mu(X\backslash A)\log\mu(X\backslash A) \\
&= \mu(A)H_{\mu_A}\left(\mathcal{P}^{(n)}\right) + \mu(X\backslash A)H_{\mu_{X\backslash A}}\left(\mathcal{P}^{(n)}\right) + H_\mu(\mathcal{Q})
\end{aligned}$$

En divisant par n et en faisant tendre n vers l'infini on en déduit

$$h_\mu(f,\mathcal{P}\vee\mathcal{Q}) = \mu(A)h_{\mu_A}(f,\mathcal{P}) + \mu(X\backslash A)h_{\mu_{X\backslash A}}(f,\mathcal{P}),$$

et finalement, comme toute paire de partitions de $(A, X\backslash A)$ donne une partition de X, on peut prendre le sup sur $\mathcal{P}$ indépendamment dans les deux termes de droite, et on obtient l'égalité souhaitée. □

Démonstration de la proposition. — On fixe $A \subset X$ tel que $\mu(A) = \nu(X\backslash A) = 1$, alors pour tout $\alpha \in]0,1[$ on a $(\alpha\mu + (1-\alpha)\nu)_A = \mu$ et $(\alpha\mu + (1-\alpha)\nu)_{X\backslash A} = \nu$, et le résultat découle du lemme précédent. □

5.3. — On dit qu'une application continue sur un espace métrique compact est **positivement expansive** s'il existe $\delta > 0$ tel que pour tous $x \neq y$ dans X, il existe $n \geqslant 0$ tel que $d(f^n(x), f^n(y)) \geqslant \delta$. Cette définition n'a de sens que si f est non-inversible: en effet si f est un homéomorphisme, pour tout $\varepsilon > 0$ il existe $x \neq y$ tel que $d(f^n(x), f^n(y)) < \varepsilon$ pour tout $n \geqslant 0$ (voir [**4**, Prop. 2.4.1]). Ainsi pour la définition d'un **homéomorphisme expansif** on demande l'existence d'un $n \in \mathbb{Z}$ tel que $d(f^n(x), f^n(y)) \geqslant \delta$.

Exercice. — Montrer que le décalage unilatéral sur $d \geqslant 2$ symboles et la multiplication par $m \geqslant 2$ sur le cercle sont des applications (positivement) expansives. Montrer que l'application induite par $\left(\begin{smallmatrix}2&1\\1&1\end{smallmatrix}\right)$ sur $\mathbb{T}^2$ est un homéomorphisme expansif.

5.4. *Proposition. — Si f est une application positivement expansive (resp. un homéomorphisme expansif) sur un espace métrique compact X, préservant une mesure borélienne*

μ, et si $\mathcal{P}$ est une partition mesurable finie en pièces de diamètre strictement inférieur à la constante d'expansivité δ, alors $\mathcal{P}$ est un générateur unilatéral (resp. bilatéral) et donc $h_\mu(f,\mathcal{P}) = h_\mu(f)$.

Démonstration. — Nous traiterons le cas unilatéral; le cas inversible est analogue. Remarquons d'abord que par compacité, pour tout $\varepsilon > 0$, il existe n_ε tel que si $d(x,y) \geqslant \varepsilon$, il existe $0 \leqslant n \leqslant n_\varepsilon$ tel que $d(f^n(x), f^n(y)) \geqslant \delta$. On en déduit que si les pièces de la partition $\mathcal{P}$ sont de diamètre strictement inférieur à δ, les pièces de $\mathcal{P} \vee f^{-1}\mathcal{P} \vee \dots f^{-n_\varepsilon}\mathcal{P}$ sont de diamètre inférieur à ε. Ainsi par la Proposition 4.5, $\mathcal{P}$ est un générateur, et on conclut. □

5.5. — On peut ainsi calculer l'**entropie du décalage** sur d symboles $\sigma : \Sigma_d^+ \to \Sigma_d^+$ (resp. du décalage bilatéral $\sigma : \Sigma_d \to \Sigma_d$) muni de sa mesure équilibrée μ_{eq}. En effet la partition $\mathcal{P}$ en cylindres de profondeur 0 est un générateur (resp. un générateur bilatéral), et $\mathcal{P}^{(n)}$ est la partition en cylindres de profondeur $n-1$. Ainsi pour toute pièce $P \in \mathcal{P}^{(n)}$ on a $\mu_{\mathrm{eq}}(P) = \frac{1}{d^n}$ et

$$H_{\mu_{\mathrm{eq}}}\left(\mathcal{P}^{(n)}\right) = -d^n \frac{1}{d^n} \log \frac{1}{d^n} = n \log d$$

et finalement $h_{\mu_{\mathrm{eq}}}(\sigma) = \log d$. Remarquer que nous avons résolu le problème posé en introduction de ce chapitre:

Corollaire. — Si $d \neq d'$ les décalages sur d et d' symboles, munis de leurs mesures équilibrées, ne sont pas mesurablement conjugués.

5.6. *Exercice.* — En utilisant la description géométrique faite au §3.4 du chapitre I, montrer que l'entropie de l'application d'Arnol'd $\left(\begin{smallmatrix} 2 & 1 \\ 1 & 1 \end{smallmatrix}\right)$ sur $\mathbb{T}^2$ relativement à la mesure de Haar vaut $\log \frac{3+\sqrt{5}}{2}$.

5.7. *Théorème. — Soit μ une mesure invariante pour le décalage σ sur Σ_d^+. Alors $h_\mu(\sigma) \leqslant \log d$, avec égalité si et seulement si μ est la mesure équilibrée.*

Ainsi μ_{eq} admet une caractérisation dynamique naturelle: c'est l'**unique mesure d'entropie maximale** de σ.

Démonstration. — On conserve les notations du § 5.5. Si μ n'est pas équilibrée, c'est qu'il existe un cylindre P de profondeur k tel que $\mu(P) \neq d^{-k}$. Par la Proposition 1.5.3, on en déduit que $H_\mu(\mathcal{P}^{(k)}) < \log d^k$. Mais par ailleurs comme la partition $\mathcal{P}$ en cylindres est génératrice, on a

$$h_\mu(f) = \lim_{n\to\infty} \frac{1}{nk} H_\mu(\mathcal{P}^{(nk)}) \leqslant \frac{1}{k} H_\mu(\mathcal{P}^{(k)})$$

par sous-additivité (Lemme 2.1), et donc $h_\mu(f) < \log d$, comme annoncé. □

5.8. *Théorème. — Si f est un homéomorphisme du cercle ou de l'intervalle, muni d'une mesure invariante μ, alors $h_\mu(f) = 0$.*

Ceci est une autre manifestation du fait qu'il n'y a pas de dynamique inversible chaotique sur le cercle (cf. le Corollaire 5.8 du chapitre IV)

Démonstration. — Nous considérons le cas du cercle, l'adaptation au cas de l'intervalle est directe. Soit $\mathcal{P}_k$ une famille emboitée de subdivisions du cercle en intervalles de longueur 2^{-k}, de sorte que la suite $(\mathcal{P}_k)$ est croissante et génératrice. D'après la Proposition 4.6 il suffit de montrer que $h_\mu(f, \mathcal{P}_k) = 0$. Mais pour tout n, la partition $f^{-n}\mathcal{P}_k$ est une partition en 2^k intervalles, de sorte que $\mathcal{P}_k^{(n)} = \mathcal{P}_k \vee \ldots \vee f^{-(n-1)}\mathcal{P}_k$ est une partition en au plus $n2^k$ intervalles (car il y a au plus $n2^k$ extrémités). Ainsi $H_\mu(\mathcal{P}_k^{(n)}) \leqslant \log(n2^k)$ et $\frac{1}{n}H_\mu(\mathcal{P}_k^{(n)})$ tend vers 0 quand $n \to \infty$. □

6. Inégalité de Ruelle

6.1. — Le cadre de cette section est la théorie ergodique des applications différentiables. Donnons nous une application C^1 d'une variété compacte M de dimension d, munie d'une mesure de probabilité invariante μ (ou bien pour simplifier une application C^1 définie au voisinage d'un compact M de $\mathbb{R}^d$, et une mesure invariante à support dans M).

Dans ce contexte on a défini au Chapitre VI le cocycle tangent, lié à la formule de dérivation des fonctions composées $Df_x^n = Df_{f^{n-1}(x)} \circ \cdots \circ Df_x$, et ses les exposants de Lyapunov $\lambda_1 \leqslant \cdots \leqslant \lambda_d$, comptés avec multiplicité, et notés χ_i quand on prend en compte les multiplicités. Plus précisément, le théorème d'Oseledets fournit pour μ presque tout x un drapeau $V_0 = \{0\} < V_1(x) < \cdots < V_r(x) = T_xM$ (ou bien $\mathbb{R}^d$) et des réels $\chi_1(x) < \cdots < \chi_r(x)$ tels que pour tout $v \in V_i(x) \backslash V_{i-1}(x)$ on a

$$\frac{1}{n} \log \|Df_x^n(v)\| \underset{n\to\infty}{\longrightarrow} \chi_i,$$

où $\|\cdot\|$ désigne la norme associée à une métrique riemannienne quelconque sur M. Lorsque μ est ergodique, les dimensions des espaces V_i, ainsi que la valeur des exposants χ_i, ne dépendent pas de x générique. On définit alors les $(\lambda_i)_{i=1,\ldots,d}$ comme étant la liste ordonnée des χ_i, répétés autant de fois que leur multiplicité, celle ci étant par définition $\dim V_i$.

6.2. *Théorème (Inégalité de Ruelle). — Si f est une application C^2 d'une variété compacte (ou une application C^2 sur un compact $M \subset \mathbb{R}^d$), μ une mesure ergodique, et les λ_i sont définis comme ci-dessus, on a*

$$h_\mu(f) \leqslant \sum_i \lambda_i^+ = \sum_i \max(\lambda_i, 0).$$

6.3. *Corollaire. — Sous les hypothèses du théorème $h_\mu(f) \leqslant d \log^+\big(\sup_M \|Df\|\big)$.*

Ce corollaire peut être démontré de manière un peu plus élémentaire, et vaut en fait pour les applications C^1 (voire lipschitziennes). On obtient en particulier que l'entropie est finie, ce qui n'était pas évident a priori! Noter que le second membre ne dépend pas de μ, donc l'inégalité vaut pour toutes les mesures invariantes (et tous les choix de norme), et donc également pour l'entropie topologique (voir le §7.1).

6.4. — On utilise souvent l'inégalité de Ruelle de la manière suivante, qui indique que la positivité de l'entropie est associée à une forme d'expansion de la dynamique (dynamique chaotique):

Corollaire. — Sous les hypothèses du théorème, si $h_\mu(f) > 0$ alors μ admet un exposant strictement positif (ou de manière équivalente, $\chi^+ = \lambda_d > 0$).

Noter qu'inversement si tous les exposants de Lyapunov sont strictement négatifs, alors μ est concentrée sur une orbite périodique: la preuve a été donnée en dimension 1 au Chapitre IV et peut être étendue en dimension arbitraire avec des méthodes similaires.

Le corollaire peut être également formulé pour une mesure non ergodique:

Corollaire. — Si f est comme dans le théorème et μ est une mesure non nécessairement ergodique telle que $h_\mu(f) > 0$ alors il existe x tel que $\chi^+(x) = \lambda_d(x) > 0$.

En effet, on écrit la décomposition ergodique (ou une décomposition de Choquet) $\mu = \int \mu_\alpha d\rho(\alpha)$, et comme l'entropie est affine (Proposition 5.2), il existe α tel que $h_{\mu_\alpha}(f) > 0$ et on applique le corollaire précédent à μ_α.

6.5. *Remarque.* — En fait l'inégalité de Ruelle elle-même peut être être adaptée au cas non ergodique, soit en utilisant la décomposition ergodique comme ci-dessus, soit en remarquant que la preuve donne directement une inégalité faisant intervenir l'exposant de Lyapunov moyen :

$$h_\mu(f) \leqslant d \int \chi^+(x) \mathrm{d}\mu(x)$$

(exercice: vérifier les détails).

6.6. — Il est naturel de se demander ce qui explique la différence entre les deux membres dans l'inégalité de Ruelle, et si l'on peut caractériser le cas d'égalité. C'est l'objet du théorème de Ledrappier-Young, un autre "classique" de la dynamique différentiable[(5)]

6.7. — Rappelons que si A est une matrice $d \times d$, on peut écrire $A = U\Delta V$, où U et V sont orthogonales et Δ est une matrice diagonale à coefficients positifs ou nuls et rangés dans l'ordre décroissant (Δ est alors uniquement déterminée). C'est la décomposition KAK de A (voir le chapitre VI). Les coefficients diagonaux $\sigma_1(A) \geqslant \cdots \geqslant \sigma_d(A)$ de Δ sont par définition les valeurs singulières de A. Dans le cas du cocycle tangent, on commence par fixer une trivialisation du fibré tangent de M sur un ensemble de mesure totale pour identifier Df_x^n à une matrice $d \times d$ (voir la remarque au §2.3 du chapitre VI). On peut alors écrire la décomposition KAK. Pour démontrer l'inégalité de Ruelle nous aurons besoin de la variante suivante du théorème d'Oseledets, qui découle directement de sa démonstration: notons $\lambda_1 \geqslant \cdots \geqslant \lambda_d$ les exposants de Lyapunov de μ, répétés autant de fois que leur multiplicité. Alors presque sûrement et dans L^1 on a

$$\lim_{n\to\infty} \frac{1}{n} \log \sigma_i(Df_x^n) = \lambda_i.$$

6.8. — La preuve est basée sur un lemme géométrique élémentaire. Si F est un sous-ensemble d'un espace métrique, nous noterons classiquement F_ε l'ensemble des points situés à distance strictement inférieure à ε de F.

[(5)] Voir *The metric entropy of diffeomorphisms. I. et II.*, Ann. Math 122, pp. 509-574. Voir également le chapitre 11 de [**12**] pour une présentation en dimension 1 (complexe).

Lemme. — *Soit A une application linéaire dans $\mathbb{R}^d$, muni de sa structure euclidienne standard. Alors il existe une constante $C = C(d)$ ne dépendant que de la dimension telle que pour tout $\rho > 0$, le nombre de boules de rayon $\rho/2$ disjointes que peut rencontrer $(A(B(0, \rho)))_{2\rho}$ est majoré par*

$$C(d) \prod_{i=1}^{n} \max(\sigma_i(A), 1).$$

Démonstration. — Par continuité il suffit de montrer le résultat lorsque A est inversible. On écrit la décomposition $A = U\Delta V$, et comme U et V sont des isométries, il suffit de démontrer le résultat pour Δ. Ensuite, on remarque que le problème est homogène en ρ, donc on peut supposer $\rho = 1$. Enfin, si $(\Delta(B(0, 1)))_2$ rencontre une boule de rayon $1/2$, alors $(\Delta(B(0, 1)))_3$ la contient. Ainsi, comme toutes ces boules ont le même volume $V(d)$, il suffit donc de majorer le volume de $(\Delta(B(0, 1)))_3$ par une quantité du type $C(d) \prod_{i=1}^{n} \max(\sigma_i(A), 1)$.

Pour cela on observe que $\Delta(B(0, 1))$ est un ellipsoïde d'équation

$$\left\{ x \in \mathbb{R}^d, \sum_{i=1}^{d} \frac{x_i^2}{\sigma_i^1} = 1 \right\}.$$

En particulier si $x \in \Delta(B(0, 1))$ on a $|x_i| \leqslant \sigma_i$ pour tout i, et donc si $y \in (\Delta(B(0, 1)))_3$ on a $|y_i| \leqslant 3 + \sigma_i$ pour tout i. Ainsi le volume de $(\Delta(B(0, 1)))_3$ est majoré par $\prod_{i=1}^{d} 2(3 + \delta_i)$, et comme $3 + \sigma_i \leqslant 4 \max(\sigma_i, 1)$ le résultat suit. □

Démonstration de l'inégalité de Ruelle. — Par commodité nous supposerons dans cette démonstration que M est un compact de $\mathbb{R}^d$ et f est définie au voisinage de ce compact. L'adaptation au cas d'une variété compacte repose sur des arguments standard, les détails (un peu pénibles) sont laissés au lecteur. Fixons $n \geqslant 0$; nous allons évaluer $h_\mu(f^n) = nh_\mu(f)$ en fonction des $\sigma_i(Df^n)$. Pour ε suffisamment petit, dès que $\|x - y\| \leqslant \varepsilon$ on a

$$\frac{1}{2} \max(\sigma_i(Df_y^n), 1) \leqslant \max(\sigma_i(Df_x^n), 1) \leqslant 2\max(\sigma_i(Df_y^n), 1) \tag{1}$$

$$\text{et} \quad \|f^n(y) - f^n(x) - Df_x^n(y - x)\| < \varepsilon. \tag{2}$$

Nous allons chercher une partition adaptée. On dit qu'un sous-ensemble fini E d'un espace métrique X est ε-séparé si pour tous $x \neq y$ dans E on a $\|x - y\| > \varepsilon$.

Exercice. — Si X est compact, tout sous-ensemble ε-séparé est fini et contenu dans un ensemble ε-séparé maximal.

Exercice. — Si E un un sous-ensemble ε-séparé maximal, alors les boules $B(e, \varepsilon/2)$, $e \in E$, sont disjointes et $\bigcup_{e \in E} B(e, \varepsilon) = X$.

Fixons $\rho > 0$ tel que f^n est bien définie sur $M_{2\rho}$. Donnons nous un ensemble ε-séparé maximal E de M_ρ et considérons la partition $\mathcal{P}_E$ associée de M en “cellules de Voronoi”, dont les pièces sont définies pour $e \in E$ par

$$P_e = \left\{ x \in M, \|x - e\| < \min_{e' \in E \setminus \{e\}} d(e', x) \right\}.$$

Quitte à translater légèrement les éléments de E (c'est pour cela qu'on ne contraint pas les points de E à appartenir à M) on a $\mu\left(\bigcup_{e \in E} \partial P_e\right) = 0$ et $\mathcal{P}_E$ définit bien une partition mesurable finie.

Affirmation: pour tout $e \in E$, $f^n P_e$ rencontre au plus $C(d) \prod_{i=1}^d \max(\sigma_i(Df_e^n), 1)$ atomes de $\mathcal{P}_E$.

En effet les boules $B(e', \varepsilon/2)$, $e' \in E$ sont disjointes et pour $x \in P_e$, on a $\|x - e\| < \varepsilon$ (voir les exercices ci-dessus). Donc par (2) on a $\|f^n(x) - f^n(e) - Df_e^n(x - e)\| < \varepsilon$, soit $f^n(x) \in f^n(e) + Df_e^n(B(x, \varepsilon))$. Comme le diamètre d'une pièce de $\mathcal{P}_E$ est majoré par 2ε, on conclut que si $P_{e'}$ rencontre $f^n(P_e)$, e' est contenu dans $(f^n(e) + Df_e^n(B(x, \varepsilon)))_{2\varepsilon}$, et on conclut par le lemme géométrique 6.8.

L'affirmation précédente implique qu'il y a au plus $C(d) \prod_{i=1}^d \max(\sigma_i(Df_e^n), 1)$ atomes de $f^{-n}\mathcal{P}_E$ qui rencontrent P_e; nous allons nous servir de cette information pour majorer l'entropie conditionnelle $H(f^{-n}\mathcal{P}_E | \mathcal{P}_E)$. À cette fin rappelons que

$$H(f^{-n}\mathcal{P}_E | \mathcal{P}_E) = \sum_{e \in E} \mu(P_e) \left(- \sum_{e' \in E} \mu(f^{-n} P_{e'} | P_e) \log \mu(f^{-n} P_{e'} | P_e) \right). \tag{3}$$

Nous avons vu à la Proposition 1.5 que si q nombres μ_i vérifient $\sum \mu_i = 1$ alors $-\sum \mu_i \log \mu_i \leqslant \log q$. En appliquant cette inégalité à la somme entre parenthèses dans (3) (dont on contrôle par l'affirmation précédente le nombre de termes non nuls), on obtient

$$\begin{aligned} H(f^{-n}\mathcal{P}_E | \mathcal{P}_E) &\leqslant \sum_{e \in E} \mu(P_e) \log \left(C(d) \prod_{i=1}^d \max(\sigma_i(Df_e^n), 1) \right) \\ &\leqslant \sum_{e \in E} \mu(P_e) \left(\log C(d) + \sum_{i=1}^d \log^+ (\sigma_i(Df_e^n)) \right). \end{aligned}$$

Par (1) on a

$$\mu(P_e) \log^+ (\sigma_i(Df_e^n)) \leqslant \int_{P_e} \left(\log 2 + \log^+ (\sigma_i(Df_x^n))\right) \mathrm{d}\mu(x)$$

et finalement on obtient

$$H(f^{-n}\mathcal{P}_E | \mathcal{P}_E) \leqslant d \log 2 + \log(C(d)) + \sum_{i=1}^d \int_M \log^+ (\sigma_i(Df_x^n)) \, \mathrm{d}\mu(x).$$

D'après le Lemme 6.9 ci-dessous on a l'inégalité $h_\mu(f^n, \mathcal{P}_E) \leqslant H(f^{-n}\mathcal{P}_E | \mathcal{P}_E)$, ainsi on on peut écrire

$$\frac{1}{n} h_\mu(f^n, \mathcal{P}_E) \leqslant \frac{1}{n} H(f^{-n}\mathcal{P}_E | \mathcal{P}_E) \leqslant \frac{1}{n} \log(2^d C(d)) + \frac{1}{n} \int_M \sum_{i=1}^d \log^+ (\sigma_i(Df_x^n)) \, \mathrm{d}\mu(x).$$

Pour conclure, observons que le membre de droite de l'inégalité ne dépend pas de ε. Si (ε_j) est une suite tendant vers 0, on peut construire par récurrence une suite croissante d'ensembles ε_j séparés maximaux E_j (et satisfaisant en outre $\mu(\partial\mathcal{P}_{E_j}) = 0$ pour tout j). Alors la suite de partitions $\mathcal{P}_{E_j}$ est croissante et le diamètre des pièces de $\mathcal{P}_{E_j}$ est majoré par $2\varepsilon_j$. C'est donc une suite génératrice (voir le §4.5) et quand $j \to \infty$ $\frac{1}{n}h_\mu(f^n, \mathcal{P}_{E_j})$ tend vers $\frac{1}{n}h_\mu(f^n) = h_\mu(f)$, et l'on obtient ainsi

$$h_\mu(f) \leqslant \frac{1}{n}\log(2^d C(d)) + \frac{1}{n}\int_M \sum_{i=1}^d \log^+\left(\sigma_i(Df_x^n)\right) \mathrm{d}\mu(x).$$

Il ne reste qu'à faire tendre n vers l'infini et utiliser la variante du théorème d'Oseledets exposée au §6.7 et l'inégalité de Ruelle est démontrée. □

6.9. *Lemme. — Si (X, μ, f) est un système dynamique mesurable et $\mathcal{P}$ est une partition mesurable finie, on a $h_\mu(f, \mathcal{P}) \leqslant H(f^{-1}\mathcal{P}|\mathcal{P})$.*

Démonstration. — On rappelle que $h_\mu(f, \mathcal{P}) = \lim_{n\to\infty} \frac{1}{n}H(\mathcal{P} \vee f^{-1}\mathcal{P} \vee \cdots f^{-(n-1)}\mathcal{P})$. Écrivons

$$H(\mathcal{P} \vee f^{-1}\mathcal{P} \vee \cdots f^{-(n-1)}\mathcal{P}) = H(\mathcal{P}) + H(f^{-1}\mathcal{P} \vee \cdots f^{-(n-1)}\mathcal{P}|\mathcal{P})$$

et utilisons la formule $H(\mathcal{P} \vee \mathcal{Q}|\mathcal{R}) = H(\mathcal{P}|\mathcal{Q} \vee \mathcal{R}) + H(\mathcal{Q}|\mathcal{R})$:

$$\begin{aligned} H(f^{-1}\mathcal{P} \vee \cdots f^{-(n-1)}\mathcal{P}|\mathcal{P}) &= H(f^{-1}\mathcal{P}|\mathcal{P}) + H(f^{-2}\mathcal{P} \vee \cdots f^{-(n-1)}\mathcal{P}|\mathcal{P} \vee f^{-1}\mathcal{P}) \\ &\leqslant H(f^{-1}\mathcal{P}|\mathcal{P}) + H(f^{-2}\mathcal{P} \vee \cdots f^{-(n-1)}\mathcal{P}|f^{-1}\mathcal{P}) \\ &= H(f^{-1}\mathcal{P}|\mathcal{P}) + H(f^{-1}\mathcal{P} \vee \cdots f^{-(n-2)}\mathcal{P}|\mathcal{P}). \end{aligned}$$

En itérant cette inégalité on aboutit à

$$H(f^{-1}\mathcal{P} \vee \cdots f^{-(n-1)}\mathcal{P}|\mathcal{P}) \leqslant (n-1)H(f^{-1}\mathcal{P}|\mathcal{P})$$

et le résultat suit. □

7. Miscellanées

7.1. — Donnons nous un homéomorphisme f d'un espace métrique compact X. À toute mesure invariante μ est associée son entropie, on a donc une application

$$\mathcal{M} \ni \mu \mapsto h_\mu(f) \in [0, \infty]$$

définie sur l'ensemble des mesures invariantes. On peut se demander: quelle est la régularité de cette application? l'ensemble des valeurs qu'elle prend? En particulier, si $h_\mu(f)$ mesure la complexité de (f, μ) il est naturel de penser la complexité "totale" de f comme étant $\sup_{\mu\in\mathcal{M}} h_\mu(f)$.

Il se trouve que cette quantité peut être caractérisée dynamiquement: c'est l'**entropie topologique** de f, qui peut être définie directement de la manière suivante. Introduisons une suite de distances (d_n) sur X par $d_n(x, y) = \max_{0\leqslant k\leqslant n-1} d(f^k(x), f^k(y))$. On dit que deux points x et y sont (n, ε) séparés si $d_n(x, y) \geqslant \varepsilon$. Pour tout $\varepsilon > 0$ on pose $N(f, \varepsilon, n)$ le cardinal maximal d'un ensemble (n, ε) séparé: par compacité celui-ci est fini. Intuitivement, c'est le nombre maximal de suites de la forme $(x, f(x), \ldots, f^{n-1}(x))$ que l'on peut distinguer dans X si on l'observe avec une résolution ε. Ce nombre croît en général exponentiellement avec n. L'entropie topologique est finalement définie comme la limite de ce taux de croissance exponentiel quand $\varepsilon \to 0$:

$$h_{\mathrm{top}}(f) = \lim_{\varepsilon \to 0} \limsup_{n\to\infty} \frac{1}{n} \log N(f, \varepsilon, n) \in [0, \infty].$$

C'est un invariant de conjugaison topologique, et le **principe variationnel** affirme que $\sup_{\mu \in \mathcal{M}} h_\mu(f) = h_{\mathrm{top}}(f)$. Dans le cas du décalage sur d symboles, d'après le Théorème 5.7 on a $h_{\mathrm{top}}(f) = \log d$. On pourra consulter l'une des références [**4, 5, 6, 14**] pour plus de détails.

Exercice. — Montrer que $\mu \mapsto h_\mu(f)$ n'est pas continue en général.

Exercice. — Montrer que l'entropie topologique d'une application lipschitzienne est finie.

7.2. — L'entropie d'une partition est définie comme une information moyenne, et il est naturel de se demander comment cette information est répartie pour la partition itérée $\mathcal{P}^{(n)}$. Le théorème suivant affirme que quand μ est ergodique cette répartition est en quelque sorte uniforme.

Théorème (Shannon-McMillan-Breiman). — Soit $(X, \mathcal{A}, \mu, f)$ un système dynamique mesuré ergodique et $\mathcal{P}$ une partition mesurable finie. Alors pour μ presque tout x on a

$$\lim_{n\to\infty} \frac{1}{n} I(\mathcal{P}^{(n)}(x)) \underset{n\to\infty}{\longrightarrow} h_\mu(f).$$

Pour la démonstration on pourra consulter le chapitre 2 de [**12**] (qui de manière générale présente de façon assez compacte bon nombre de thèmes avancés sur l'entropie métrique).

◇

IX. MESURES INVARIANTES ABSOLUMENT CONTINUES SUR L'INTERVALLE

Nous avons vu au chapitre II qu'un système dynamique topologique admet toujours une mesure invariante. Néanmoins cette mesure peut être très singulière, et ne pas décrire le comportement statistique d'un ensemble "substantiel" d'orbites. Dans le cas d'un système dynamique défini sur une variété, un objectif naturel serait de chercher une description de presque toute orbite au sens de la mesure de Lebesgue. On dit qu'une mesure invariante μ sur une variété M (disons compacte) est une **mesure physique** s'il existe un ensemble de mesure de Lebesgue positive de points $x \in M$ tels que pour toute fonction continue φ

$$\frac{1}{n}\sum_{k=1}^{n} \varphi(f^k(x)) \underset{n\to\infty}{\longrightarrow} \int \varphi \, \mathrm{d}\mu.$$

L'ensemble de ces points x est par définition le bassin $\mathcal{B}(\mu)$ de μ. La terminologie est claire: les mesures physiques sont celles qui se "voient" en pratique si on fait des "observations" sur X; elles sont donc d'une grande importance pour les applications concrètes des systèmes dynamiques. Si μ est une mesure ergodique absolument continue (i.e. absolument continue par rapport à la mesure de Lebesgue), c'est une mesure physique mais il y a d'autres exemples: le plus simple est celui d'une mesure atomique portée par un point fixe attractif.

De manière plus générale, on peut se demander si un système dynamique admet une (ou peut être plusieurs) mesure(s) invariante(s) "naturelle(s)". C'est un thème important dans la théorie contemporaine des systèmes dynamiques. Au delà des mesures physiques et des mesures absolument continues, on peut chercher des mesures maximisant certaines fonctionnelles dynamiques, comme l'entropie. Ces mesures maximisantes s'appellent des **états d'équilibre** (voir par exemple [**12**, Chap. 3]).

Dans ce chapitre nous allons montrer que certaines applications dilatantes de l'intervalle admettent des mesures ergodiques absolument continues par rapport à la mesure de Lebesgue. C'est un cas particulier d'un ensemble de résultats beaucoup plus généraux, en dimension arbitraire, qui sera l'occasion de rencontrer quelques concepts importants, comme celui d'opérateur de transfert, ainsi que d'appréhender le rôle clé des estimées de distortion(1).

(1) Voir notamment K. Krzyżewsk et W. Szlenk, *On invariant measures for expanding differentiable mappings.* Studia Math. 33 (1969) pour l'équivalent de ces constructions pour les applications dilatantes des variétés compactes.

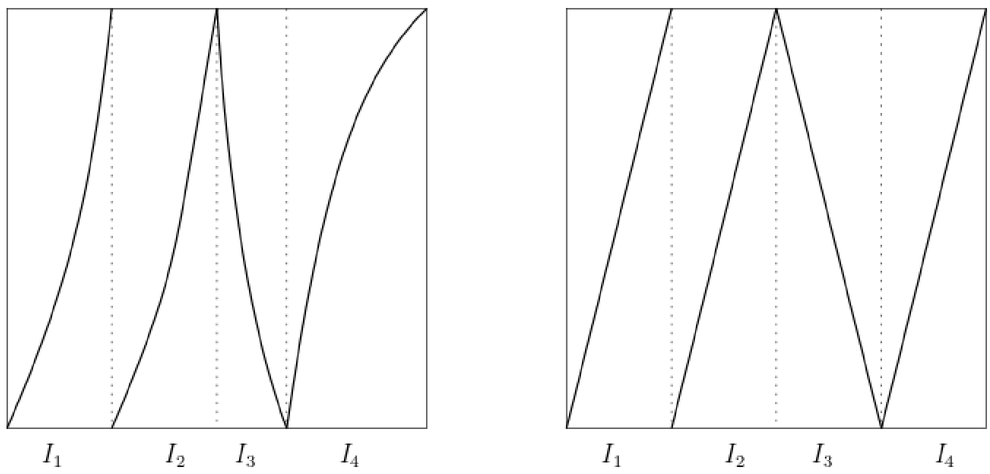


FIGURE 1. Deux applications markoviennes dilatantes de même profil.

1. Applications markoviennes de l'intervalle

1.1. — Dans tout ce chapitre nous considérons une application $f : I \to I$, où $I \subset \mathbb{R}$ est un segment, et nous normalisons la situation en fixant $I = [0,1]$. On suppose donnée une partition $\mathcal{P} = \{I_1, \ldots, I_d\}$ de I en d sous intervalles I_j, rangés dans l'ordre croissant. Comme on travaille mod. 0 on supposera par commodité les intervalles I_j ouverts. Pour tout j on suppose que $f|_{I_j}$ s'étend en un difféomorphisme de $\overline{I_j}$ sur I (propriété de Markov "à branches complètes"). On notera $f_j = f|_{I_j}$.

Les valeurs de f aux extrémités de ces intervalles n'ont pas particulièrement d'importance (on peut par exemple décider que f est continue à gauche). En tout état de cause on ne demande pas que f soit continue sur I. Enfin nous supposerons qu'il existe $\beta > 1$ tel que pour tout $j \in \{1, \ldots, d\}$, $|f_j'| = |(f|_{I_j})'| \geqslant \beta$ (propriété de dilatation uniforme).

Une transformation f satisfaisant ces hypothèses est donc par définition une "application markovienne à branches complètes uniformément dilatante". Dans la suite nous abrègerons cette terminologie en **application markovienne dilatante**.

1.2. — Il est à noter que la plupart des résultats ci-dessous (notamment la construction et les propriétés de la mesure physique) s'étendent au cas d'applications markoviennes plus générales, où l'on demande simplement que $f(I_j)$ soit une réunion de I_k, et que la chaine de Markov associée soit primitive.

En revanche, la propriété de dilatation uniforme est essentielle: si par exemple on déforme f de sorte que 0 est un point fixe tel que $f'(0) = 1$ et f est dilatante hors de l'origine (applications dites "intermittentes") alors il n'y a plus de mesure physique (finie).

Un autre cadre auquel ces résultats se généralisent sans difficulté est celui des applications dilatantes du cercle (voir la section 5 ci-après), et même plus généralement de toute variété compacte (voir [**6**]).

1.3. — Il est facile de voir en arguant comme au § 3.5 du chapitre I que f est essentiellement conjuguée à un décalage sur d symboles. Pour montrer les propriétés de rigidité de la section 4, nous aurons besoin d'un résultat plus précis (voir le Théorème 1.4 ci-dessous). Soit E_0 l'ensemble des extrémités des intervalles I_j et $E = E(f)$ l'ensemble dénombrable des préimages de E_0 par un itéré de f: $E_n = \bigcup_{k\leqslant n} f^{-k}(\bigcup \partial I_j)$ et $E = \bigcup_n E_n$. Pour une suite $(j_0, \dots, j_n) \in \{1, \dots d\}^{n+1}$ on pose

$$I_{j_0,\dots,j_n} = f_{j_0}^{-1} f_{j_1}^{-1} \cdots f_{j_{n-1}}^{-1} I_{j_n},$$

c'est-à-dire l'intervalle des points dont l'itinéraire dans la partition $\mathcal{P}$ commence par $(j_0, \cdots, j_n)$. Comme au chapitre précédent on note $\mathcal{P}^{(n)}$ la partition $\bigvee_{i=0}^{k-1} f^{-i}\mathcal{P}$, qui est exactement la partition en les intervalles $I_{j_0,\dots,j_n}$.

Remarquer que la longueur de $(I_{j_0,\dots j_n})$ est majorée par $\beta^{-(n+1)}$. Ainsi l'application d'itinéraire $\iota : I\backslash E \to \{1, \dots d\}^{\mathbb{N}}$ est injective, et on voit facilement que son image est le complémentaire d'un ensemble dénombrable. Il découle également que la partition $\mathcal{P}$ est génératrice (au sens de l'entropie, cf. la définition 4.7 du chapitre précédent).

1.4. — Pour aller plus loin, définissons le **profil** d'une application markovienne comme étant la suite des signes des dérivées des f_j. Par exemple le profil des exemples de la figure 1 est $(+, +, -, +)$.

Théorème. — Deux applications markoviennes dilatantes de même profil sont topologiquement conjuguées. Plus précisément, si f et g sont des applications markoviennes de classe C^1 par morceaux et de même profil, il existe un homéomorphisme croissant ϕ de I tel que $\phi \circ f = g \circ \phi$ hors de $E(f)$. En outre ϕ est Hölder.

Démonstration. — Si f est une application markovienne dilatante, considérons la partition $\mathcal{P}^{(n)}$ et rangeons ses intervalles en ordre croissant, on obtient ainsi une suite ordonnée d'intervalles $(I_0^{(n)}, I_1^{(n)}, \dots, I_{d^{n+1}}^{(n)})$. Remarquer que l'ensemble des bords de ces intervalles est $E_n(f)$. Si le profil est $(+, \cdots, +)$, l'ordre ainsi obtenu sur les $I_{j_0,\dots j_n}$ est l'ordre lexicographique, mais pour un profil différent on obtient un autre ordre, que l'on pourrait expliciter. Il est clair que cet ordre ne dépend que du profil de f. Si maintenant on considère f et g comme dans l'énoncé, pour tout $n \geqslant 0$ posons ϕ_n l'unique application croissante $I \to I$ affine par morceaux qui pour tout $1 \leqslant q \leqslant d^{n+1}$ envoie $I_q^{(n)}(f)$ sur $I_q^{(n)}(g)$ (on pourrait la définir juste par son action sur l'ensemble des extrémités $E_n(f)$). Lors du passage de n à $n+1$, chaque intervalle $I_q^{(n)}(f)$ est subdivisé en d nouveaux intervalles, et l'application ϕ_n, qui est linéaire de $I_q^{(n)}(f)$ dans $I_q^{(n)}(g)$ est "découpée" en une application affine par morceaux (voir la figure 2). Noter que ϕ_n et ϕ_{n+1} coincident sur $\partial I_q^{(n)}(f)$. Plus généralement pour $m \geqslant n$, $\phi_m|_{E_n(f)} = \phi_n|_{E_n(f)}$, et donc pour tout $e \in E(f)$, la suite $(\phi_n(e))$ est stationnaire, donc convergente.

Pour conclure la démonstration il suffit de montrer que la suite (ϕ_n) converge uniformément vers une application strictement croissante $\phi : I \to I$. En effet ϕ sera automatiquement une conjugaison pour la raison suivante: pour tout $1 \leqslant q \leqslant d^{n+1}$, $I_q^{(n)}(f)$ est un intervalle de profil $I_{j_{0,q},\dots,j_{n,q}}(f)$ et comme la combinatoire des intervalles de $\mathcal{P}^{(n)}$ ne dépend que du profil, on a $I_q^{(n)}(g) = I_{j_{0,q},\dots,j_{n,q}}(g)$ (i.e. pour les mêmes indices). Ensuite $f(I_q^{(n)}(f)) =$

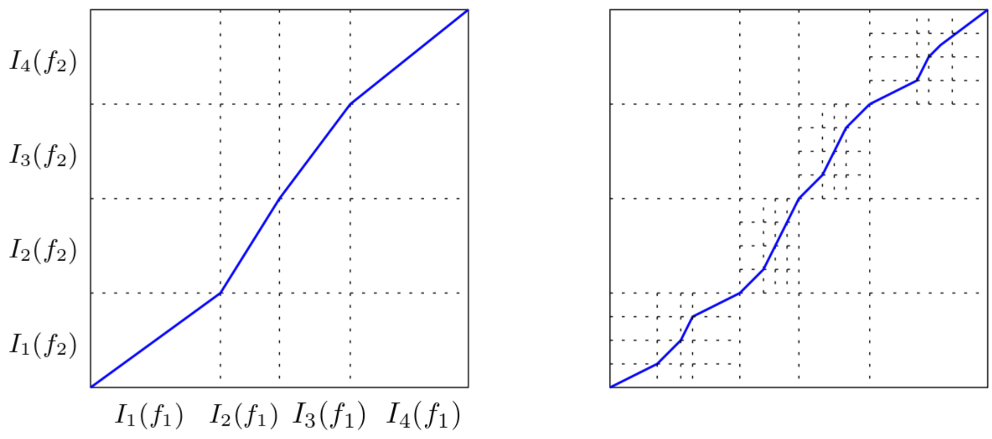


Figure 2. Deux étapes de la construction de ϕ.

$I_{j_{1,q},\dots,j_{n,q}}(f)$ est un intervalle de la forme $I_{q'}^{(n-1)}(f)$ et donc pour les mêmes raisons on a et $g(I_q^{(n)}(g)) = I_{j_{1,q},\dots,j_{n,q}}(g) = I_{q'}^{(n-1)}(g)$, donc $\phi_n(f(I_q^{(n)}(f))) = g(I_q^{(n)}(g)) = g(\phi_n(I_q^{(n)}(f)))$, et pour tout $m \geqslant n$ on a également $\phi_m(f(I_q^{(n)}(f))) = g(\phi_m(I_q^{(n)}(f)))$. Comme (ϕ_m) converge vers l'homéomorphisme ϕ, on conclut que $\phi(f(I_q^{(n)}(f))) = g(\phi(I_q^{(n)}(f)))$, et, finalement, comme tout $x \notin E(f)$ est l'intersection décroissante d'une suite de $I_{q_n}^{(n)}$ on conclut que $\phi(f(x)) = g(\phi(x)$.

Pour montrer que (ϕ_n) converge uniformément, nous allons montrer qu'elle est localement uniformément Hölder. Ainsi par le théorème d'Ascoli elle admet une sous suite uniformément convergente, mais la limite ϕ est unique car $\phi_n|_{E(f)}$ converge simplement et $E(f)$ est dense. En outre, ϕ sera strictement croissante car $\phi|_{E(f)}$ est strictement croissante, et également Hölder. Pour $k = 1, 2$, fixons $1 < \beta_k \leqslant \gamma_k$ tels que $\beta_k \leqslant |(f_k)'| \leqslant \gamma_k$ sur I. Pour $x, y \in I$ soit $n = n(x, y)$ l'unique entier tel que $\gamma_1^{-(n+2)} \leqslant |x - y| < \gamma_1^{-(n+1)}$. Comme les pièces de $\mathcal{P}^{(n)}(f)$ sont des intervalles de longueur $\geqslant \gamma_1^{-(n+1)}$, x et y sont soit dans la même pièce, soit dans des pièces voisines. Ainsi $\phi_n(x)$ et $\phi_n(y)$ sont soit dans la même pièce de $\mathcal{P}^{(n)}(g)$, soit dans des pièces voisines, de même que $\phi_m(x)$ et $\phi_m(y)$ pour tout $m \geqslant n$. On en déduit que

$$|\phi_m(x) - \phi_m(y)| \leqslant 2\beta_2^{-(n+1)} \leqslant C\,|x - y|^{\log \beta_2/\log \gamma_1}\,. \tag{1}$$

Ce n'est pas encore tout à fait l'hypothèse du théorème d'Ascoli car n dépend de x et y. Pour s'en sortir, fixons $\varepsilon > 0$ et δ tel que $C\delta^{\log \beta_2/\log \gamma_1} < \varepsilon$, où C est comme dans (1). Le raisonnement précédent dit qu'il existe $n = n(\delta)$ (et donc $n = n(\varepsilon)$) tel que tel que si $|x - y| \approx \delta$ et $m \geqslant n$, alors $|\phi_m(x) - \phi_m(y)| < \varepsilon$. Mais comme ϕ_m est croissante, sa variation sur un intervalle de taille $< \delta$ est controlée par sa variation sur un intervalle de taille δ, et on en déduit que si $|x - y| < \delta$, on a également $|\phi_m(x) - \phi_m(y)| < \varepsilon$. Finalement quitte à diminuer un peu δ, on a également la majoration $|\phi_k(x) - \phi_k(y)| < \varepsilon$ pour les indices $k < n$, et l'équicontinuité est démontrée. Pour le caractère Hölder de la limite, il suffit de faire tendre m vers l'infini dans l'estimée (1). $\square$

Remarque. — Pour montrer la convergence uniforme et l'existence de ϕ, on aurait pu utiliser le théorème de Dini plutôt que le théorème d'Ascoli, mais la continuité Hölder de ϕ est une information intéressante, et illustre un phénomène typique en dynamique hyperbolique.

2. Construction de la mesure physique

Dans cette section nous gardons les notations de la section précédente, et travaillons sous l'hypothèse que les branches f_j de f sont de classe $C^{1+\alpha}$ pour un $\alpha > 0$. On rappelle que ceci signifie que f_j est de classe C^1 et f_j' est hölderienne d'exposant α. Notre objectif est de démontrer le théorème suivant.

2.1. *Théorème. — Si f est une application markovienne dilatante à branches de classe $C^{1+\alpha}$, elle admet une mesure invariante μ absolument continue sur I, dont la densité est continue et strictement positive. En outre μ est mélangeante.*

Le théorème ergodique montre alors:

Corollaire. — La mesure μ est l'unique mesure physique de f (en particulier l'unique mesure invariante absolument continue), et son bassin est de mesure totale.

En, particulier, pour Leb presque tout $x \in I$, la suite $f^n(x)$ s'équidistribue selon μ.

2.2. — Notons Leb la mesure de Lebesgue sur I.

Lemme. — Si $h \in L^1(I)$ on a $f_\star(h\,\mathrm{Leb}) = (\mathcal{L}h)\mathrm{Leb}$, où

$$\mathcal{L}h(x) = \sum_{j=1}^{d} \frac{h(f_j^{-1}(x))}{|f_j'(f_j^{-1}(x))|} = \sum_{y \in f^{-1}(x)} \frac{h(y)}{|f'(y)|}. \tag{2}$$

L'opérateur $\mathcal{L}$ s'appelle l'**opérateur de transfert** (noter que dans la dernière expression de l'égalité (2) on ne tient pas compte d'éventuels y situés au bord des I_j). Remarquer que la conservation de la masse par $f_\star$ implique que $\|\mathcal{L}h\|_{L^1(I,\mathrm{Leb})} = \|h\|_{L^1(I,\mathrm{Leb})}$.

Démonstration. — C'est juste la formule du changement de variables:

$$\begin{aligned}\int \varphi\, f_\star(h\,\mathrm{Leb}) &= \int_I \varphi \circ f(x)h(x)dx = \sum_{j=1}^{d} \int_{I_j} \varphi \circ f(x)h(x)dx \\ &= \sum_{j=1}^{d} \int_I \varphi(y)\frac{h(f_j^{-1}(y))}{|f_j'(f_j^{-1}(y))|}dy = \int_I \varphi(x)\mathcal{L}h(x)\,dx.\end{aligned}$$

□

2.3. — On voit donc que tout point fixe h de l'opérateur de transfert définit une mesure f-invariante absolument continue $h\,\mathrm{Leb}$. L'assertion d'existence dans le Théorème 2.1 découle donc de la proposition suivante:

Proposition. — Il existe une fonction continue h sur I telle que $\mathcal{L}h = h$. En outre h est strictement positive et hölderienne d'exposant α.

2.4. — Comme souvent en dynamique différentiable sur l'intervalle, le point technique clé est de démontrer une **estimée de distortion** (cf. la section 5 du chapitre IV). Nous aurons besoin de la version précise suivante:

Lemme. — Il existe une constante D telle que pour tous $x, y \in I$ et toute suite $(j_0, \dots, j_{n-1})$, on a

$$\left|\frac{(f^n)'(f_{j_0}^{-1} \circ \cdots \circ f_{j_{n-1}}^{-1}(x))}{(f^n)'(f_{j_0}^{-1} \circ \cdots \circ f_{j_{n-1}}^{-1}(x'))}\right| \leqslant 1 + D\left|x - x'\right|^{\alpha}.$$

En particulier,

$$\left|\frac{(f^n)'(y)}{(f^n)'(y')}\right| \leqslant D' = 1 + D$$

sur toute pièce de $\mathcal{P}^{(n-1)}$.

Démonstration. — Soient x et x' comme dans l'énoncé, posons $y = f_{j_0}^{-1} \circ \cdots \circ f_{j_{n-1}}^{-1}(x)$, (resp. $y' = f_{j_0}^{-1} \circ \cdots \circ f_{j_{n-1}}^{-1}(x')$) et pour $0 \leqslant k \leqslant n-1$, $y_k = f^k(y)$ (resp. $y'_k = f^k(y')$). On a alors

$$\frac{(f^n)'(y)}{(f^n)'(y')} = \prod_{k=0}^{n-1} \frac{f'(y_k)}{f'(y'_k)}. \tag{3}$$

Mais $y_k = f_{j_k}^{-1} \circ \cdots \circ f_{j_{n-1}}^{-1}(x)$ et de même pour y'_k. On en déduit que $|y_k - y'_k| \leqslant \beta^{-(n-k)}|x - x'|$ et donc comme f est $C^{1+\alpha}$,

$$|f'(y_k) - f'(y'_k)| \leqslant C_1 \beta^{-\alpha(n-k)} |x - x'|^{\alpha},$$

et finalement

$$\left|\frac{f'(y_k)}{f'(y'_k)}\right| \leqslant 1 + C_2 \beta^{-\alpha(n-k)} |x - x'|^{\alpha}.$$

En entrant cette estimation dans (3) on obtient

$$\begin{aligned}
\left|\frac{(f^n)'(y)}{(f^n)'(y')}\right| &\leqslant \prod_{k=0}^{n-1} \left(1 + C_2 \beta^{-\alpha(n-k)} |x - x'|^{\alpha}\right) \\
&= \exp\left(\sum_{k=0}^{n-1} \log\left(1 + C_2 \beta^{-\alpha(n-k)} |x - x'|^{\alpha}\right)\right) \\
&\leqslant \exp\left(\sum_{k=0}^{n-1} C_2 \beta^{-\alpha(n-k)} |x - x'|^{\alpha}\right) \quad \text{car } \log(1+u) \leqslant u \text{ sur } \mathbb{R}_+ \\
&\leqslant \exp\left(C_2 (1 - \beta^{-\alpha})^{-1} |x - x'|^{\alpha}\right) \\
&\leqslant 1 + D|x - x'|^{\alpha} \quad \text{car } e^u \leqslant 1 + \frac{e^M - 1}{M} u \text{ sur } [0, M],
\end{aligned}$$

comme annoncé. □

Démonstration de la Proposition 2.3. — Nous allons dans un premier temps montrer que la suite de fonctions $\mathcal{L}^n\mathbf{1}$ est équicontinue et vérifie $1/c \leqslant \mathcal{L}^n\mathbf{1} \leqslant c$ pour une certaine constante $c > 0$. Par une récurrence immédiate, on voit que

$$\mathcal{L}^n\mathbf{1}(x) = \sum_{y\in f^{-n}(x)} \frac{1}{|(f^n)'(y)|} = \sum_{j_0,\dots,j_{n-1}=1}^{d} \frac{1}{\left|(f^n)'(f_{j_0}^{-1}\circ\cdots\circ f_{j_{n-1}}^{-1}(x))\right|}.$$

Pour tous $x, x' \in I$, d'après le lemme de distortion on a pour tous $j_0, \dots, j_{n-1}$,

$$\frac{(D+1)^{-1}}{\left|(f^n)'(f_{j_0}^{-1}\circ\cdots\circ f_{j_{n-1}}^{-1}(x))\right|} \leqslant \frac{1}{\left|(f^n)'(f_{j_0}^{-1}\circ\cdots\circ f_{j_{n-1}}^{-1}(x'))\right|} \leqslant \frac{D+1}{\left|(f^n)'(f_{j_0}^{-1}\circ\cdots\circ f_{j_{n-1}}^{-1}(x))\right|},$$

et donc

$$\frac{1}{D+1}\mathcal{L}^n(x) \leqslant \mathcal{L}^n(x') \leqslant (D+1)\mathcal{L}^n(x).$$

Mais comme $\mathcal{L}^n\mathbf{1}$ est la densité de la mesure $f^n_*\mathrm{Leb}$, on a également $\int_I \mathcal{L}^n\mathbf{1}(x)\,dx = 1$, et on en déduit que pour tout $x \in I$, $(D+1)^{-1} \leqslant \mathcal{L}^n(x) \leqslant D+1$.

Pour l'équicontinuité le raisonnement est analogue: on répète cette estimation en utilisant cette fois la première estimation du Lemme 2.4 pour aboutir à

$$\frac{1}{1+D\,|x-x'|^\alpha} \leqslant \frac{\mathcal{L}^n(x')}{\mathcal{L}^n(x)} \leqslant 1 + D\,|x-x'|^\alpha\,,$$

et finalement

$$|\mathcal{L}^n(x) - \mathcal{L}^n(x')| = \mathcal{L}^n(x)\left|1 - \frac{\mathcal{L}^n(x')}{\mathcal{L}^n(x)}\right| \leqslant D\cdot D\,|x-x'|^\alpha\,,$$

ce qui établit l'équicontinuité.

Finalement, pour trouver la densité invariante, nous appliquons l'argument habituel de moyennisation: soit $h_n = \frac{1}{n}\sum_{k=1}^n \mathcal{L}^k\mathbf{1}$. C'est une suite de fonctions uniformément bornées et uniformément hölderiennes d'exposant α. Par le théorème d'Ascoli elle admet une valeur d'adhérence h pour la norme uniforme, de même régularité, et clairement, $\mathcal{L}h = h$, ce qui conclut la preuve. □

2.5. — Pour établir l'ergodicité nous aurons besoin d'une version (simplifiée) du théorème de densité de Lebesgue. Nous avons vu qu'à une application markovienne dilatante f est associée une suite de partitions $\mathcal{P}^{(n)}$ en intervalles (ouverts) de diamètre tendant vers 0. Soit $\mathcal{F}_n$ la tribu engendrée par $\mathcal{P}^{(n)}$.tribu engendrée par les $\mathcal{F}_n$ coincide mod. 0 avec la tribu borélienne (plus précisément, elle contient les ouverts de $I\backslash E(f)$).

Théorème. — *Si* $h \in L^1(I, \mathrm{Leb})$ *alors* $\mathbb{E}(h|\mathcal{F}_n)$ *tend vers* h Leb-*p.p.*

Démonstration. — Nous n'aurons besoin de ce résultat que pour une fonction caractéristique donc par simplicité supposons que $h \in L^2$. Par construction $(\mathbb{E}(h|\mathcal{F}_n))$ est une martingale (dite martingale de Doob) adaptée à la filtration $\mathcal{F}_n$, donc $\mathbb{E}(h|\mathcal{F}_n)$ converge p.s. et dans L^2 vers une fonction $k \in L^1$. Comme pour $m \geqslant n$ on a

$$\mathbb{E}(\mathbb{E}(h|\mathcal{F}_m)|\mathcal{F}_n) = \mathbb{E}(h|\mathcal{F}_n),$$

en faisant $m \to \infty$ et en utilisant la convergence L^2 on en déduit que $\mathbb{E}(k|\mathcal{F}_n) = \mathbb{E}(h|\mathcal{F}_n)$ pour tout n, donc $h = k$ p.p. et le résultat est démontré. □

2.6. *Proposition. — La mesure absolument continue* $\mu = h\,\mathrm{Leb}$ *construite à la Proposition 2.3 est exacte, donc mélangeante.*

Démonstration. — D'après le §II.6, il faut démontrer que si A est un ensemble mesurable tel que $\mu(A) > 0$, alors $\mu(f^n(A)) \to 1$ quand $n \to \infty$. Comme $c^{-1}\mathrm{Leb} \leqslant \mu \leqslant c\mathrm{Leb}$, il suffit en fait de d'établir ce résultat pour la mesure de Lebesgue. Fixons $\varepsilon > 0$. Avec les notations précédentes on a $\mathbb{E}(\mathbf{1}_A|\mathcal{F}_n) \to 1_A$ p.p. donc pour tout n assez grand il existe une pièce $P \in \mathcal{P}^{(n)}$ telle que $\mathrm{Leb}(A \cap P) \geqslant (1-\varepsilon)\mathrm{Leb}(P)$, ou de manière équivalente, $\mathrm{Leb}(A^\complement \cap P) \leqslant \varepsilon\mathrm{Leb}(P)$. Mais par ailleurs $f^n : P \to I$ est un difféomorphisme de distortion bornée par une constante uniforme D, donc, par la formule du changement de variable on a que pour tout $y_0 \in P$,

$$\frac{1}{D}\,|(f^n)'(y_0)|\,\mathrm{Leb}(P) \leqslant \int_P |(f^n)'(y)|\,\mathrm{d}y = \mathrm{Leb}(f^n(P)) = 1 \leqslant D\,|(f^n)'(y_0)|\,\mathrm{Leb}(P).$$

En particulier $|(f^n)'(y_0)| \leqslant \frac{D}{\mathrm{Leb}(P)}$. On en déduit que

$$\mathrm{Leb}(f^n(A^\complement \cap P)) = \int_{A^\complement \cap P} |(f^n)'(y)|\,\mathrm{d}y \leqslant D\frac{\mathrm{Leb}(A^\complement \cap P)}{\mathrm{Leb}(P)} \leqslant D\varepsilon,$$

et donc $\mathrm{Leb}(f^n(A)) \geqslant 1 - D\varepsilon$, et la proposition suit. □

2.7. *Remarque.* — La preuve de la Proposition 2.3 montre que la densité de la mesure absolument continue est α-hölderienne. Plus généralement si les branches de f sont de classe C^r on peut montrer que la densité est C^{r-1}. Pour les applications dilatantes du cercle c'est un théorème de Krzyzewsky[(2)] et Sacksteder[(3)] On peut le démontrer directement par récurrence sur r en étudiant les dérivées successives de $\mathcal{L}h$. L'exercice suivant propose une preuve directe du fait que la densité est analytique lorsque les branches de f sont analytiques.

Exercice. — Soit f une application markovienne dilatante telle que les branches f_j sont réelles-analytiques. On suppose que pour tout $j \in \{1, \ldots, d\}$, f_j admet une extension analytique à un voisinage de $\overline{I_j}$ dans $\mathbb{R}$.

1. Montrer qu'il existe un voisinage V de I dans $\mathbb{C}$ et pour tout j un voisinage U_j de $\overline{I_j}$ dans $\mathbb{C}$, tel que $U_j \Subset V$, et une extension holomorphe de f_j à U_j telle que $f_j : U_j \to V$ est un difféomorphisme dilatant.
2. Soit $(\sigma_1, \ldots, \sigma_d)$ le profil de f. Montrer que l'opérateur de transfert défini par
$$\mathcal{L}h(x) = \sum_{j=1}^{d} \frac{h(f_j^{-1}(x))}{\sigma_j f_j'(f_j^{-1}(x))}$$
s'étend à l'espace $H^\infty(V)$ des fonctions holomorphes bornées sur V, et en déduire que la densité de la mesure invariante est analytique.

(2) *Some results on expanding mappings.* Asterisque 50 (1977)
(3) *The measures invariant under an expanding map.* Lectures Notes in Math. 392. Springer.

3. Rigidité différentiable

Dans cette section nous donnons une application un peu inattendue de l'existence (ou plutôt de la régularité) de la mesure absolument continue à une question de rigidité.

3.1. — Rappelons qu'une application $g : I \to \mathbb{R}$ est **absolument continue** si pour tout $\varepsilon > 0$ il existe $\delta > 0$ tel que pour toute collection d'intervalles disjoints $([x_j, y_j])_{j \geqslant 0}$ telle que $\sum_j |y_j - x_j| \leqslant \delta$ on a $\sum_j |g(x_j) - g(y_j)| \leqslant \varepsilon$. Par exemple une application lipschitzienne est absolument continue. La propriété clé (qui découle directement de la définition) des applications absolument continues est qu'elles préservent les ensembles de mesure de Lebesgue nulle.

3.2. — Nous avons vu au Théorème 1.4 que deux applications markoviennes dilatantes de même profil sont Hölder conjuguées. Le théorème suivant montre que si la conjugaison est absolument continue (par exemple Lipschitz), elle est alors automatiquement beaucoup plus régulière. C'est ce qu'on appelle un résultat de **rigidité différentiable**.

Théorème. — Soient f et g deux applications markoviennes dilatantes à branches $C^{1+\alpha}$ (resp. C^k, $k \geqslant 2$, resp. analytiques) sur I, de même profil. Si f et g sont conjuguées par un homéomorphisme ϕ absolument continu, alors ϕ est $C^{1+\alpha}$ (resp. C^k, resp. analytique).

Ce théorème est une version simplifiée d'un résultat de Shub et Sullivan(4). Il faut comprendre que l'existence d'une conjugaison régulière (disons C^1) entre deux applications $C^{1+\alpha}$ est une propriété très spéciale. Par exemple une telle conjugaison préserve les multiplicateurs (dérivées) des points périodiques. Donc une petite perturbation "générique" d'une application markovienne dilatante ne lui est pas en général C^1 conjuguée (on peut facilement modifier la dérivée en un point fixe par exemple) alors que par le Théorème 1.4 elle lui est Hölder-conjuguée.

3.3. *Lemme. — Si f admet une mesure de probabilité invariante $h\,\mathrm{Leb}$ avec h et hölderienne d'exposant α (resp. resp. C^{k-1}, resp. analytique) et strictement positive, il existe un difféomorphisme $\psi : I \to I$ de classe $C^{1+\alpha}$ (resp. C^k, resp. analytique) tel que $\psi f \psi^{-1}$ préserve la mesure de Lebesgue.*

Démonstration. — Si $\psi : I \to I$ est un C^1-difféomorphisme, l'application $\psi f \psi^{-1}$ préserve $\psi_*(h\,\mathrm{Leb})$. En appliquant la formule du changement de variable comme au Lemme 2.2 on voit que

$$\psi_*(h\,\mathrm{Leb}) = \frac{h \circ \psi^{-1}}{|\psi' \circ \psi^{-1}|}\mathrm{Leb}.$$

Posons $\psi : x \mapsto \int_0^x h(t)dt$. Comme $h > 0$ et $\int_0^1 h(t)dt = 1$, ceci définit un difféomorphisme croissant de classe $C^{1+\alpha}$ (resp. C^k, resp. analytique) de I qui répond au problème. $\square$

(4) *Expanding endomorphisms of the circle revisited.* Ergodic Theory Dynam. Systems 5 (1985).

Démonstration du Théorème 3.2. — Supposons f et g à branches de classe $C^{1+\alpha}$, telles que $g = \phi f \phi^{-1}$, avec ϕ absolument continue. D'après le Théorème 2.1 et la Proposition 2.3, pour $k = 1, 2$, f_k admet une mesure invariante $h_k\,\mathrm{Leb}$ à densité $h_k > 0$ höldérienne d'exposant α. Soit ψ_k comme au lemme précédent et $g_k = \psi_k f_k \psi_k^{-1}$ qui est une application markovienne dilatante à branches $C^{1+\alpha}$. De $g = \phi f \phi^{-1}$ on déduit la relation

$$(\psi_2\phi\psi_1^{-1})\, g_1\, (\psi_2\phi\psi_1^{-1})^{-1} = g_2,$$

qui définit un homéomorphisme $\Psi := \psi_2\phi\psi_1^{-1}$ conjuguant g_1 et g_2, qui est absolument continu par composition. Si A est un ensemble de mesure de Lebesgue nulle on a $(\Psi^{-1})_\star\mathrm{Leb}(A) = \mathrm{Leb}(\Psi(A)) = 0$. Ainsi $(\Psi^{-1})_\star\mathrm{Leb}$ est une mesure absolument continue, et qui est g_1-invariante. D'après le corollaire du Théorème 2.1, celle ci est unique et on a donc $(\Psi^{-1})_\star\mathrm{Leb} = \mathrm{Leb}$. Or un homéomorphisme Ψ de I préservant la mesure de Lebesgue est linéaire: en effet si Ψ est croissant on a $\Psi^{-1}([0,x]) = [0, \Psi^{-1}(x)]$ et donc si $\mathrm{Leb}(\Psi^{-1}([0,x])) = \mathrm{Leb}([0,x])$ on déduit $\Psi^{-1}(x) = x$ donc $\Psi(x) = x$; de même si Ψ est décroissant on obtient $\Psi(x) = 1 - x$. On conclut que $\phi = \psi_2^{-1}\Psi\psi_1$ est de classe $C^{1+\alpha}$.

Dans le cas où f et g sont de classe C^k ou analytiques, on raisonne de la même façon en utilisant la Remarque 2.7. □

4. Entropie et encore de la rigidité

4.1. *Théorème (Formule de Rokhlin). — Soit f une application markovienne dilatante à branches $C^{1+\alpha}$ et μ sa mesure absolument continue. Alors*

$$h_\mu(f) = \int \log|f'|\,\mathrm{d}\mu.$$

Ceci montre en particulier que pour la mesure absolument continue on a égalité dans l'inégalité de Ruelle. Un théorème de Ledrappier affirme que c'est en fait une caractérisation de μ(5).

Démonstration. — Comme la partition $\mathcal{P}$ est génératrice (au sens de l'entropie, cf. le chapitre précédent), d'après le théorème de Kolmogorov-Sinai, on a

$$h_\mu(f) = h_\mu(f, \mathcal{P}) = \lim_{n\to\infty} -\frac{1}{n} \sum_{P\in\mathcal{P}^{(n)}} \mu(P)\log\mu(P). \tag{4}$$

Comme μ est absolument continue et à densité bornée, on a $c^{-1}\mathrm{Leb}(P) \leqslant \mu(P) \leqslant c\mathrm{Leb}(P)$ pour un c uniforme, donc en utilisant $\sum_{P\in\mathcal{P}^{(n)}} \mu(P) = 1$ on obtient

$$-\frac{1}{n} \sum_{P\in\mathcal{P}^{(n)}} \mu(P)\log\mu(P) = \frac{1}{n} \sum_{P\in\mathcal{P}^{(n)}} \mu(P)\log(\mathrm{Leb}(P)) + O\left(\frac{1}{n}\right).$$

(5) *Some properties of absolutely continuous measures on an interval.* Ergodic Theory Dynam. Syst. 1 (1981) 77–93.

En outre par le lemme de distortion (et un raisonnement analogue à celui de la Proposition 2.6), on voit qu'à une constante multiplicative uniforme près on a pour tout $x \in P$, $\mathrm{Leb}(P) \asymp |(f^n)'(x)|^{-1}$, ainsi

$$\left|\log(\mathrm{Leb}(P)) - \log \frac{1}{|(f^n)'(x)|}\right| \leqslant O(1),$$

donc

$$\mu(P)\log(\mathrm{Leb}(P)) = \int_P \log \frac{1}{(f^n)'(x)} \mathrm{d}\mu(x) + O(1),$$

et finalement

$$-\frac{1}{n}\sum_{P\in\mathcal{P}^{(n)}} \mu(P)\log\mu(P) = \frac{1}{n}\int \log|(f^n)'(x)|\,\mathrm{d}\mu(x) + O\left(\frac{1}{n}\right). \tag{5}$$

D'après le théorème de Birkhoff (voir le §5 du chapitre IV) $\frac{1}{n}\log|(f^n)'(x)|$ converge p.p. et dans $L^1(\mu)$ vers l'exposant de Lyapunov moyen $\int \log|f'|\,\mathrm{d}\mu$, et en vertu des égalités (4) et (5), on obtient $h_\mu(f) = \int \log|f'|\,\mathrm{d}\mu$, comme annoncé. □

4.2. — Une application markovienne dilatante est mesurablement conjuguée à un décalage sur d symboles, elle admet donc une autre mesure invariante naturelle qui est sa mesure équilibrée μ_{eq}, qui donne la même masse $\frac{1}{d^n}$ à tous les cylindres de profondeur n, ou de manière équivalente à tous les atomes de $\mathcal{P}^{(n-1)}$. La mesure équilibrée admet aussi la caractérisation suivante (Théorème VIII.5.7): c'est l'unique mesure d'entropie maximale $\log d$.

4.3. — Il est naturel de se demander quand les mesures μ_{eq} et μ coincident. C'est le cas lorsque f est affine par morceaux et de pente constante, c'est à dire que $|f'| = d$ (ou de manière équivalente les I_j sont de longueur $1/d$). En effet il est clair que dans ce cas la mesure de Lebesgue est à la fois f-invariante et équilibrée.

Remarquer que par le Théorème 1.4, toute application markovienne dilatante de degré d est topologiquement conjuguée à l'unique application affine par morceaux f_0 de pente constante et de même profil que f. Si cette conjugaison est absolument continue (et donc $C^{1+\alpha}$ en vertu du Théorème 3.2), on a alors $\mu = \mu_{\mathrm{eq}}$: en effet si la conjugante est notée ϕ (i.e. $\phi f_0 \phi^{-1} = f$), on a $\phi_*\mathrm{Leb} = \mu$ car ϕ préserve l'absolue continuité et $\phi_*\mathrm{Leb} = \mu_{\mathrm{eq}}$ car l'entropie est un invariant de conjugaison mesurable.

Le théorème de rigidité suivant dit que cette situation est la seule possible.

4.4. *Théorème. — Si f est une application markovienne dilatante à branches de classe $C^{1+\alpha}$ (resp. C^k, $k \geqslant 2$, resp. analytiques) sur I, telle que $\mu = \mu_{\mathrm{eq}}$, alors f est $C^{1+\alpha}$ (resp. C^k , resp. analytiquement) conjuguée à son modèle affine par morceaux de pente constante.*

Démonstration. — Supposons que les branches sont de classe $C^{1+\alpha}$; les autres cas se traitent de façon analogue grâce à la Remarque 2.7. Soit $\psi : I \to I$ le difféomorphisme croissant $C^{1+\alpha}$ défini au Lemme 3.3, tel que $g := \psi f \psi^{-1}$ est une application markovienne dilatante $C^{1+\alpha}$ préserve la mesure de Lebesgue. On a par construction $\psi_*\mu = \mathrm{Leb}$, et donc également $\psi_*\mu_{\mathrm{eq}} = \mathrm{Leb}$. D'après le Théorème 1.4 il existe un homéomorphisme croissant ϕ

conjugant f et l'unique application markovienne f_0 affine par morceaux de pente constante de même profil que f, i.e. $f_0 = \phi f \phi^{-1}$. Ainsi $\theta = \phi \circ \psi^{-1}$ est un homéomorphisme croissant conjugant f_0 et g. Comme l'entropie est un invariant de conjugaison mesurable (donc topologique), θ doit envoyer l'unique mesure d'entropie maximale de f_0 sur celle de g, ce qui se traduit en $\theta_* \mathrm{Leb} = \mathrm{Leb}$, et donc comme au Théorème 3.2 on conclut que $\theta = \mathrm{id}$. Ainsi $\phi = \psi$, $g = f_0$, et le théorème est démontré. □

5. Le cas du cercle

L'ensemble des résultats de ce chapitre s'étend au cas des applications dilatantes du cercle. On peut pour cela soit adapter facilement les démonstrations (ce qui est assez simple) ou bien directement utiliser ce qui a été fait sur l'intervalle. Dans cette courte section nous expliquons comment mettre en œuvre cette deuxième approche (quelques détails sont laissés en exercice).

5.1. — On identifie le cercle $\mathbb{S}^1$ à $\mathbb{R}/\mathbb{Z}$ et on note $\pi : \mathbb{R} \to \mathbb{S}^1$ l'application naturelle. Si $f : \mathbb{S}^1 \to \mathbb{S}^1$ est une application différentiable, on peut définir sa dérivée en tout point en relevant localement par π et remarquant que la valeur de cette dérivée ne dépend pas du choix de la préimage (en termes plus savants: le fibré tangent du cercle est trivial). On dira que f est dilatante si $|f'(x)| > 1$ pour tout x. Il y a donc deux possibilités: soit $f'(x) > 1$ pour tout x (cas croissant), soit $f'(x) < -1$ pour tout x (cas décroissant): c'est clair si f est C^1, mais c'est le cas également si f n'est supposée que différentiable (pourquoi?). Dans la suite nous nous focalisons sur le cas croissant, l'adaptation au cas décroissant étant laissé à la sagacité de la lectrice.

5.2. — Pour tout $x \in \mathbb{S}^1$ on notera $\hat{x}$ son unique relevé à $[0, 1[$. Si f est une application dilatante du cercle, posons $y_0 = \widehat{f(0)}$. On peut relever f dans un voisinage de l'origine à une application $\hat{f}$ définie au voisinage de 0 dans $\mathbb{R}$ et telle que $\hat{f}(0) = y_0$. On se convainc aisément que cette application s'étend de proche en proche à $[0, 1]$ et que $\hat{f}(1) = y_0 + d$ pour un certain $d \in \mathbb{Z}$. Si f est croissante on a $d \geqslant 1$ et comme f est dilatante on a en fait $d \geqslant 2$ (pourquoi?). On étend ensuite $\hat{f}$ par périodicité en posant pour $k \in \mathbb{Z}$ $\hat{f}(y + k) = \hat{f}(y) + dk$. Par définition d est le **degré** de f (terminologie bien sûr compatible avec le degré défini en topologie)

En comparant la croissance de $\hat{f}$ en $\pm\infty$ à celle de $x \mapsto x$, en en utilisant la dilatation on voit que $\hat{f}$ a exactement un point fixe $\hat{\alpha}$. Posons $\alpha = \pi(\hat{\alpha})$. En conjuguant f par la rotation R_α on peut supposer que $\alpha = 0$, ce que nous ferons désormais.

5.3. — Définissons maintenant un (autre) relevé g de f à $[0, 1]$ en posant $g(y) = \widehat{f(\pi(y))}$. Ceci définit une application dilatante croissante de profil $(+, \ldots, +)$, qui vaut 0 aux points de coupure donc est continue à droite, et qui admet d branches, où d est le degré de f (pourquoi?). On remarque que outre 0, g admet $d - 2$ points fixes, donc f a $d - 1$ points fixes. La projection π réalise une semiconjugaison entre g et f, qui est injective hors d'un ensemble fini, donc induit une conjugaison bimesurable. Ceci implique que les résultats de la section 2 valent pour f: *si f est $C^{1+\alpha}$, elle admet une unique mesure invariante absolument continue à densité C^{α} et strictement positive, qui est mélangeante.*

5.4. — Pour étendre les résultats de rigidité, on remarque déjà que si ϕ est un homéomorphisme croissant de $[0,1]$, il induit un unique homéomorphisme $\tilde{\phi}$ croissant du cercle tel que $\tilde{\phi}(0) = 0$. Si maintenant f et g sont des applications dilatantes croissante du cercle de même degré, quitte à les conjuguer par des rotations on peut supposer qu'elles fixent 0 et on note g_1 et g_2 leurs relevés respectifs à $[0,1]$. Par le Théorème 1.4, il existe un homéomorphisme hölderien ϕ tel $g_2 = \phi g_1 \phi^{-1}$. On vérifie alors que $g = \tilde{\phi} f \tilde{\phi}^{-1}$: en effet cette égalité vaut sur un sous-ensemble dense et tout est continu. Autrement dit: *deux applications dilatantes du cercle de même degré et de même sens de variation sont Hölder-conjuguées.* On peut maintenant étendre les résultats de rigidité des sections 3 et 4: *si $\tilde{\phi}$ est absolument continu, alors il est $C^{1+\alpha}$.* Une petite difficulté est la démonstration ne montre le caractère $C^{1+\alpha}$ de $\tilde{\phi}$ que hors du point fixe 0. Pour la résoudre, il suffit de répéter l'argument en changeant de point fixe.

De même on obtient que *si pour une application f dilatante $C^{1+\alpha}$ du cercle de degré d, la mesure absolument continue est d'entropie maximale, alors f est $C^{1+\alpha}$ conjuguée à $x \mapsto dx$.* Comme précédemment, les mêmes résultats valent dans les catégories C^k et analytique.

◇

X. SYSTÈMES DYNAMIQUES ALÉATOIRES

1. Vocabulaire et premières définitions

1.1. — Soit X un espace mesurable (qui va rapidement être supposé métrique localement compact). L'objet de ce paragraphe est de formaliser la notion de système dynamique aléatoire sur X. L'idée, naturelle, est la suivante: étudier des compositions aléatoires de transformations de X de la forme $f_n \circ \cdots \circ f_1 \circ f_0$, où les f_i sont choisies selon une certaine loi. Noter que nous avons déjà considéré une telle situation au chapitre VII, où les f_i étaient des homographies sur $\mathbb{P}^1(\mathbb{R})$.

Une motivation initiale pour ce formalisme est de modéliser l'itération d'un système dynamique soumis à un bruit aléatoire: on itère ainsi une suite aléatoire f_{ε_i} de perturbations d'une transformation f. Néanmoins, motivé notamment par des questions de mathématiques fondamentales, le cadre général s'est largement écarté de cette idée initiale et permet d'englober des situations beaucoup plus variées.

1.2. — On peut modéliser mathématiquement ceci de la manière suivante: on se donne $(\Omega, \mathcal{A})$ un espace mesurable, et à chaque $\omega \in \Omega$ est associé une transformation mesurable $f_\omega : X \to X$. On suppose en outre que $(\omega, x) \mapsto f_\omega(x)$ est mesurable. On peut alors considérer l'espace $\Omega^{\mathbb{N}} =: \Sigma_+$ des suites de variables aléatoires $\underline{\omega} = (\omega_n)_{n \geqslant 0}$, muni de sa tribu produit $\mathcal{A}^{\otimes \mathbb{N}}$ et d'une loi de probabilité $\mathbb{P}$, et il s'agit d'étudier les compositions de la forme $f_{\omega_{n-1}} \circ \cdots \circ f_{\omega_1} \circ f_{\omega_0}$.

Introduisons l'application F_+ de $\Sigma_+ \times X$ dans lui même définie par

$$F_+(\underline{\omega}, x) = (\sigma \underline{\omega}, f_{\omega_0}(x)),$$

où $\sigma : \Sigma_+ \to \Sigma_+$ est le décalage. D'après un calcul analogue à celui effectué au chapitre VI, on a

$$F_+^n(\underline{\omega}, x) = (\sigma^n(\underline{\omega}), f_{(\sigma^{n-1}\underline{\omega})_0} \circ \cdots \circ f_{\omega_0}(x)) = (\sigma^n(\underline{\omega}), f_{\omega_{n-1}} \circ \cdots \circ f_{\omega_0}(x)),$$

et on voit que l'étude de notre système dynamique aléatoire se ramène à celle du système dynamique “déterministe” $(\Sigma_+ \times X, F)$.

On voit par exemple que les cocycles linéaires ou plus généralement les G-cocycles introduits au chapitre VI sont de cette forme.

Nous ne considèrerons dans ce chapitre **que le cas où les transformations f_ω sont choisies de façon indépendante et de même loi** ν, et réserverons le terme “système dynamique aléatoire” à cette situation. Dans le cas où les f_ω sont de même loi mais non nécessairement indépendantes, on croise parfois la terminologie “système dynamique stationnaire”.

Une situation déjà très intéressante et qui exprime une bonne partie de la complexité du cas général est celle où Ω est un ensemble fini: il s'agit donc de choisir aléatoirement une transformation f dans une famille finie et d'itérer les compositions aléatoires ainsi obtenues.

1.3. *Définition.* — Un **système dynamique (mesurable) aléatoire** sur un espace mesurable $(X, \mathcal{B})$ est la donnée:

- d'un espace probabilisé $(\Omega, \mathcal{A}, \nu)$;
- et d'une application qui à $\omega \in \Omega$ associe une transformation $f_\omega : X \to X$ de sorte que $(\omega, x) \mapsto f_\omega(x)$ soit mesurable par rapport à la tribu produit $\mathcal{F}_+ = \mathcal{A}^{\otimes \mathbb{N}} \otimes \mathcal{B}$.

On lui associe alors le **produit semi-direct** F sur $\Sigma_+ \times X$ (rappelons que $\Sigma_+ = \Omega^{\mathbb{N}}$) défini par

$$F_+(\underline{\omega}, x) = (\sigma \underline{\omega}, f_{\omega_0}(x)), \text{ où } \underline{\omega} = (\omega_n)_{n \geqslant 0}.$$

Dans tout le chapitre, pour alléger les notations, nous identifierons souvent implicitement ν et son image par $\omega \mapsto f_\omega$, de sorte que ν sera parfois vue comme une mesure sur l'ensemble des applications de X dans X. Ce point de vue sera particulièrement utile lorsque nous considèrerons un groupe ou semigroupe G agissant sur X, ν pouvant être vue comme une mesure sur G. Par exemple, on identifiera parfois $\underline{\omega}$ avec la suite $(f_{\omega_n})_{n \geqslant 0}$, dans ce cas on peut simplement écrire $\underline{\omega} = (f_n)_{n \geqslant 0}$.

Comme pour les cocycles, on utilisera la notation

$$f^n_{\underline{\omega}} = f_{\omega_{n-1}} \circ \cdots \circ f_{\omega_0}.$$

Si en outre X est un espace topologique et si pour presque tout ω, f_ω est continue on parlera de système dynamique aléatoire topologique. Noter que dans ce cas la topologie considérée sur $\Sigma_+ \times X$ est la topologie produit (et le cas échéant la tribu est la tribu borélienne associée).

1.4. — En pratique X sera toujours un espace métrique et $\mathcal{B}$ la tribu borélienne. Il est à noter que la tribu produit sur $\Omega^{\mathbb{N}}$ est engendrée par les cylindres, c'est à dire les sous-ensembles de la forme $\prod_{i \in \mathbb{N}} \Omega_i$, où $\Omega_i = \Omega$ sauf pour un nombre fini d'indices i, pour lesquels Ω_i est un élément de la tribu $\mathcal{A}$. Par souci de simplicité nous noterons les cylindres $\Omega_{i_1} \times \cdots \times \Omega_{i_k}$ en négligeant les indices pour lesquels $\Omega_i = \Omega$. Si Ω est un espace métrique compact et $\mathcal{A}$ sa tribu borélienne, alors, muni de sa topologie produit, $\Omega^{\mathbb{N}}$ est compact (et métrisable), et la tribu produit des tribus boréliennes coincide avec la tribu borélienne associée à la topologie produit (voir la section 6.4 de [**2**]).

Si l'on souhaite travailler dans des espaces plus généraux, comme nous allons faire un certain nombre de constructions faisant intervenir des produits infinis, il est néanmoins sage de se limiter au cas des espaces probabilités standard (voir si besoin le II.2.7 pour la définition).

1.5. — La théorie des systèmes dynamiques aléatoires a des liens profonds avec celle des marches aléatoires sur les groupes. Soit G un groupe localement compact à base dénombrable (typiquement un groupe de type fini) et ν une mesure de probabilité sur G. Supposons que G agisse mesurablement sur un espace mesurable $(X, \mathcal{B})$, et notons $\rho : G \to \mathrm{Aut}(X)$ le morphisme associé. Alors on a un système dynamique aléatoire induit par (G,ν) sur X, et les trajectoires $g_n\cdots g_0$ de la marche aléatoire donnent lieu aux itérations aléatoires $\rho(g_n)\circ\cdots\circ\rho(g_0)$.

1.6. — Plus généralement, il est utile dans ce contexte d'utiliser le vocabulaire des **semigroupes**. On rappelle qu'un semigroupe est un ensemble muni d'une loi de composition associative et possédant un élément neutre (la "multiplication"). L'exemple qui va nous intéresser est l'ensemble $\mathcal{F}(X,X)$ des applications (resp. mesurables, continues, etc.) de X dans X muni de la composition. Une **action** d'un semigroupe G sur X est simplement un morphisme de G dans $\mathcal{F}(X,X)$.

Un **semigroupe topologique** est un semigroupe G muni d'une topologie telle que la multiplication $G\times G\to G$ est continue. Si X est métrique compact, l'ensemble $\mathcal{C}(X,X)$ des applications continues, muni de la topologie de la convergence uniforme, est un semigroupe topologique. On peut dans ce cas parler de mesures sur G, implicitement muni de sa tribu borélienne.

1.7. — Un système dynamique aléatoire est un cas particulier de chaîne de Markov, dont les probabilités de transition sont données par $P(x,\cdot)=\int\delta_{f_\omega(x)}\mathrm{d}\nu(\omega)$, autrement dit $P(x,B)=\nu(\{\omega\in\mathcal{A},\ f_\omega(x)\in B\})$. On peut donc appliquer les idées et le vocabulaire de la théorie des chaînes de Markov. Toutefois (1) l'approche par les systèmes dynamiques apporte un éclairage nouveau, et des problématiques légèrement différentes et (2) dans le cas de probabilités de transition discrètes sur un espace continu, la plupart des concepts usuels de la théorie des chaînes de Markov (irréductibilité, etc.) s'appliquent difficilement.

Il est à noter que sous des hypothèses de régularité très faibles sur X et les probabilités de transition, toute chaîne de Markov peut en fait être représentée par un système dynamique aléatoire (voir le livre de Kifer [**8**, §I.1.1])

2. Mesures stationnaires

Remarque préliminaire. — Les résultats de cette section valent pour des chaînes de Markov générales. Attention néanmoins à la terminologie: nous appellerons mesure stationnaire ce qui est habituellement désigné dans la théorie des chaînes de Markov sous le nom de mesure invariante. Le terme "mesure invariante" sera réservé pour une notion plus restrictive (voir la section suivante).

2.1. — On considère un système dynamique aléatoire général comme au §1.3. Définissons des opérateurs de transition agissant respectivement sur les fonctions et les mesures sur X: si φ est une fonction sur X on pose

$$P^\bullet\varphi : x\mapsto P^\bullet\varphi(x)=\int\varphi(f_\omega(x))\mathrm{d}\nu(\omega)$$

et si μ est une mesure σ-finie sur $(X,\mathcal{B})$, pour $B\in\mathcal{B}$ on pose

$$P_\bullet\mu(B)=\int\mu(f_\omega^{-1}B)\mathrm{d}\nu(\omega).$$

On voit donc que $P_\bullet\mu=\int(f_\omega)_*\mu\,\mathrm{d}\nu(\omega)$ et en reprenant les notations du Chapitre VII, on peut écrire $P_\bullet\mu=\nu*\mu$. En appliquant le théorème de Fubini à $\{(\omega,x),f_\omega(x)\in B\}$ on déduit que pour $\varphi = \mathbf{1}_B$

$$\langle \mu, P^\bullet \varphi \rangle = \langle P_\bullet \mu, \varphi \rangle,$$

relation qui s'étend à toutes les fonctions mesurables bornées par les arguments habituels: prendre des combinaisons linéaires, des limites monotones et séparer en partie positive et négative.

2.2. *Définition.* — Une mesure[(1)] $P_\bullet$-invariante sera dite **stationnaire** (ou si le besoin s'en fait sentir, ν-stationnaire). Une fonction $P^\bullet$-invariante est dite ν-**harmonique**.

2.3. — Dans le contexte d'une action d'un groupe (G, ν) comme au §1.5, l'opérateur de transition $\mu \mapsto P_\bullet \mu = \int \rho(g)_* \mu \mathrm{d}\nu(g)$ devient un opérateur de convolution et on note $P_\bullet \mu = \nu * \mu$ (cf. le chapitre VII).

2.4. *Proposition. — Si $(X, (f_\omega), \nu)$ est un système dynamique aléatoire comme ci-dessus, une mesure de probabilité μ sur X est ν-stationnaire si et seulement si la mesure $\nu^{\mathbb{N}} \times \mu$ est F_+-invariante, où F_+ est le produit semi-direct associé.*

Démonstration. — L'invariance of $\nu^{\mathbb{N}} \times \mu$ est équivalente à l'égalité

$$(\nu^{\mathbb{N}} \times \mu)(F_+^{-1}(C \times A)) = \nu^{\mathbb{N}} \times \mu(C \times A) = \left(\prod_{j=0}^{N} \nu(C_j)\right) \cdot \mu(A), x \tag{1}$$

pour tous les cylindres $C = C_0 \times \cdots \times C_N$ dans $\Omega^{\mathbb{N}}$ et tout $A \in \mathcal{B}$. Par définition

$$F_+^{-1}(C \times A) = \left\{(\omega, x) \in \Omega^{\mathbb{N}} \times X \ ; \ f_{\omega_N} \in C_{N-1}, \dots, f_{\omega_1} \in C_0, f_{\omega_0}(x) \in A\right\},$$

ainsi

$$(\nu^{\mathbb{N}} \times \mu)(F_+^{-1}(C \times A)) = \left(\prod_{j=0}^{N} \nu(C_j)\right) \cdot (\nu \times \mu) \left(\{(\omega, x) \ ; \ f_\omega(x) \in A\}\right)$$

Finalement par le théorème de Fubini

$$(\nu \times \mu) \left(\{(\omega, x) \ ; \ f_\omega(x) \in A\}\right) = \int \mathbf{1}_{f_\omega^{-1}(A)}(x) \, \mathrm{d}\nu(\omega) \, \mathrm{d}\mu(x) = \int \mu(f_\omega^{-1}(A)) \, \mathrm{d}\nu(\omega)$$

et on conclut que $\nu^{\mathbb{N}} \times \mu$ est F_+ invariante si et seulement si pour tout $A \in \mathcal{B}$ $\mu(A) = \int \mu(f_\omega^{-1}(A)) \, \mathrm{d}\nu(f_\omega)$, c'est à dire que μ est stationnaire. □

2.5. — On a dans ce contexte un analogue du théorème de Krylov-Bogolyubov.

Proposition. — Si X est un espace métrique compact et f_ω est continue pour presque tout ω, alors le système dynamique aléatoire $(X, (f_\omega), \nu)$ admet une mesure stationnaire.

Démonstration. — On raisonne exactement comme au Théorème II.1.6. Pour cela on montre d'abord une propriété de continuité de $P_\bullet$: si (μ_n) est une suite de mesures de probabilité et $\mu_n \rightharpoonup \mu$ alors $P_\bullet \mu_n \rightharpoonup P_\bullet \mu$. En effet pour presque tout ω, $(f_\omega)_* \mu_n \rightharpoonup (f_\omega)\mu$, donc si φ est une fonction continue sur X on a par théorème de convergence dominée

$$\langle P_\bullet \mu_n, \varphi \rangle = \int \langle (f_\omega)_* \mu_n, \varphi \rangle \, \mathrm{d}\nu(\omega) \underset{n \to \infty}{\longrightarrow} \int \langle (f_\omega)_* \mu, \varphi \rangle \, \mathrm{d}\nu(\omega) = \langle P_\bullet \mu, \varphi \rangle.$$

---

(1) Rappelons que dans ce cours, sauf mention du contraire, toutes les mesures sont supposées finies.

Ensuite on prend une probabilité μ_0 quelconque sur X, on pose $\mu_n = \frac{1}{n}\sum_{k=0}^{n-1} P_\bullet^k \mu_0$ et par linéarité et continuité de $P_\bullet$ on voit que toute valeur d'adhérence de la suite μ_n est $P_\bullet$-invariante, d'où le résultat. □

2.6. *Remarque.* — Pour un système dynamique aléatoire topologique sur un compact X comme ci-dessus l'ensemble des mesures de probabilité stationnaires est un compact convexe de $\mathcal{C}(X,\mathbb{R})^*$ pour la topologie faible. En particulier par le théorème de Choquet toute probabilité stationnaire s'écrit comme une combinaison convexe de probabilités stationnaires extrémales. Comme dans le cas classique, celles-ci sont appelées ergodiques et seront étudiées plus en détail au §5 ci-dessous.

3. Mesures invariantes

3.1. — Considérons un système dynamique aléatoire $(X,(f_\omega),\nu)$. Nous dirons qu'une mesure μ sur X est **ν-p.s. invariante**, ou encore **presque invariante** ou plus simplement **invariante**, si pour ν-presque tout ω on a $(f_\omega)_*\mu = \mu$.

Remarquer que cette notion n'aurait pas vraiment de sens pour une chaîne de Markov. La propriété suivante est évidente:

Proposition. — Toute mesure invariante est stationnaire

3.2. — La réciproque de cette proposition est fausse: en effet plaçons nous sur $X = [0,1]$ et définissons

$$f_1 : x \mapsto \frac{x}{2},\ f_2 : x \mapsto \frac{1+x}{2} \text{ et } \nu = \frac{1}{2}\left(\delta_{f_1} + \delta_{f_2}\right).$$

Alors la mesure de Lebesgue sur $[0,1]$ est stationnaire, mais non invariante. Il est important de remarquer que cette non-invariance est liée à une propriété de **contraction** de la dynamique (sur ce sujet voir également la discussion du §3.12, ainsi que le chapitre suivant).

Exercice. — (cf. le chapitre VII) Soit ν une mesure de probabilité sur $\mathrm{GL}_2(\mathbb{R})$ telle que le sous-groupe engendré par $\mathrm{Supp}(\nu)$ soit non élémentaire. Soit μ l'unique mesure ν-stationnaire sur $\mathbb{P}^1(\mathbb{R})$. Montrer que μ n'est pas invariante.

3.3. — La tension entre invariance et stationnarité est un aspect fondamental de l'étude ergodique des systèmes dynamiques aléatoires, et qui joue un rôle particulièrement important en théorie géométrique des groupes (via les marches aléatoires). Il est notamment intéressant de caractériser les situations où toute mesure stationnaire (finie) est automatiquement invariante. Nous dirons qu'un système dynamique aléatoire ayant cette propriété est **raide** (**stiff** en anglais; cette terminologie est due à Furstenberg). Commençons par un exemple simple.

3.4. *Proposition. — Si $(X,(f_\omega),\nu)$ est un système dynamique aléatoire sur un ensemble fini, tel que pour tout ω, f_ω est bijective. Alors toute mesure de probabilité stationnaire est invariante.*

Démonstration. — Soit μ une mesure de probabilité stationnaire sur X, il s'agit de montrer que μ est invariante. D'après le théorème de Choquet (que l'on applique ici en dimension finie), il suffit de considérer le cas où μ est extrémale. Remarquer que l'ensemble des applications de X dans X est fini, donc ν est une mesure à support fini. Si l'on écrit $\mu = \sum_{x\in X} \mu_x \delta_x$ avec $0 \leqslant \mu_x \leqslant 1$ pour tout x, alors pour $f : X \to X$ on obtient

$$f_\star\mu = \sum_{x\in X} \mu_x \delta_{f(x)} = \sum_{x'\in X} \mu\left(\left\{f^{-1}(x')\right\}\right)\delta_{x'}$$

et donc

$$P_\bullet\mu = \sum_{x\in X} \mu_x \int \delta_{f_\omega(x)} \mathrm{d}\nu(\omega) = \sum_{x'\in X}\left(\int \mu\left(\left\{f_\omega^{-1}(x')\right\}\right)\mathrm{d}\nu(\omega)\right)\delta_x = \mu = \sum_{x\in X}\mu_x\delta_x. \tag{2}$$

Remarquons déjà que pour tout $x \in \mathrm{Supp}(\mu)$ la première égalité dans (2) implique que pour tout $\omega \in \mathrm{Supp}(\nu)$, $f_\omega(x) \in \mathrm{Supp}(\mu)$. Posons $\mu_0 = \min\{\mu_x,\ x \in \mathrm{Supp}(\mu)\}$ et $X_0 = \{x,\ \mu_x = \mu_0\}$. Alors par définition de μ_0, pour tout $(\omega, x) \in \mathrm{Supp}(\mu) \times \mathrm{Supp}(\nu)$ on a $\mu(\{f_\omega^{-1}(x)\}) \geqslant \mu_0$, et donc par (2) il vient que pour pour tout $x \in X_0$ et tout $\omega \in \mathrm{Supp}(\nu)$, on a $\mu(\{f_\omega^{-1}(x)\}) = \mu_0$, soit $f_\omega^{-1}(x) \subset X_0$, et donc $f_\omega^{-1}(X_0) \subset X_0$. Comme $\int \mu(f_\omega^{-1}(X_0))\mathrm{d}\nu(\omega) = \mu(X_0)$ on obtient que pour tout $\omega \in \mathrm{Supp}(\nu)$, $f_\omega^{-1}(X_0) = X_0$ et donc X_0 est f_ω-invariant. Comme μ est extrémale, on en déduit que $\mu = \mu|_{X_0}$ et finalement, comme pour tout $x \in X_0$ et presque tout ω on a $\mu(\{f_\omega^{-1}(x)\}) = \mu_0 = \mu(\{x\})$, on conclut que μ est invariante. □

Remarque. — Le résultat ne vaut pas sans l'hypothèse de bijectivité, comme le montre l'exemple $X = \{0, 1\}$, $f_0 : X \mapsto \{0\}$, $f_1 : X \mapsto \{1\}$, $\nu = \frac{1}{2}(\delta_{f_0} + \delta_{f_1})$ et $\mu = \frac{1}{2}(\delta_0 + \delta_1)$.

3.5. — Un thème important en théorie géométrique des groupes est de savoir si la propriété de raideur peut être assurée par les propriétés du semigroupe Γ engendré par les f_ω (resp. du groupe G si celles-ci sont inversibles). Il peut s'agir de propriétés algébriques, topologiques, ou métriques de ce groupe. L'idée fondamentale est que pour tout sous-ensemble mesurable B de X, la fonction $h = \gamma \mapsto \mu(\gamma^{-1}B)$ est une fonction ν-harmonique bornée sur Γ: elle vérifie $h(\gamma) = \int h(\gamma\gamma')\mathrm{d}\nu(\gamma')$ (exercice!). Il en est de même pour la fonction $\gamma \mapsto \int \varphi(\gamma x)\mathrm{d}\mu(x)$ pour φ une fonction mesurable bornée sur X. Il s'agit donc d'assurer que **toute fonction ν-harmonique bornée est constante** (propriété dite de Liouville).

Nous allons dans les paragraphes suivants voir quelques exemples de tels résultats. Pour plus d'informations sur ce sujet on peut consulter l'article classique de Kaimanovich et Vershik [**7**].

3.6. *Théorème (Choquet-Deny). — Soit G un semigroupe* ***commutatif*** *localement compact et à base dénombrable agissant sur un espace métrique X localement compact et séparable, et ν une mesure de probabilité sur G. Alors pour le système dynamique aléatoire engendré par ν, toute mesure stationnaire sur X est invariante.*

L'hypothèse topologique sur G est satisfaite par exemple pour un système dynamique aléatoire engendré par un nombre fini de transformations f_ω.

Démonstration. — Soit Γ le semigroupe engendré par les f_ω et montrons qu'une fonction ν-harmonique bornée sur Γ est p.s. constante. Remarquer qu'il suffit de considérer le cas de fonctions du type $\gamma \mapsto \int \varphi(\gamma x) \mathrm{d}\mu(x)$ avec φ continue à support compact, donc on peut se limiter aux fonctions harmoniques continues sur Γ. Les hypothèses sur G et X garantissent que le théorème de représentation de Choquet s'applique[2] Ainsi, d'après le théorème de Choquet il suffit de démontrer le résultat pour une fonction ν-harmonique extrémale. Soit h une telle fonction. On observe que pour tout $\gamma_0 \in \Gamma$, $h_{\gamma_0} : \gamma \mapsto h(\gamma\gamma_0)$ est également harmonique. En effet

$$h_{\gamma_0}(\gamma) = h(\gamma\gamma_0) = h(\gamma_0\gamma) = \int h(\gamma_0\gamma\gamma')\mathrm{d}\nu(\gamma') = \int h_{\gamma_0}(\gamma\gamma')\mathrm{d}\nu(\gamma').$$

Mais par ailleurs l'harmonicité de h signifie que $h = \int h_{\gamma'}\mathrm{d}\nu(\gamma')$, et comme h est extrémale on a $h = h_{\gamma'}$ pour presque tout γ'. D'après les remarques ci-dessus on obtient que pour toute fonction $\varphi \in \mathcal{C}_c(X)$ on a presque sûrement $\int \varphi(\gamma x)\mathrm{d}\mu(x) = \int \varphi(x)\mathrm{d}\mu(x)$. Appliquant ceci à une famille dénombrable dense de fonctions continues à support compact, on conclut que pour presque tout γ, pour toute fonction $\varphi \in \mathcal{C}_c(X)$, $\int \varphi(\gamma x)\mathrm{d}\mu(x) = \int \varphi(x)\mathrm{d}\mu(x)$. En d'autres termes, pour ν-presque tout γ on a $\gamma_\star\mu = \mu$: μ est invariante. □

3.7. *Exemple:* $\times 2 \times 3$. — Considérons le semigroupe Γ engendré par les applications $M_2 : x \mapsto 2x$ et $M_3 : x \mapsto 3x$ sur le cercle $\mathbb{T} = \mathbb{R}/\mathbb{Z}$. D'après le théorème précédent, pour toute mesure de probabilité sur ce semi-groupe, toute mesure stationnaire est invariante par M_2 et M_3. On vérifie facilement que la mesure de Haar est un exemple d'une telle mesure invariante. On peut également trouver des mesures invariantes atomiques: en effet pour tout $x \in \mathbb{Q}/\mathbb{Z}$, l'orbite $\Gamma \cdot x$ est finie (car les dénominateurs n'augmentent pas) et porte donc une mesure invariante. Un des problèmes les plus célèbres de la théorie ergodique, la "conjecture $\times 2 \times 3$ de Furstenberg" est la réciproque de cette assertion: *toute mesure invariante par M_2 et M_3 sur le cercle est une combinaison de la mesure de Haar et de mesures atomiques.* Parmi les résultats positifs en direction de cette conjecture on peut citer les deux résultats suivants:

– (Furstenberg 1967) Tout ensemble infini invariant par M_2 et M_3 est dense dans $\mathbb{T}$, ou de manière équivalente, si $x \notin \mathbb{Q}/\mathbb{Z}$, $\{2^m 3^n x,\ m, n \geqslant 0\}$ est dense dans $\mathbb{T}$.
– (Rudolph 1990) Si μ est une mesure M_2 et M_3 invariante, qui est ergodique pour ce semi-groupe (voir le §5) et d'entropie strictement positive pour M_2 ou M_3 alors μ est la mesure de Haar.

3.8. *Remarque.* — On dit qu'un groupe a la **propriété de Choquet-Deny** si pour toute mesure ν engendrant G en tant que semigroupe, toute fonction ν-harmonique bornée est constante. La caractérisation algébrique des groupes de type fini ayant cette propriété a été obtenue récemment[3]: un groupe de type fini a la propriété de Choquet-Deny si et seulement s'il est virtuellement nilpotent, c'est à dire qu'il admet un sous-groupe nilpotent d'indice fini.

---

[2] C'est un point en fait un peu délicat, pour les détails voir Deny *Sur l'équation de convolution $\mu = \mu \star \sigma$.* Séminaire Brelot-Choquet-Deny. Théorie du potentiel, tome 4 (1959-1960), exp. no 5, p. 1-11.

[3] Voir: J. Frisch, Y. Hartman, O. Tamuz, P.V. Ferdowski, *Choquet-Deny groups and the infinite conjugacy class property*, Ann. Math 190 (2019), 307–320.

3.9. *Théorème. — Soit G un groupe* **compact** *agissant topologiquement sur un espace X localement compact et ν une mesure de probabilité sur G. Alors toute mesure ν-stationnaire sur X est invariante.*

Avant d'aborder la preuve, voici un résultat préliminaire:

3.10. *Lemme. — Un sous-semigroupe fermé d'un groupe compact est un sous-groupe*

Démonstration. — En effet soit G un groupe compact, Γ un sous semi-groupe fermé, et $\gamma \in \Gamma$. Soit n_j une sous-suite telle que $\gamma^{n_j} \to \gamma_\infty$. Alors $\gamma^{n_{j+1}-n_j} \to \mathrm{id}$, et donc $\gamma^{n_{j+1}-n_j-1} \to \gamma^{-1}$. Comme $n_{j+1} - n_j \geqslant 1$, on voit que $\gamma^{n_{j+1}-n_j-1} \in \Gamma$, et ainsi $\gamma^{-1} \in \Gamma$, comme annoncé. $\square$

Démonstration du Théorème 3.9. — Comme au théorème précédent, il suffit de montrer que toute fonction ν-harmonique de la forme $\gamma \mapsto \int \varphi(\gamma x) d\mu(x)$, $\varphi \in \mathcal{C}_c(X)$ est constante sur Γ_ν. Remarquer qu'une telle fonction est continue par rapport à γ. L'argument va reposer ici sur le **principe du maximum**. Soit donc h une fonction harmonique continue sur G et Γ_ν l'adhérence du sous-semigroupe engendré par $\mathrm{Supp}(\nu)$. D'après le lemme précédent, Γ_ν est un sous-groupe. Soit $\gamma \in \Gamma_\nu$ tel que $h(\gamma) = \max_{\Gamma_\nu} h$. Comme $h(\gamma) = \int h(\gamma\gamma') d\nu(\gamma')$ pour ν-presque tout γ' on en déduit $h(\gamma\gamma') = h(\gamma)$. Comme h est continue on en déduit que cette relation vaut pour tout $\gamma' \in \mathrm{Supp}(\nu)$, et par itération et passage à l'adhérence, elle vaut pour tout $\gamma' \in \Gamma_\nu$. On conclut que h est constante sur Γ_ν, et donc comme au théorème précédent que pour ν-presque tout γ, $\gamma_*\mu = \mu$, ce qu'il fallait démontrer. $\square$

3.11. *Remarque.* — Dans les deux derniers théorèmes, si ν est à support fini on peut se dispenser de l'hypothèse de compacité locale sur X et supposer que X est un espace mesurable général. *Exercice:* vérifier cette affirmation.

3.12. — Plutôt que des hypothèses générales sur le groupe, on peut également chercher des hypothèses sur son **action** impliquant la propriété de raideur des mesures stationnaires. Voici un énoncé qui généralise le Théorème 3.9: on dit qu'une action d'un groupe G sur un espace métrique X est **distale** si pour tous $x_1 \neq x_2$ dans X, $\inf_g d(gx_1, gx_2) > 0$. Furstenberg[(4)], a montré que toute action distale sur un espace métrique compact est raide.

Plus récemment, Bourgain, Furman, Lindenstrauss et Mozes[(5)] ont montré que si ν est une mesure à support fini sur $\mathrm{SL}(2, \mathbb{Z})$ engendrant un sous groupe non-élémentaire Γ alors l'action naturelle de ν sur le tore $\mathbb{T}^2$ est raide. En outre, les seules mesures invariantes sont la mesure de Haar et des mesures atomiques portées par $\Gamma \cdot x$, où $x \in \mathbb{Q}^2/\mathbb{Z}^2$. (Il s'agit donc en quelque sorte de la résolution de l'analogue de la conjecture $\times 2 \times 3$ dans ce contexte.) D'après l'exercice 3.2, il s'agit bien d'une propriété de l'action et non du groupe. Ce résultat a été la source de nombreux autres développements sur ce sujet.

(4) *Stiffness of group actions*, in *Lie groups and ergodic theory (Mumbay 1996)*, Tata Inst. Fund. Res. pp 105–117

(5) *Stationary measures and equidistribution for orbits of nonabelian semigroups on the torus.* J. Amer. Math. Soc., 24(2011):231–280.

3.13. — Une notion intimement liée à ces questions est celle de moyennabilité. Un groupe topologique (localement compact à base dénombrable) G est **moyennable** si toute action continue de G sur un espace métrique compact admet une mesure invariante. D'après ce qui précède, les groupes abéliens et les groupes compacts sont moyennables, $\mathrm{SL}_2(\mathbb{Z})$ et les groupes libres ne le sont pas.

Exercice. — Soit G un groupe discret. Par définition, une suite de Følner est une suite de sous-ensembles finis $F_n \subset G$ tels que pour tout $g \in G$

$$\frac{\#(gF_n \cap F_n)}{\#F_n} \underset{n\to\infty}{\longrightarrow} 1.$$

Montrer que si G admet une suite de Følner alors G est moyennable.

4. Argument de martingale, proximalité, et extension naturelle

Dans cette section nous considérons un système dynamique aléatoire $((f_\omega), \nu)$ sur un espace métrique compact X, muni d'une mesure de probabilité (borélienne) stationnaire μ.

4.1. — Le lemme suivant a déjà été rencontré au chapitre VII (voir le §3.7), où il a joué un rôle crucial (on peut en fait généraliser la forme plus forte du Lemme 3.7, voir l'exercice 4.4 ci-dessous).

Lemme. — Pour $\nu^{\mathbb{N}}$-presque tout $\underline{\omega} = (\omega_n)_{n\geqslant 0}$, la suite de mesures $(f_{\omega_0})_* \cdots (f_{\omega_n})_*\mu$ converge vers une mesure $\mu_{\underline{\omega}}$. En outre, la dépendance $\underline{\omega} \mapsto \mu_{\underline{\omega}}$ est borélienne, et

$$\int \mu_{\underline{\omega}} \mathrm{d}\nu^{\mathbb{N}}(\underline{\omega}) = \mu.$$

Démonstration. — La convergence découle du théorème de convergence des martingales, exactement comme au VII.3.7 [6]. La deuxième affirmation vient du fait que $\omega \mapsto (f_{\omega_0})_* \cdots (f_{\omega_n})_*\mu$ est une application borélienne de Ω muni de sa tribu produit naturelle vers l'espace des mesures de probabilité sur X, qui est compact (exercice!) et la troisième découle de ce que pour tout n on a $\mathbb{E}\left((f_{\omega_0})_* \cdots (f_{\omega_n})_*\mu\right) = \mu$ et du théorème de convergence dominée (appliquer ceci à une fonction continue). □

4.2. *Lemme. — Avec les notations ci-dessus, μ est p.s. invariante si et seulement si pour $\nu^{\mathbb{N}}$-presque tout $\underline{\omega}$ on a $\mu_{\underline{\omega}} = \mu$.*

Démonstration. — Le sens direct découle directement du lemme précédent. Réciproquement si pour presque tout $\underline{\omega}$ on a $\mu_{\underline{\omega}} = \mu$ alors μ est invariante: en effet par le Lemme 4.1, pour ν-presque tout α on a $(f_\alpha)_*\mu_{\underline{\omega}} = \mu_{\alpha\underline{\omega}}$, où $\alpha\underline{\omega}$ désigne la suite $(\alpha, \omega_0, \omega_1, \ldots)$. Donc si presque sûrement $\mu_{\underline{\omega}} = \mu$, on en déduit que pour ν-presque tout α, $(f_\alpha)_*\mu = \mu$. □

[6] Attention les notations sont légèrement différentes

4.3. — En général, les mesures $\mu_{\underline{\omega}}$ sont singulières par rapport à μ[7]. Le cas extrême est le suivant: on dit que le système dynamique aléatoire est **proximal** si pour presque tout ω, $\mu_{\underline{\omega}}$ est une masse de Dirac. Ceci traduit une propriété de **contraction** de la dynamique et se produit par exemple pour l'exemple du §3.2 ou pour l'action projective d'un produit aléatoire de matrices 2×2, comme on l'a vu au chapitre VII. Nous verrons d'autres exemples de ce type au chapitre suivant.

4.4. *Exercice.* — [8]

(a) En s'inspirant de la preuve du Lemme 3.7 du chapitre VII, montrer que pour ν-presque tout α et pour $\nu^{\mathbb{N}}$-presque tout $\underline{\omega}$ on a

$$(f_{\omega_0})_* \cdots (f_{\omega_n})_*(f_\alpha)_*\mu \rightharpoonup \mu_{\underline{\omega}}.$$

(b) En déduire une nouvelle preuve du théorème de Choquet-Deny 3.6.

4.5. *Exercice.* — Si $(X, (f_\omega), \nu)$ est comme au 4.1, on a un système dynamique aléatoire induit sur l'espace $\mathcal{M}(X)$ des mesures de probabilité sur X: f_ω induit $(f_\omega)_*$. La famille $(\mu_{\underline{\omega}})$ induit une variable aléatoire sur $\mathcal{M}(X)$ dont la loi sera notée $\widehat{\mu}$. Montrer que $\widehat{\mu}$ est stationnaire pour $(\mathcal{M}(X), (f_\omega)_*, \nu)$ et que le système dynamique aléatoire $(\mathcal{M}(X), (f_\omega)_*, \nu, \widehat{\mu})$ est proximal.

4.6. — Soit G un groupe localement compact muni d'une mesure de probabilité ν, agissant sur un espace mesuré (X, μ), avec μ stationnaire. Pour $\phi \in L^\infty(X, \mu)$, la fonction $g \mapsto \int \phi(g \cdot x) \mathrm{d}\mu(x)$ est ν-harmonique et bornée. L'exercice précédent montre qu'à partir de toute telle action on peut en construire une proximale. Furstenberg a démontré (sous quelques hypothèses naturelles sur ν) qu'il existe une (X, μ)-action proximale “universelle”, ayant la propriété que pour toute fonction ν-harmonique bornée h sur G il existe $\hat{h} \in L^\infty(X, \mu)$ essentiellement unique telle que

$$h(g) = \int \hat{h}(g \cdot x) \mathrm{d}\mu(x).$$

Noter que cette équation est similaire à la représentation de Poisson pour les fonctions harmoniques dans le disque unité. Par analogie l'espace obtenu s'appelle le **bord de Poisson** de (G, ν). Par exemple d'après le théorème de Choquet-Deny si G est abélien le bord de Poisson est réduit à un point.

4.7. — Considérons maintenant le décalage bilatéral σ sur $\Sigma = \Omega^{\mathbb{Z}}$ (nous changeons la notation $\underline{\omega}$ en ξ pour distinguer suites finies et suites bi-infinies). On peut comme précédemment considérer le produit semi-direct F sur $\Omega^{\mathbb{Z}} \times X$ défini par

$$F(\xi, x) = (\sigma\xi, f_{\xi_0}(x)).$$

[7] C'est même le cas dès que μ n'est pas invariante, voir Furstenberg et Glasner *Stationary dynamical systems*, in *Dynamical numbers—interplay between dynamical systems and number theory*, Contemp. Math., 532, AMS (2010), pp. 1–28.

[8] Voir [**1**, 2.21 et 2.22] pour la solution

Remarquer que si les f_ω sont inversibles, contrairement à F_+ ce système dynamique est maintenant inversible. Par un argument identique à celui du Lemme 4.1, pour $\nu^{\mathbb{Z}}$-presque tout ξ, la limite

$$m_\xi := \lim_{n\to+\infty}(f_{\xi_{-1}}\circ\cdots\circ f_{\xi_{-n}})_*\mu = \lim_{n\to+\infty}(f^n_{\sigma^{-n}\xi})_*\mu \tag{3}$$

existe. On munit $\Sigma\times X$ de la tribu produit $\mathcal{F}=\mathcal{A}^{\otimes\mathbb{Z}}\otimes\mathcal{B}$ et on y définit une mesure m par

$$m=\int\delta_\xi\otimes m_\xi\,\mathrm{d}\nu^{\mathbb{Z}}(\xi),$$

en d'autres termes, si φ est une fonction mesurable bornée sur $\Sigma\times X$,

$$\int\varphi\,\mathrm{d}m=\int\varphi(\xi,x)\,\mathrm{d}m_\xi(x)\mathrm{d}\nu^{\mathbb{Z}}(\xi).$$

4.8. *Proposition. — La mesure m est l'unique mesure F-invariante sur $\Sigma\times X$ se projetant sur $\nu^{\mathbb{N}}\times\mu$ par la projection naturelle $\Sigma\times X\to\Sigma_+\times X$. En particulier on a*

$$\nu^{\mathbb{Z}}\times\mu \text{ est } F\text{-invariante} \Leftrightarrow m=\nu^{\mathbb{Z}}\times\mu \tag{4}$$

et en outre

$$m=\nu^{\mathbb{Z}}\times\mu\Leftrightarrow\mu \text{ est p.s. invariante.} \tag{5}$$

4.9. *Remarque.* — Dans le cas inversible le système dynamique $(\Sigma\times X, m, F)$ est en fait (mesurablement conjugué à) l'extension naturelle de $(\Sigma_+\times X,\nu^{\mathbb{N}}\times\mu,F_+)$ (notion définie au §2.7 du chapitre II).

Démonstration. — Commençons par montrer que $m=\lim_{n\to\infty}(F^n)_*(\nu^{\mathbb{Z}}\times\mu)$. En effet si φ est une fonction continue on a

$$\begin{aligned}\langle F^n_*(\nu^{\mathbb{Z}}\times\mu),\varphi\rangle &= \langle\nu^{\mathbb{Z}}\times\mu,\varphi\circ F^n\rangle\\ &=\int\varphi\left(\sigma^n\xi, f_{\xi_{n-1}}\circ\cdots\circ f_{\xi_0}(x)\right)\mathrm{d}\nu^{\mathbb{Z}}(\xi)\mathrm{d}\mu(x)\\ &=\int\varphi\left(\xi', f_{\xi'_{-1}}\circ\cdots\circ f_{\xi'_{-n}}(x)\right)\mathrm{d}\nu^{\mathbb{Z}}(\xi')\mathrm{d}\mu(x)\end{aligned}$$

où dans la dernière égalité on a fait le changement de variable $\xi'=\sigma^n(\xi)$ et utilisé l'invariance de $\nu^{\mathbb{Z}}\times\mu$ par $(\xi,x)\mapsto(\sigma^n(\xi),x)$. On en déduit le résultat souhaité par la convergence (3) et le théorème de convergence dominée.

Le fait que $m=\lim_{n\to\infty}(F^n)_*(\nu^{\mathbb{Z}}\times\mu)$ implique immédiatement que m est F-invariante.

Notons $p:\Sigma\times X\to\Sigma_+\times X$ la projection naturelle. On vérifie simplement que que $p\circ F=F_+\circ p$. Ainsi

$$p_*m=\lim_{n\to\infty}p_*F^n_*(\nu^{\mathbb{Z}}\times\mu)=\lim_{n\to\infty}(F^n_+)_*p_*(\nu^{\mathbb{Z}}\times\mu)=\nu^{\mathbb{N}}\times\mu.$$

Pour démontrer l'unicité, soit $\widetilde{m}$ une mesure F-invariante sur $\Sigma\times X$ telle que $p_*\widetilde{m}=\nu^{\mathbb{N}}\times\mu$. Pour démontrer que $\widetilde{m}=m$, il suffit de le vérifier sur un ensemble de la forme $C\times A$, où $C=\Omega_{i_1}\times\cdots\times\Omega_{i_k}$ est un cylindre. Alors par invariance de $\widetilde{m}$ on a pour tout n,

$\widetilde{m}(C \times A) = \widetilde{m}(F^{-n}(C \times A))$. Remarquer que pour $(\omega, x) \in \Sigma_+ \times X$, $p^{-1}(\omega, x)$ est l'ensemble des (ξ, x), où ξ a les mêmes coordonnées positives (i.e. le même "futur") que ω. Plus généralement, la tribu engendrée par p est la tribu des évènements "ne dépendant que du futur". Nous affirmons que si $n > \max |i_j|$ alors $F^{-n}(C \times A)$ appartient à cette tribu: en effet si $F^n(\xi, x)$ appartient à $\Omega_{i_1} \times \cdots \times \Omega_{i_k} \times A$ et ξ' a le même futur que ξ, alors il en est de même pour $F^n(\xi', x)$. Comme $p_*\widetilde{m} = p_*m$, m et $\widetilde{m}$ coincident sur les ensembles ne dépendant que du futur et Il s'ensuit que $\widetilde{m}(F^{-n}(C \times A)) = m(F^{-n}(C \times A))$. Par invariance, on en déduit que $\widetilde{m}(C \times A) = m(C \times A)$ et l'unicité est démontrée. La première implication dans (4) découle immédiatement de cette unicité, et l'implication réciproque est évidente.

Pour (5), remarquons que si μ est p.s. invariante, la convergence (3) implique que p.s. $m_\xi = \mu$ et donc $m = \nu^{\mathbb{Z}} \times \mu$. Finalement, supposons que $\nu^{\mathbb{Z}} \times \mu$ est F-invariante, et pour simplifier la rédaction, supposons que Ω est fini (le cas général est laissé en exercice au lecteur). Fixons $f \in \Omega$ et posons C le cylindre composé des suites ξ telles que $\xi_{-1} = f$. Alors $\nu^{\mathbb{Z}}(C) = \nu(\{f\})$. Pour $A \in \mathcal{B}$, on a

$$\begin{aligned} F^{-1}(C \times A) &= \{(\xi, x),\ \sigma\xi \in C,\ f_{\xi_0}(x) \in A\} \\ &= \left\{(\xi, x),\ \xi_0 = f,\ x \in f^{-1}(A)\right\} \\ &= \sigma^{-1}(C) \times f^{-1}(A). \end{aligned}$$

Ainsi

$$\begin{aligned} (\nu^{\mathbb{Z}} \times \mu)(F^{-1}(C \times A)) &= (\nu^{\mathbb{Z}} \times \mu)(\sigma^{-1}(C) \times f^{-1}(A)) = \nu(\{f\})\mu(f^{-1}A) \\ &= (\nu^{\mathbb{Z}} \times \mu)(C \times A) = \nu(\{f\})\mu(A) \end{aligned}$$

et on a bien $\mu(f^{-1}A) = \mu(A)$ pour presque tout f. □

4.10. *Exercice.* — Soient $f_1, \ldots, f_k$ un nombre fini de contractions affines de $[0, 1]$ dans lui-même (c'est à dire de la forme $f_i(x) = a_i x + b_i$, avec $|a_i| < 1$). Soit ν une mesure de probabilité sur l'espace des applications affines de $\mathbb{R}$, de la forme $\nu = \sum_{i=1}^{k} \nu_i \delta_{f_i}$, avec $\nu_i > 0$. On considère le système dynamique aléatoire engendré par $((f_i), \nu)$ sur $[0, 1]$.

(a) Montrer que ce système dynamique aléatoire est proximal.
(b) Montrer qu'il admet une unique probabilité stationnaire μ.
(c) Montrer que si μ n'est pas réduite à une masse de Dirac, elle est diffuse et non invariante.
(d) Calculer l'exposant de Lyapunov de $((f_i), \nu, \mu)$ (cf. le chapitre suivant).

5. Ergodicité

Comme à la section précédente on se donne un système dynamique aléatoire $((f_\omega), \nu)$ sur un espace métrique compact X, muni d'une mesure de probabilité stationnaire μ. [9]

[9] Les résultats des §§ 5.1-5.4 valent en fait dans un cadre d'une action sur un espace mesurable.

5.1. — Opérons d'abord un retour sur la notion de fonction ν-harmonique. Rappelons que φ est ν-harmonique si $P^\bullet\varphi = \varphi$, c'est à dire pour tout x, $\varphi(x) = \int \varphi(f_\omega(x))\mathrm{d}\nu(\omega)$. Nous dirons que φ est **μ-presque ν-harmonique** (ou plus simplement **presque harmonique**) si cette relation a lieu μ-p.p. De même un sous-ensemble mesurable A sera dit **ν-invariant** (resp. **μ-presque ν-invariant**) si $\mathbf{1}_A$ est ν-harmonique (resp. $\mu-$presque ν-harmonique).

Remarquer que $P^\bullet\mathbf{1}_A = \mathbb{P}(f_\omega(\cdot) \in A)$, donc A est presque invariant si (et seulement si) pour μ-presque tout $x \in A$, ν-p.s. on a $f_\omega(x) \in A$. En d'autres termes, pour presque tout $(x,\omega) \in A \times \Omega$ on a $x \in f_\omega^{-1}(A)$. Donc pour ν-presque tout ω on a $A \subset f_\omega^{-1}(A)$ à un ensemble de mesure nulle près, donc $\mu(A) \leqslant \mu(f_\omega^{-1}(A))$. Mais par stationnarité on a $\mathbb{E}(\mu(f_\omega^{-1}(A))) = \mu(A)$. Ceci montre que A *est ν-presque invariant si (et seulement si) pour ν-presque tout ω, $\mu(A \triangle f_\omega^{-1}(A)) = 0$.*

5.2. *Exercice.* — Montrer que si A est presque invariant, alors $\mu|_A$ définit une mesure stationnaire.

5.3. *Lemme.* — *Toute fonction presque harmonique bornée est presque partout égale à une fonction harmonique.*

Démonstration. — Soit φ telle que $P^\bullet\varphi = \varphi$ μ-p.p., et soit E un ensemble de mesure totale où ces fonctions coincident. Remarquons déjà que pour μ-presque tout x, pour ν-presque tout ω on a $f_\omega(x) \in E$. Ceci découle du théorème de Fubini et de la relation de stationnarité
$$\int \mathbf{1}_E(f_\omega(x))\mathrm{d}\nu(\omega)\mathrm{d}\mu(x) = \int \mathbf{1}_E(x)\mathrm{d}\mu(x) = 1.$$
Soit E_2 l'ensemble des x satisfaisant la propriété précédente. Pour $x \in E \cap E_2$, on considère la relation $\varphi(x) = \int \varphi(f_\omega(x))\mathrm{d}\nu(\omega)$, et pour presque tout ω, comme $f_\omega(x) \in E$ on a
$$\varphi(f_\omega(x)) = \int \varphi(f_{\omega'}(f_\omega(x)))\mathrm{d}\nu(\omega')$$
et par Fubini, en intégrant par rapport à ω', on voit que $(P^\bullet)^2\varphi = P^\bullet\varphi$ p.p. On itère cette relation et pour tout $x \in X$ on pose
$$\varphi_\infty(x) = \liminf_{n\to\infty}(P^\bullet)^n\varphi(x).$$
Par ce qui précède, $\varphi = \varphi_\infty$ μ-p.p, et donc φ_∞ est presque harmonique. Le lemme de Fatou implique que $P^\bullet\varphi_\infty \leqslant \varphi_\infty$. Comme précédemment on peut itérer cette relation, et pour tout x on appelle $\psi(x)$ la limite décroissante de $(P^\bullet)^n\varphi_\infty(x)$. Finalement par le théorème de convergence monotone on a $P^\bullet\psi(x) = \psi(x)$ pour tout x, ce qu'il fallait démontrer. $\square$

Remarque. — Contrairement au cas des systèmes dynamiques classiques, si φ est une fonction indicatrice, on ne peut pas en général choisir pour ψ une fonction indicatrice. Autrement dit, en général un sous-ensemble presque invariant ne coincide pas p.p. avec un ensemble invariant. Pour comprendre le problème, revenons au cas d'un système dynamique déterministe: si A est essentiellement invariant on fabrique un ensemble invariant en posant $A_1 = \bigcap_{n\geqslant 0} f^{-n}(A)$, puis A_2 l'intersection décroissante des $f^{-n}(A_1)$, qui coincide avec A mod. 0 et est invariant. Dans le cas aléatoire, il faudrait faire l'intersection non-dénombrable des $f_{\omega_n}^{-1} \cdots f_{\omega_0}^{-1}(A)$, et l'argument ne conclut pas.

Exercice. — Fabriquer un contre-exemple.[10]

5.4. *Proposition-Définition.* — *Soit $(X, (f_\omega), \nu)$ un système dynamique aléatoire muni d'une probabilité stationnaire μ comme ci dessus. Alors les conditions suivantes sont équivalentes:*

(i) *tout sous-ensemble presque invariant est de mesure 0 ou 1;*
(ii) *toute fonction mesurable harmonique bornée est p.s. constante;*
(iii) *toute fonction L^∞ presque harmonique est constante (i.e. constante comme fonction dans L^∞);*
(iv) *toute fonction L^1 presque harmonique est constante (idem);*

Si l'une de ces conditions est satisfaite, on dira que le système dynamique aléatoire est ***ergodique****.*

Démonstration. — Les implications $(iv) \Rightarrow (iii) \Rightarrow (ii)$ et $(iii) \Rightarrow (i)$ sont évidentes. L'implication $(ii) \Rightarrow (iii)$ découle du Lemme 5.3.

Montrons que (i) implique (iii). Soit φ une fonction presque harmonique. Le point clé est de démontrer que l'ensemble $\{\varphi \geqslant 0\}$ est presque invariant. En appliquant ce résultat à $\varphi - c$, on obtient de même que pour tout $c \in \mathbb{R}$, $\{\varphi \geqslant c\}$ est presque invariant, donc de mesure 0 ou 1 par (i). On conclut alors comme dans le cas déterministe (voir le chapitre II). Pour démontrer l'invariance de $\{\varphi \geqslant 0\}$, remarquons que comme $P^\bullet\varphi = \varphi$ (dans L^∞), on a $P^\bullet|\varphi| \leqslant |\varphi|$. Mais $\langle \mu, P^\bullet|\varphi|\rangle = \langle \mu, |\varphi|\rangle$, donc on en déduit que $P^\bullet|\varphi| = |\varphi|$ μ-p.p. Comme $\varphi^+ = \frac{1}{2}(\varphi + |\varphi|)$, φ^+ est également presque harmonique. Pour conclure que $\{\varphi \geqslant 0\}$ est presque invariant, nous allons comparer les relations $P^\bullet\varphi = \varphi$ et $P^\bullet\varphi^+ = \varphi^+$. En effet, soit $E = \{\varphi \geqslant 0\}$. Pour tout x on a $\varphi(x) \leqslant \varphi^+(x)$ avec égalité ssi $x \in E$. Maintenant pour presque tout $x \in E$ on a par hypothèse

$$\int \varphi(f_\omega(x))\mathrm{d}\nu(\omega) = \int \varphi^+(f_\omega(x))\mathrm{d}\nu(\omega),$$

et ainsi pour ν-presque tout ω, $\varphi(f_\omega(x)) = \varphi^+(f_\omega(x))$, soit $f_\omega(x) \in E$. Par Fubini, on en déduit que ν-p.s. $E \subset f_\omega^{-1}(E)$ mod. 0 et donc E est presque invariant.

Pour montrer que (iii) implique (iv) on part d'une fonction $\varphi \in L^1$ telle que $P^\bullet\varphi = \varphi$ dans L^1 et en raisonnant comme au paragraphe précédent, on montre que pour tout $c \in \mathbb{R}$ les fonctions $\max(\varphi, c) = (\varphi - c)^+ + c$ et $\min(\varphi, c)$ sont presque harmoniques. Ainsi pour tous $c_1 < c_2$, $\min(\max(\varphi, c_1), c_2)$ est presque harmonique, donc p.s. constante, et on conclut. □

5.5. — Voici un énoncé utile.

[10] Voir [**1**, Rem. 2.7] pour la solution.

Proposition. — Soit $(X, (f_\omega), \nu)$ *un système dynamique aléatoire; on suppose que les* f_ω *sont p.s. inversibles. Si une probabilité stationnaire* μ *a un atome en* x_0*, alors il existe un ensemble fini* E *presque invariant contenant* x_0*, et* $\mu|_E$ *est invariante.*

Démonstration. — Soit $\mu_{\max}$ la masse maximale d'un atome de μ, et

$$E_{\max} = \{x, \mu(\{x\}) = \mu_{\max}\}.$$

qui est un sous-ensemble fini. Pour $x \in E_{\max}$, comme les f_ω sont inversibles, on a

$$\mu(\{x\}) = \int \mu(f_\omega^{-1}(\{x\}))\mathrm{d}\nu(\omega) = \int \mu(\{f_\omega^{-1}(x)\})\mathrm{d}\nu(\omega).$$

Comme pour tout y on a $\mu(\{y\}) \leqslant \mu_{\max}$, on en déduit que pour tout $x \in E_{\max}$, pour ν-presque tout ω, $f_\omega^{-1}(x)$ appartient à $E_{\max}$, et donc $E_{\max}$ est presque invariant. Ainsi $\mu|_{E_{\max}}$ définit une mesure stationnaire, qui est invariante par la Proposition 3.4. On réitère ensuite ce raisonnement avec la mesure stationnaire $\mu - \mu|_{E_{\max}}$, ce qui absorbera x_0 en un nombre fini d'étapes. □

5.6. — Considérons maintenant le produit semi-direct (non-inversible) F_+ associé à notre système dynamique aléatoire. Rappelons qu'il s'agit de la transformation $\nu^{\mathbb{N}} \times \mu$-invariante sur $\Omega^{\mathbb{N}} \times X$ définie par $F_+(\omega, x) = (\sigma\omega, f_{\omega_0}(x))$.

Théorème. — Soit $(X, (f_\omega), \nu, \mu)$ *un système dynamique aléatoire mesuré comme précédemment. Les conditions suivantes sont équivalentes:*

(i) *la mesure stationnaire* μ *est ergodique;*
(ii) *la mesure* $\nu^{\mathbb{N}} \times \mu$ *est* F_+*-ergodique.*

Le corollaire suivant est connu sous le nom de **théorème ergodique aléatoire**. Noter que le cas des mesures invariantes est déjà non trivial: il s'agit en fait de la version initiale du théorème ergodique aléatoire, due à Ulam, Von Neumann et Kakutani. Le cas général est dû à Ohno.

5.7. *Corollaire. — Sous les hypothèses du théorème précédent, si* $\varphi \in L^1(X, \mu)$*, pour* $(\nu^{\mathbb{N}} \times \mu)$*-presque tout* $(\underline{\omega}, x) = ((f_n), x) \in \Sigma_+ \times X$ *et dans* $L^1(\nu^{\mathbb{N}} \times \mu)$ *on a*

$$\frac{1}{n}\sum_{k=1}^{n} \varphi(f_{\omega_{n-1}} \circ \cdots \circ f_{\omega_1} \circ f_{\omega_0}(x)) \underset{n\to\infty}{\longrightarrow} \int \varphi \,\mathrm{d}\mu.$$

Dans le cas proximal, ce résultat contraste avec la distribution des points du type $f_{\omega_0} \cdots f_{\omega_n}(x)$ qui ont tendance à converger vers les mesures atomiques $\mu_{\underline{\omega}}$.

Démonstration. — Il suffit d'appliquer le théorème de Birkhoff à la fonction φ, vue comme une fonction ne dépendant que de la variable x dans $\Sigma_+ \times X$. □

Démonstration du théorème. — Le sens $(ii) \Rightarrow (i)$ est facile. En effet, soit $A \subset X$ est un sous-ensemble μ-presque ν-invariant, i.e. pour ν-presque tout ω, $\mu(f_\omega^{-1}A \triangle A) = 0$ et considérons $\widehat{A} = \Sigma_+ \times A$. Alors

$$F_+^{-1}(\widehat{A}) = \{(\underline{\omega}, x),\ f_{\omega_0}(x) \in A\} = \bigcup_{\underline{\omega} \in \Sigma_+} \{\underline{\omega}\} \times f_{\omega_0}^{-1}(A),$$

et par le théorème de Fubini on voit que $(\nu^{\mathbb{N}} \times \mu)\big(F_+^{-1}\widehat{A}\triangle\widehat{A}\big) = 0$. Comme $\nu^{\mathbb{N}} \times \mu$ est F_+-ergodique, on en déduit que $(\nu^{\mathbb{N}} \times \mu)\big(\widehat{A}\big) = \mu(A)$ vaut 0 ou 1, et le résultat suit.

Pour l'implication réciproque, donnons nous une fonction φ mesurable bornée et F_+-invariante sur $\Sigma_+ \times X$. Pour produire une fonction sur X on pose simplement

$$\check{\varphi} : x \longmapsto \int \varphi\,(\underline{\omega}, x)\,\mathrm{d}\nu^{\mathbb{N}}(\underline{\omega}).$$

Alors

$$\begin{aligned} P^{\bullet}\check{\varphi}(x) &= \int \check{\varphi}(f_\alpha(x))\mathrm{d}\nu(\alpha) = \int \varphi\,(\underline{\omega}, f_\alpha(x))\,\mathrm{d}\nu^{\mathbb{N}}(\underline{\omega})\mathrm{d}\nu(\alpha) \\ &= \int \varphi\,(\underline{\omega}', x)\,\mathrm{d}\nu^{\mathbb{N}}(\underline{\omega}') = \check{\varphi}(x) \end{aligned}$$

où dans la dernière intégrale on a posé $\underline{\omega}' = (\alpha, \omega_0, \ldots)$ (i.e. la concaténation de α et $\underline{\omega}'$) et on a utilisé la structure produit de $\nu^{\mathbb{N}}$ et l'invariance de φ. Ainsi $\check{\varphi}$ est ν-harmonique et par ergodicité de μ on déduit que $\check{\varphi}$ est μ-p.p. égale à la constante $\mathbb{E}(\varphi)$.

On raffine cette idée en introduisant pour $n \geqslant 0$

$$\check{\varphi}^{(n)}(\underline{\omega}, x) = \int \varphi(\underline{\omega}, x)\mathrm{d}\nu(\omega_n)\mathrm{d}\nu(\omega_{n+1})\cdots$$

L'invariance de φ dit que p.p. on a

$$\varphi(\underline{\omega}, x) = \varphi(\sigma^n\underline{\omega}, f_{\underline{\omega}}^n(x)) \text{ où } f_{\underline{\omega}}^n = f_{\omega_{n-1}} \circ \cdots \circ f_{\omega_0}$$

et donc

$$\begin{aligned} \check{\varphi}^{(n)}(\underline{\omega}, x) &= \int \varphi(\sigma^n\underline{\omega}, f_{\underline{\omega}}^n(x))\mathrm{d}\nu(\omega_n)\mathrm{d}\nu(\omega_{n+1})\cdots \\ &= \int \varphi(\underline{\omega}', f_{\underline{\omega}}^n(x))\mathrm{d}\nu^{\mathbb{N}}(\underline{\omega}') \text{ oú on fait le changement de variable } \underline{\omega}' = \sigma^n\underline{\omega} \\ &= \check{\varphi}(f_{\underline{\omega}}^n(x)) = \mathbb{E}(\varphi) \text{ p.p.} \end{aligned}$$

On peut par ailleurs présenter $\check{\varphi}^{(n)}$ comme une espérance conditionnelle en introduisant la tribu $\mathcal{F}_{n,+}$ engendrée par les variables $\omega_0, \ldots, \omega_{n-1}$ et x, et alors $\check{\varphi}^{(n)} = \mathbb{E}(\varphi|\mathcal{F}_{n,+})$. Remarquer que la tribu engendrée par les $\mathcal{F}_{n,+}$ est $\mathcal{F}_+$. Ainsi par le théorème de convergence des martingales, $\check{\varphi}^{(n)}$ converge $(\nu^{\mathbb{N}} \times \mu)$-p.p. vers φ. Du calcul précédent on déduit ainsi que φ est $(\nu^{\mathbb{N}} \times \mu)$-p.p. égale à une constante et l'ergodicité est établie. □

5.8. *Corollaire. — Une mesure stationnaire est ergodique si et seulement si c'est un point extrémal du convexe des mesures stationnaires.*

Démonstration. — Il est clair que si μ n'est pas extrémale parmi les mesures stationnaires alors $\nu^{\mathbb{N}} \times \mu$ n'est pas extrémale parmi les mesures F_+-invariantes. On en déduit la chaîne d'implications:

$$\mu \text{ ergodique } \Leftrightarrow \nu^{\mathbb{N}} \times \mu \text{ ergodique } \Leftrightarrow \nu^{\mathbb{N}} \times \mu \text{ extrémale } \Rightarrow \mu \text{ extrémale.}$$

Réciproquement si μ n'est pas ergodique il existe un ensemble μ-presque invariant non-trivial A et on vérifie simplement que la mesure $\mathbf{1}_A\mu$ est stationnaire (cf. l'exercice 5.2).

Ainsi la décomposition $\mu = \mathbf{1}_A\mu + \mathbf{1}_{A^c}\mu$ exprime μ comme une combinaison convexe non-triviale de mesures stationnaires. □

5.9. — On peut également énoncer le théorème ergodique aléatoire en termes du système dynamique inversible $(\Sigma \times X, F, m)$ introduit à la section 4:

Théorème. — Sous les hypothèses précédentes on a l'équivalence:

(i) *la mesure stationnaire μ est ergodique;*
(ii) *la mesure m est F-ergodique.*

Dans le cas inversible ce résultat découle directement du fait général que l'extension naturelle d'un système ergodique est ergodique (cf. la Remarque 4.9). La démonstration ci-dessous est essentiellement l'adaptation de ce fait général à notre contexte.

Démonstration. — L'implication $(ii) \Rightarrow (i)$ est facile, puisque si $E \subset \Sigma_+ \times X$ est un sous-ensemble F_+-invariant, alors $p^{-1}(E)$ est F-invariant (où $p : \Sigma \times X \to \Sigma_+ \times X$ est l'application naturelle), donc $m(p^{-1}(E)) = (\nu^{\mathbb{N}} \times \mu)(E)$ vaut 0 ou 1.

Pour l'implication réciproque, nous allons utiliser la caractérisation suivante de l'ergodicité (exercice): *m est ergodique si et seulement si pour toute fonction mesurable bornée φ, les moyennes de Birkhoff convergent m-p.p. vers une fonction constante p.p., si et seulement si $\mathbb{E}(\varphi|\mathcal{I})$ est égale à une constante m-p.p.* (on rappelle que $\mathcal{I}$ désigne la tribu des sous-ensembles F-invariants).

Pour $n \in \mathbb{Z}$, notons $\mathcal{F}_n$ la tribu engendrée par les ensembles de la forme $C \times A$, où C est un "cylindre" ne faisant intervenir que les coordonnées d'indice supérieur ou égal à n et $A \in \mathcal{A}$. En d'autres termes, une fonction est $\mathcal{F}_n$ mesurable si elle ne dépend que des variables ξ_k, $k \geqslant n$ et de la variable x. On a $\mathcal{F}_n \subset \mathcal{F}_{n-1}$ et la tribu engendrée par les $\mathcal{F}_n$ est la tribu produit $\mathcal{F}$ sur $\Sigma \times X$. En outre on vérifie simplement que pour tout n, $F^{-1}(\mathcal{F}_n) = \mathcal{F}_{n+1}$.

Soit maintenant φ une fonction mesurable bornée sur $\Sigma \times X$. Fixons $n \geqslant 0$ et montrons d'abord que $\mathbb{E}(\varphi|\mathcal{I})$ est constante dans le cas particulier où φ est $\mathcal{F}_{-n}$-mesurable. Si $n = 0$ c'est évident car une fonction $\mathcal{F}_0$-mesurable sur $\Sigma \times X$ s'identifie naturellement à une fonction mesurable pour la tribu produit de $\Sigma_+ \times X$ et m se projette sur la mesure ergodique $\nu^{\mathbb{N}} \times \mu$ (exercice: remplir les détails de ce raisonnement). Pour n quelconque on remarque que pour $k \geqslant n$ la fonction $\varphi \circ F^k$ est $\mathcal{F}_0$-mesurable, donc les moyennes de Birkhoff de $\varphi \circ F^k$ convergent m-p.p. vers une fonction constante. Mais alors il en est de même pour celles de φ (pourquoi?) et on conclut que $\mathbb{E}(\varphi|\mathcal{I}) = c$, m-p.p. Noter que $c = \mathbb{E}(\varphi)$

Pour une fonction φ mesurable générale, le théorème de convergence des martingales et le fait que la filtration $(\mathcal{F}_n)$ croît vers $\mathcal{F}$ assurent que $\mathbb{E}(\varphi|\mathcal{F}_{-n})$ converge m-p.p. vers φ. Ainsi par les propriétés de convergence de l'espérance conditionnelle et le fait que $\mathbb{E}(\varphi|\mathcal{F}_{-n})$ est $\mathcal{F}_{-n}$-mesurable, on déduit que

$$\mathbb{E}(\varphi|\mathcal{I}) = \lim_{n\to\infty} \mathbb{E}(\mathbb{E}(\varphi|\mathcal{F}_n)|\mathcal{I}) = \mathbb{E}(\varphi),$$

ce qui termine la preuve. □

6. Théorème ergodique de Breiman

L'objet de cette section est de démontrer le théorème suivant, dû à Breiman (qui vaut dans le contexte plus général des chaînes de Markov, voir le livre de Benoist-Quint [**1**]).

6.1. *Théorème. — Soit X un espace métrique compact et $((f_\omega), \nu)$ un système dynamique aléatoire topologique sur X. Alors pour tout x dans X, pour $\nu^{\mathbb{N}}$-presque tout $\underline{\omega} \in \Omega^{\mathbb{N}}$, toute valeur d'adhérence de la suite*

$$\frac{1}{n}\sum_{k=1}^{n} \delta_{f^k_{\underline{\omega}}(x)} \text{ (on rappelle que } f^k_{\underline{\omega}} = f_{\omega_{k-1}} \circ \cdots \circ f_{\omega_0}\text{)}$$

est une mesure stationnaire.

Noter que si on moyenne par rapport à $\underline{\omega}$, c'est à dire que l'on s'intéresse à la convergence de $\frac{1}{n}\sum_{k=1}^{n} P^k_\bullet \delta_x$, le résultat est facile et contenu dans la Proposition 2.5.

6.2. — Le corollaire suivant s'applique par exemple aux produits aléatoires de matrices.

Corollaire. — Si $(X, (f_\omega), \nu)$ admet une unique probabilité stationnaire μ, alors pour tout $x \in X$, pour $\nu^{\mathbb{N}}$-presque tout $\underline{\omega}$, quand $n \to \infty$ on a

$$\frac{1}{n}\sum_{k=1}^{n} \delta_{f^k_{\underline{\omega}}(x)} \rightharpoonup \mu.$$

6.3. — La démonstration du théorème de Breiman repose sur la généralisation suivante, due à Kolmogorov, de la loi forte des grands nombres.

Proposition. — Dans un espace probabilisé, soit $(\mathcal{F}_n)_{n\geqslant 1}$ une filtration et $(X_n)_{n\geqslant 1}$ une suite adaptée de variables aléatoires (i.e pour tout n, X_n est $\mathcal{F}_n$-mesurable) uniformément bornées dans L^2, et telles que

$$\text{pour tout } n \geqslant 1,\ \mathbb{E}(X_n|\mathcal{F}_{n-1}) = 0.$$

Alors $\displaystyle\frac{1}{n}\sum_{k=1}^{n} X_k$ tend vers 0 p.p. et dans L^2.

Démonstration. — On pose $S_n = \sum_{k=1}^{n} \frac{1}{k} X_k$. L'hypothèse implique que (S_n) est une martingale relativement à la filtration $(\mathcal{F}_n)$. En outre (S_n) est bornée dans L^2 car pour tous $j < k$ on a

$$\mathbb{E}(X_k X_j) = \mathbb{E}(\mathbb{E}(X_k X_j|X_j)) = \mathbb{E}(X_j \mathbb{E}(X_k|X_j)) = 0$$

et donc

$$\mathbb{E}(S_n^2) = \sum_{k=1}^{n} \frac{\mathbb{E}(X_k^2)}{k^2} \leqslant \sup_{k\geqslant 0} \mathbb{E}(X_k^2) \sum_{k=1}^{\infty} \frac{1}{k^2}.$$

Ainsi S_n converge presque sûrement et dans L^2 et on conclut par le lemme élémentaire suivant (pour la convergence dans L^2 on applique le lemme à la suite dans L^2). □

6.4. *Lemme (dit "de Kronecker"). — Soit $(v_n)_{n\geqslant 1}$ une suite réelle telle que la série $\sum_{k\geqslant 1}\frac{v_k}{k}$ converge. Alors*

$$\frac{1}{n}\sum_{k=1}^{n} v_k \underset{n\to\infty}{\longrightarrow} 0.$$

Démonstration du lemme. — On pose $V_n = \sum_{k=1}^{n}\frac{v_k}{k}$, qui converge vers une limite ℓ. On écrit alors

$$\frac{1}{N}\sum_{n=1}^{N} V_n = \frac{1}{N}\sum_{n=1}^{N}\sum_{k=1}^{n}\frac{v_k}{k} = \frac{1}{N}\sum_{k=1}^{N}\frac{N-k+1}{k}v_n = \frac{N+1}{N}V_N - \frac{1}{N}\sum_{k=1}^{N} v_k$$

et comme par le théorème de Cesarò, $\frac{1}{N}\sum_{n=1}^{N} V_n$ tend également vers ℓ, le résultat suit. □

Démonstration du Théorème 6.1. — Le point $x \in X$ étant fixé, posons $\mu_n = \frac{1}{n}\sum_{k=1}^{n}\delta_{f^k_{\underline{\omega}}(x)}$. Il s'agit de montrer que pour presque tout $\underline{\omega}$, $P_\bullet\mu_n - \mu_n \rightharpoonup 0$. En considérant une suite dense de $\mathcal{C}(X)$, cela revient donc à montrer que pour une fonction continue φ sur X fixée, $\langle P_\bullet\mu_n - \mu_n, \varphi\rangle$ tend vers 0 avec n pour presque tout $\underline{\omega}$. Pour $n \geqslant 1$ on introduit la fonction

$$\varphi_n : \underline{\omega} \longmapsto \varphi(f^n_{\underline{\omega}}(x)) - P^\bullet\varphi(f^{n-1}_{\underline{\omega}}(x)).$$

Soit $\mathcal{F}_n$ la tribu engendrée par $\omega_0, \ldots, \omega_{n-1}$. Par indépendance, prendre l'espérance conditionnelle par rapport à $\mathcal{F}_{n-1}$ revient à intégrer par rapport aux variables ω_{n-1}, ω_n, etc. En écrivant $f^n_{\underline{\omega}}(x) = f_{\omega_{n-1}}\circ\cdots\circ f_{\omega_0}(x)$, on a donc

$$\begin{aligned}\mathbb{E}(\varphi\left(f^n_{\underline{\omega}}(x)\right)|\mathcal{F}_{n-1}) &= \int\varphi\left(f^n_{\underline{\omega}}(x)\right)\mathrm{d}\nu(\omega_{n-1}) = \int\varphi\left(f_{\omega_{n-1}}(y)\right)\mathrm{d}\nu(\omega_{n-1})\\ &= P^\bullet\varphi(y) = P^\bullet\varphi(f^{n-1}_{\underline{\omega}}(x)) \text{ où on a posé } y = f^{n-1}_{\underline{\omega}}(x)\end{aligned}$$

et donc $\mathbb{E}(\varphi_n|\mathcal{F}_{n-1}) = 0$. [11] Nous sommes donc dans les conditions d'application de la Proposition 6.3, qui assure que $\frac{1}{n}\sum_{k=1}^{n}\varphi_k$ tend vers 0 p.s. Ainsi

$$\begin{aligned}\langle\mu_n - P_\bullet\mu_n, \varphi\rangle &= \frac{1}{n}\sum_{k=1}^{n}\varphi(f^k_{\underline{\omega}}(x)) - \frac{1}{n}\sum_{k=1}^{n}P^\bullet\varphi(f^k_{\underline{\omega}}(x))\\ &= \frac{1}{n}\sum_{k=1}^{n}\left(\varphi(f^k_{\underline{\omega}}(x)) - P^\bullet\varphi(f^{k-1}_{\underline{\omega}}(x))\right) + O\left(\frac{1}{n}\right)\\ &= \frac{1}{n}\sum_{k=1}^{n}\varphi_k(\underline{\omega}) + O\left(\frac{1}{n}\right) \underset{n\to\infty}{\longrightarrow} 0,\end{aligned}$$

où on a fait un changement d'indice à la deuxième ligne. Le théorème est démontré. □

◇

[11] Cet argument est une variation sur le fait bien connu que si φ est harmonique, alors $(\varphi(f^n_{\underline{\omega}}(x)))_{n\geqslant 0}$ est une martingale.

XI. DYNAMIQUE ALÉATOIRE SUR LE CERCLE

Dans ce chapitre nous allons introduire les exposants de Lyapunov en dynamique aléatoire différentiable et en déduire des techniques pour l'analyse des mesures stationnaires, en se concentrant sur le cas des actions par difféomorphismes sur le cercle. Nous allons ainsi partiellement étendre l'analyse des mesures stationnaires par l'action projective d'un produit aléatoire de matrices 2×2 sur le cercle (espace projectif de dimension 1), faite au chapitre VII, à des dynamiques beaucoup plus générales.

1. Exposants de Lyapunov

1.1. — On se donne une variété M de dimension d et un système dynamique aléatoire $((f_\omega), \nu)$ sur M, muni d'une mesure de probabilité stationnaire μ. Comme précédemment, nous identifions en général ν et son image par $\omega \mapsto f_\omega$. Nous supposerons que les f_ω sont des difféomorphismes de classe C^1. On reprend les notations du chapitre précédent et on considère alors le système dynamique (non-inversible) associé sur $(\Sigma_+ \times M, F_+)$ comme au chapitre précédent, muni de sa mesure invariante $\nu^{\mathbb{N}} \times \mu$. On peut également considérer un système dynamique induit sur le "fibré tangent" $\Sigma_+ \times TM$, défini par

$$(\underline{\omega}, x, v) \longmapsto (\sigma\underline{\omega}, f_{\omega_0}(x), \mathrm{d}(f_{\omega_0})_x(v)),$$

où l'on a noté (x, v) l'élément courant de TM (cf. le §2.3 du chapitre VI), que nous désignerons simplement sous le nom de "cocycle tangent". On fixe une métrique Riemannienne sur M qui permet de donner un sens à $\|Df_x\|$.

Comme expliqué au §VI.2.3, on peut simplifier cette présentation en "faisant des coupures" c'est à dire en retirant à M une hypersurface H de μ-mesure nulle, telle que $M \backslash H$ soit contractible et le fibré tangent restreint à $M \backslash H$ devient alors trivial: $TM|_{M\backslash H} \simeq (M \backslash H) \times \mathbb{R}^d$. On est alors en présence d'un $\mathrm{GL_d}(\mathbb{R})$-cocycle au sens habituel.

1.2. — Si la condition d'intégrabilité

$$\int_{\Omega \times M} \left(\log^+ \|Df_x\| + \log^+ \left\|Df_x^{-1}\right\|\right) \mathrm{d}\nu(f)\mathrm{d}\mu(x) < \infty \tag{1}$$

est satisfaite, alors en appliquant le théorème d'Oseledets on peut définir pour presque tout $(\underline{\omega}, x)$ les exposants de Lyapunov $\lambda_1(\underline{\omega}, x) > \cdots > \lambda_k(\underline{\omega}, x)$ associés au drapeau $V_1(\underline{,}x) > \cdots > V_k(\underline{\omega}, x)$ de $\{\underline{\omega}\} \times T_xM$ et tels que pour $v \in V_i \backslash V_{i-1}$ on a $\frac{1}{n} \log Df^n_{\underline{\omega}}(x) \to \lambda_i$ quand $n \to \infty$. Ce sont par définition les exposants de Lyapunov du système dynamique aléatoire $((f_\omega), \nu, \mu)$. Si M est compacte, la condition (1) est impliquée par la condition plus simple

$$\int_{\Omega} \left(\log^+ \|f\|_{C^1} + \log^+ \left\|f^{-1}\right\|_{C^1}\right) \mathrm{d}\nu(f) < \infty. \tag{2}$$

Dans toute la suite de ce chapitre nous supposerons que M est compacte et, pour simplifier l'exposition, que ν est une mesure à support fini, de sorte que la condition (1) est automatiquement satisfaite.

Exercice. — Vérifier que la condition (1) implique que les hypothèses du théorème d'Oseledets sont effectivement satisfaites pour le cocycle tangent.

1.3. — Dans ce chapitre notre objet d'étude principal est le cas où $M = \mathbb{S}^1 \simeq \mathbb{R}/\mathbb{Z}$ est le cercle. Son fibré tangent est trivial (pourquoi?) et on travaillera donc avec la dérivée $f'(x)$ plutôt qu'avec la différentielle. Dans ce cas, on peut présenter ce qui précède de manière beaucoup plus élémentaire. L'exposant de Lyapunov est

$$\chi(\underline{\omega}, x) = \lim_{n\to\infty} \frac{1}{n} \log \left|(f^n_{\underline{\omega}})'(x)\right| = \lim_{n\to\infty} \frac{1}{n} \log \left|(f_{\omega_{n-1}} \circ \cdots f_{\omega_0})'(x)\right|,$$

dont l'existence est assurée par le théorème de Birkhoff appliqué à la fonction $(\underline{\omega}, x) \mapsto \log |f'_\omega(x)|$ sur $\Sigma_+ \times X$: en effet

$$\frac{1}{n} \log \left|(f_{\omega_{n-1}} \circ \cdots f_{\omega_0})'(x)\right| = \frac{1}{n} \sum_{k=0}^{n-1} \left|f'_{\omega_k}(f_{\omega_{k-1}} \circ \cdots f_{\omega_0}(x))\right|.$$

Si μ est ergodique celui-ci est presque sûrement égal à une constante qui sera notée $\chi(\mu)$.

1.4. — Dans ce contexte (μ ergodique) on a une formule pour $\chi(\mu)$ identique à celle rencontrée au chapitre VII (formule de Furstenberg)

$$\chi(\mu) = \int \log |(f_\omega)'(x)| \, \mathrm{d}\nu(\omega) \mathrm{d}\mu(x). \tag{3}$$

Exercice. — Démontrer cette formule.

2. Propriété de contraction

Considérons donc un système dynamique aléatoire engendré par une mesure ν sur le groupe $\mathrm{Diff}^{1+\alpha}(\mathbb{S}^1)$ des difféomorphismes de classe $C^{1+\alpha}$ du cercle, $\alpha > 0$, muni d'une probabilité stationnaire μ. Par souci de simplicité nous supposerons toujours ν **à support fini** (en particulier la condition d'intégrabilité (1) est satisfaite). Nous noterons Γ_ν le **semigroupe** engendré par $\mathrm{Supp}(\nu)$ (et si besoin nous noterons $\langle \mathrm{Supp}(\nu) \rangle$ le groupe engendré).

Notre premier résultat est une version du théorème de la variété stable de Pesin.

2.1. *Proposition.* — *Sous les hypothèses précédentes, si* $(\underline{\omega}, x)$ *est tel que* $\chi(\underline{\omega}, x) < 0$, *alors il existe un intervalle* $I \subset \mathbb{S}^1$ *centré en* x *tel que pour tout* $y \in I$, *la distance* $\mathrm{d}(f^n_{\underline{\omega}}(y), f^n_{\underline{\omega}}(x))$ *tend vers 0, et plus précisément:*

$$\lim_{n\to\infty} \frac{1}{n} \log \mathrm{d}(f^n_{\underline{\omega}}(y), f^n_{\underline{\omega}}(x)) = \chi(\underline{\omega}, x).$$

Démonstration. — Une adaptation directe de la Proposition 5.9 du chapitre IV (laissée en exercice à la lectrice) montre qu'il existe un intervalle I tel que pour tout $y \in I$,

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathrm{d}(f^n_{\underline{\omega}}(y), f^n_{\underline{\omega}}(x)) \leqslant \chi(\underline{\omega}, x) + \varepsilon.$$

Pour montrer que cette limite est égale à $\chi(\underline{\omega}, x)$, on remarque déjà que l'on a également une estimation pour la distortion, c'est à dire qu'il existe une constante $D > 0$ tel que pour tout $n \geqslant 0$ et pour tout $y \in I$

$$\frac{1}{D} \leqslant \frac{\left|(f^n_{\underline{\omega}})'(y)\right|}{\left|(f^n_{\underline{\omega}})'(x)\right|} \leqslant D.$$

En effet, l'inégalité de droite est contenue dans la preuve de la Proposition IV.5.9 et pour l'inégalité de gauche, une fois que l'on sait que les orbites $f^n_{\underline{\omega}}(x)$ et $f^n_{\underline{\omega}}(y)$ sont exponentiellement asymptotiques, on peut inverser les rôles de x et y. On applique finalement le théorème des accroissements finis: pour $y \in I$,

$$\mathrm{d}(f^n_{\underline{\omega}}(y), f^n_{\underline{\omega}}(x)) = \left|(f^n_{\underline{\omega}})'(c)\right| \mathrm{d}(x, y) \geqslant \frac{1}{D} \left|(f^n_{\underline{\omega}})'(x)\right| \mathrm{d}(x, y)$$

et le résultat suit. □

2.2. — Nous dirons qu'un intervalle $I \subset \mathbb{S}^1$ est **contractable** si pour $\underline{\omega}$ dans un ensemble Ω' de mesure positive, le diamètre de $f^n_{\underline{\omega}}(I)$ tend vers 0 quand $n \to \infty$. Un point $x \in \mathbb{S}^1$ est dit contractable s'il admet un voisinage contractable. L'ensemble $\mathcal{C}$ des points contractables est par définition ouvert. Remarquer que si $f = f_{\omega_0} \in \mathrm{Supp}(\nu)$ et I est contractable, alors $f^{-1}_{\omega_0}(I)$ l'est aussi: en effet les concaténations du type $\omega_0\underline{\omega}$, avec $\underline{\omega} \in \Omega'$ contractent $f^{-1}_{\omega_0}(I)$ et forment un ensemble de mesure positive. Ainsi $\mathcal{C}$ est un ouvert $\mathrm{Supp}(\nu)^{-1}$ invariant, donc Γ^{-1}-invariant, et son complémentaire, le fermé $\mathcal{NC}$ des points non-contractables est donc Γ-invariant.

La proposition suivante découle directement de la Proposition 2.1 et du théorème de Fubini.

Proposition. — Si μ est une probabilité stationnaire ergodique d'exposant de Lyapunov $\chi(\mu) < 0$, alors $\mu(\mathcal{C}) = 1$.

2.3. — Une observation qui a déjà joué un rôle important au chapitre VII est que la propriété de contraction ne fait intervenir aucune mesure stationnaire, même si on utilise l'exposant de Lyapunov d'une mesure stationnaire pour la vérifier. Cela va nous permettre de démontrer des propriétés d'unicité, grâce au lemme suivant.

Lemme. — Si I est un intervalle contractable, il existe au plus une mesure stationnaire μ telle que $\mu(I) > 0$.

Démonstration. — Par décomposition ergodique il suffit de montrer le résultat pour μ ergodique. Supposons que μ_1 et μ_2 sont des mesures stationnaires ergodiques telles que $\mu_1(I) > 0$ et $\mu_2(I) > 0$, et soit $\Omega' \subset \Omega$ tel que pour $\underline{\omega} \in \Omega'$ on a $\mathrm{diam}(f^n_{\underline{\omega}}(I)) \to 0$. Fixons une fonction continue φ sur le cercle. Par le théorème de Birkhoff (voir le Corollaire X.5.7) pour $i = 1, 2$, pour $(\nu^{\mathbb{N}} \times \mu_i)$-presque tout $(\underline{\omega}, x)$ on a

$$\frac{1}{n} \sum_{k=0}^{n-1} \varphi(f_{\underline{\omega}}^k(x)) \underset{n\to\infty}{\longrightarrow} \int \varphi \, \mathrm{d}\mu_i.$$

Par Fubini, pour μ_i-presque tout x, cette propriété vaut pour $\nu^{\mathbb{N}}$-presque tout ω. En particulier on peut choisir $\underline{\omega} \in \Omega'$, et $(x_1, x_2) \in I^2$ tels que

$$\frac{1}{n} \sum_{k=0}^{n-1} \varphi(f_{\underline{\omega}}^k(x_1)) \underset{n\to\infty}{\longrightarrow} \int \varphi \, \mathrm{d}\mu_1 \text{ et } \quad \frac{1}{n} \sum_{k=0}^{n-1} \varphi(f_{\underline{\omega}}^k(x_2)) \underset{n\to\infty}{\longrightarrow} \int \varphi \, \mathrm{d}\mu_2.$$

Mais comme φ est continue et $\mathrm{d}(f_{\underline{\omega}}^n(x_1), f_{\underline{\omega}}^n(x_2)$ tend vers 0, il vient $\int \varphi \, \mathrm{d}\mu_1 = \int \varphi \, \mathrm{d}\mu_2$, et comme φ est arbitraire, $\mu_1 = \mu_2$. □

2.4. — On déduit des considérations précédentes un premier résultat d'unicité non trivial, qui sera généralisé un peu plus loin.

Corollaire. — Soit ν *une mesure à support fini sur* $\mathrm{Diff}^{1+\alpha}(\mathbb{S}^1)$. *On suppose que l'action du semigroupe* Γ *engendré par* $\mathrm{Supp}(\nu)$ *sur* $\mathbb{S}^1$ *est minimale (i.e. toutes les orbites sont denses) et qu'il existe une mesure* ν*-stationnaire ergodique d'exposant strictement négatif. Alors il existe une unique mesure* ν*-stationnaire.*

Démonstration. — Il suffit de montrer qu'il existe une unique mesure stationnaire ergodique. Comme Γ agit de façon minimale, le fermé invariant $\mathcal{NC}$ est soit vide, soit égal à $\mathbb{S}^1$. L'existence d'une mesure d'exposant < 0 garantit que $\mathcal{C}$ est non vide, donc $\mathcal{NC} = \varnothing$ et $\mathcal{C} = \mathbb{S}^1$. Par ailleurs, si μ est une mesure stationnaire la relation $\mu = \int (f_\omega)_* \mu \, \mathrm{d}\nu(\omega)$ implique que pour tout $f \in \mathrm{Supp}(\nu)$, $f(\mathrm{Supp}(\mu)) \subset \mathrm{Supp}(\mu)$, ainsi $\mathrm{Supp}(\mu)$ est Γ-invariant et par minimalité, $\mathrm{Supp}(\mu) = \mathbb{S}^1$. Finalement, si l'on considère deux mesures stationnaires ergodiques μ_1 et μ_2, et $x \in \mathbb{S}^1$ quelconque, il existe un intervalle ouvert $I \ni x$ contractable, et le Lemme 2.3 implique que $\mu_1 = \mu_2$, ce qui termine la preuve. □

3. Principe d'invariance

Nous conservons dans cette section les hypothèses de la section précédente. Le "principe d'invariance" affirme que pour un système dynamique aléatoire sur le cercle sans mesure invariante, on doit avoir contraction. Plus précisément:

3.1. *Théorème. — Soit* ν *est une mesure à support fini sur* $\mathrm{Diff}^{1+\alpha}(\mathbb{S}^1)$. *Alors si* μ *est une mesure stationnaire ergodique on a :*

1. *soit* μ *est* Γ_ν*-invariante (et donc* $\langle \mathrm{Supp}(\nu) \rangle$*-invariante);*
2. *soit* $\chi(\mu) < 0$.

En particulier s'il n'y a pas de mesure invariante, toute mesure stationnaire ergodique a un exposant de Lyapunov strictement négatif.

Ce théorème est une synthèse de travaux de nombreux mathématiciens parmi lesquels on peut citer Furstenberg, Ledrappier, Baxendale, Crauel, etc. La présentation est empruntée à Malicet[1]. Cette présentation dans le cadre des difféomorphismes $C^{1+\alpha}$ vise à faire le lien avec les chapitres précédents, mais il sera clair à la lecture de la preuve que le résultat vaut dans le cas C^1, et même pour des homéomorphismes, en remplaçant l'exposant de Lyapunov (infinitésimal) par un exposant de contraction (local).

Les deux alternatives du théorème ne sont pas mutuellement exclusives: leur intersection peut se produire pour une mesure finie Γ-invariante. En effet si μ est une mesure stationnaire ergodique d'exposant $\chi(\mu) < 0$, alors pour presque tout $\underline{\omega}$, la mesure $\mu_{\underline{\omega}}$ définie au Lemme lem:martingale du chapitre précédent est une somme finie de masses de Dirac (exercice![2]). Si en outre μ est invariante, alors pour tout ω, on a $\mu_{\underline{\omega}} = \mu$ et on conclut.

3.2. — Par un argument de décomposition ergodique, on voit plus généralement que si μ est une mesure stationnaire quelconque, on a que soit μ est invariante, soit

$$\int \chi(\underline{\omega}, x) \mathrm{d}\nu^{\mathbb{N}}(\underline{\omega}) \mathrm{d}\mu(x) = \int \log |(f_\omega)'(x)| \, \mathrm{d}\nu(\omega) \mathrm{d}\mu(x) < 0.$$

3.3. — Comme on l'a déjà expliqué, l'existence d'une mesure Γ_ν-invariante est un phénomène "spécial", donc il faut considérer la deuxième alternative du théorème comme le cas typique. Il est souvent assez simple d'écarter l'existence d'une telle mesure invariante. Supposons par exemple qu'il existe $\gamma \in \langle \operatorname{Supp} \nu \rangle$ dont la dynamique est de type "nord-sud", c'est à dire qu'il existe des points N et S sur $\mathbb{S}^1$ tels que pour tout $x \in \mathbb{S}^1 \backslash N$ on a $\gamma^n(x) \to S$. Alors toute mesure ν-invariante appartient est une combinaison de δ_N et δ_S. Donc s'il existe $\gamma' \in \langle \operatorname{Supp} \nu \rangle$ ne préservant ni N, ni S, ni $\{N, S\}$, il n'existe pas de mesure ν-invariante.

3.4. — Voici une autre exemple d'obstruction à l'existence d'une mesure invariante:

Proposition. — Soit Γ *un groupe d'homéomorphismes croissants du cercle, dont toutes les orbites sont denses. S'il existe une mesure* Γ*-invariante, alors* Γ *est conjugué à groupe de rotations. En particulier* Γ *est abélien.*

Démonstration. — Soit μ une mesure de probabilité invariante. Comme toute orbite de Γ est infinie, μ n'a pas d'atomes, et puisque celles-ci sont denses, $\operatorname{Supp}(\mu) = \mathbb{S}^1$. Alors, comme μ est diffuse, il existe un homéomorphisme h du cercle tel que $h_\star \mathrm{Leb} = \mu$: il suffit en effet de prendre g la fonction de répartition de μ, définie par $\mu([0, x]) = g(x) = \mathrm{Leb}([0, g(x)])$, qui est un homéomorphisme croissant du cercle, et on pose ensuite $h := g^{-1}$. On vérifie alors directement que $h\Gamma h^{-1}$ est un groupe d'homéomorphismes croissants et qui préservent la mesure de Lebesgue, c'est à dire des rotations. □

Remarque. — Pour un groupe d'homéomophismes quelconques agissant de manière minimale, ce résultat s'applique au sous-groupe d'indice 2 des éléments de Γ préservant l'orientation.

---

[1] *Random Walks on* $\mathrm{Homeo}(S^1)$. Comm. in Math. Phys. 356 (2017), 1083–1116

[2] Pour la solution, voir Le Jan, Annales de l'IHP B 23 (1987), 111–120, Proposition 2.

3.5. — La démonstration du principe d'invariance repose sur l'introduction d'une entropie $h(\mu, \nu)$ qui vérifie les propriétés suivantes:

1. $h(\mu, \nu) \geqslant 0$ avec égalité si et seulement si μ est invariante;
2. $h(\mu, \nu) \leqslant \max(-\chi(\mu), 0)$ (analogue de l'inégalité de Ruelle).

Le théorème se déduit immédiatement de ces deux propriétés. Voir la section 5 pour les détails.

4. Conséquences dynamiques

4.1. — Le Théorème 3.1, utilisé en lieu et place de la section 3 du chapitre VII fournit une nouvelle approche pour la positivité de l'exposant de Lyapunov pour les produits aléatoires de matrices 2×2 (il s'agit en fait de l'approche originale de Furstenberg, qui a été abstraite et systématisée ensuite par Ledrappier).

4.2. — En combinant le Corollaire 2.4 et le principe d'invariance, on obtient:

Corollaire. — Soit ν une mesure à support fini sur $\mathrm{Diff}^{1+\alpha}(\mathbb{S}^1)$. *On suppose que l'action du semigroupe Γ engendré par* $\mathrm{Supp}(\nu)$ *sur $\mathbb{S}^1$ est minimale et qu'il n'existe pas de mesure Γ-invariante. Alors il existe une unique mesure ν-stationnaire (qui est contractante).*

Par le théorème ergodique de Breiman, on voit ainsi que pour un point arbitraire x du cercle, pour presque toute trajectoire de la marche aléatoire engendrée par ν sur Γ, la suite des mesures empiriques aux points de l'orbite $f_{\omega_n} \circ \cdots \circ f_{\omega_0}(x)$ s'équidistribue vers l'unique mesure stationnaire μ: on a ainsi pour ce système résolu le problème fondamental de la théorie ergodique!

4.3. — Afin de discuter l'hypothèse de minimalité dans l'énoncé précédent, on peut s'appuyer sur l'alternative suivante pour la dynamique topologique d'un groupe d'homéomorphismes du cercle:

Proposition. — Si Γ est un sous-groupe du groupe $\mathrm{Homeo}(\mathbb{S}^1)$ *des homéomorphismes du cercle, l'une (exactement) des propriétés suivantes a lieu:*

(i) Γ *a une orbite finie;*
(ii) toutes les orbites de Γ *sont denses;*
(iii) il existe un unique ensemble invariant minimal $\Lambda \subset \mathbb{S}^1$, homéomorphe à un ensemble de Cantor, et sur lequel toutes les orbites s'accumulent.

Démonstration. — L'ensemble des fermés Γ-invariants non-vides est ordonné par l'inclusion, et par compacité, toute intersection décroissante de fermés invariants non-vide admet un élément minimal. Le lemme de Zorn fournit donc un fermé invariant minimal Λ. La frontière $\partial\Lambda$ et l'ensemble dérivé Λ' (i.e. l'ensemble des points d'accumulation) sont alors également invariants, et contenus dans Λ par hypothèse. On tombe alors dans une des situations suivantes:

– soit $\Lambda' = \varnothing$ et donc Λ est fini;
– soit $\partial\Lambda = \varnothing$ et donc $\Lambda = \mathbb{S}^1$;

– soit $\partial\Lambda = \Lambda' = \Lambda$ et donc Λ est un compact parfait d'intérieur vide, et donc c'est un ensemble de Cantor.

Il reste à démontrer que dans le troisième cas Λ est l'unique ensemble minimal et que toutes les autres orbites accumulent Λ. Remarquer que la deuxième de ces assertions implique la première. Soit donc x un point arbitraire du cercle et montrons que $\overline{\Gamma(x)} \supset \Lambda$. Si $x \in \Lambda$, par minimalité $\overline{\Gamma(x)} = \Lambda$ et il n'y a rien à démontrer. Si $x \notin \Lambda$, on note I la composante connexe (intervalle) de $\Lambda^{\complement}$ contenant x, et on fixe un point arbitraire $y \in \Lambda$. Soit $a \in \Lambda$ une extrémité de I. Comme Λ est minimal, il y a une suite $(\gamma_n) \in \Gamma^{\mathbb{N}}$ telle que $\gamma_n(a)$ converge vers y. Comme en outre Λ est parfait, on peut supposer que cette suite est injective. Pour tout n, $\gamma_n(a)$ est une extrémité de $\gamma_n(I)$, et il y a au plus une autre valeur n' telle que $\gamma_{n'}(a)$ est une extrémité de $\gamma_n(I)$. On peut donc extraire une sous-suite (n_j) telle que les intervalles $\gamma_{n_j}(I)$ sont distincts, et en particulier leur longueur tend vers 0. Ainsi $\mathrm{d}(\gamma_{n_j}(x), \gamma_{n_j}(a)) \to 0$, et $(\gamma_{n_j}(x))$ converge vers y, ce qu'il fallait démontrer. □

Dans le cas *(iii)*, on désigne souvent Λ sous le nom de "**minimal exceptionnel**". Néanmoins l'existence d'un tel Λ n'a rien d'exceptionnel! Il apparaît par exemple si Γ est engendré par l'action projective de deux transformations A et B de $\mathrm{GL}_2(\mathbb{R})$ de type

$$A = \begin{pmatrix} \lambda & 0 \\ 0 & \lambda^{-1} \end{pmatrix}, \text{ avec } \lambda \gg 1, \text{ et } B = PAP^{-1},$$

où P est une rotation non-triviale (c'est un exemple de **groupe de Schottky**). Paradoxalement, le cas le plus difficile à analyser dans la taxonomie de la proposition précédente est le cas où Γ a une orbite finie.

4.4. — En combinant la Proposition 4.3 au Théorème 3.1, on obtient ainsi:

Corollaire. — Sous les hypothèses du Théorème 3.1, supposons en outre que Γ_ν est un sous-groupe (par exemple si Supp(ν) *est symétrique), et que Γ_ν n'admet aucune mesure invariante. Alors il existe une unique probabilité ν-stationnaire (qui est contractante).*

Démonstration. — Il suffit de démontrer qu'il y a une unique probabilité stationnaire ergodique. Soit donc μ une telle mesure. S'il n'y a pas de mesure invariante, le premier cas de la Proposition 4.3 est exclu. D'après le Théorème 3.1, $\chi(\mu) < 0$, et d'après le Lemme 2.3, il suffit de montrer que toutes les mesures stationnaires ergodiques ont le même support. Le cas où Γ agit de façon minimale sur $\mathbb{S}^1$ a été traité au §2.4. Dans le cas où il y a un minimal exceptionnel, si μ est une mesure stationnaire ergodique, on remarque d'abord que comme Λ est l'unique ensemble minimal, il doit être contenu dans Supp(μ). Pour établir que Supp$(\mu) = \Lambda$, il suffit de montrer que pour toute composante connexe I de $\mathbb{S}^1 \backslash \Lambda$ on a $\mu(I) = 0$. Pour cela on suppose par l'absurde que $\mu(\Lambda^{\complement}) > 0$ et on imite l'argument de la Proposition 5.5 du chapitre précédent: on note $\mu_{\max} > 0$ la mesure maximale d'une composante connexe $\mathbb{S}^1 \backslash \Lambda$, et on prend I une composante telle que $\mu(I) = \mu_{\max}$. Par stationnarité, on obtient que l'orbite de I est finie, ce qui contredit par exemple le fait que les extrémités de I ont une une orbite dense dans Λ. □

5. Démonstration du principe d'invariance

On se place donc sous les hypothèses du Théorème 3.1, c'est à dire que ν est une mesure à support fini sur $\mathbb{S}^1$, et on fixe une mesure ν-stationnaire ergodique μ. Si μ a un atome, d'après la Proposition 5.5 du chapitre précédent, elle est invariante, et il n'y a rien à démontrer. Nous supposerons donc que μ est diffuse. Dans toute la démonstration, la notation I désigne un intervalle du cercle.

5.1. — La relation de stationnarité s'écrit $\mu = \int (f_\omega)_* \mu \,\mathrm{d}\nu(\omega)$. Considérons la famille (finie) de mesures de probabilité $(f_\omega)^* \mu = (f_\omega^{-1})_* \mu$. Remarquer que si $I \subset \mathbb{S}^1$ est un intervalle, $(f_\omega)^* \mu(I) = \mu(f_\omega(I))$: cette quantité mesure donc comment I est contracté (ou dilaté) par f_ω. Considérons la décomposition de Radon-Nikodym

$$(f_\omega^{-1})_* \mu = J(\omega, x)\mu + \tilde{\mu}_\omega, \text{ où } J(\omega, x) = \frac{\mathrm{d}(f_\omega^{-1})_* \mu}{\mathrm{d}\mu}(x)$$

et $\tilde{\mu}_\omega$ est la composante singulière. Remarquer que si ν (ou plus généralement $\mathrm{Supp}(\nu)$ est symétrique on a $(f_\omega^{-1})_* \mu \ll \mu$ donc $\tilde{\mu}_\omega = 0$.

On introduit une entropie, semble-t-il inspirée par la “divergence de Kullback-Leibler”:

$$h = h(\mu, \nu) = \int -\log J(\omega, x) \mathrm{d}\nu(\omega) \mathrm{d}\mu(x) \left(= \int -\log J(\omega_0, x) \mathrm{d}\nu^{\mathbb{N}}(\underline{\omega}) \mathrm{d}\mu(x) \right).$$

Dans cette définition on sous-entend que si $-\log J$ n'est pas intégrable on pose $h = +\infty$.

5.2. *Proposition.* —

1. *L'entropie h est bien définie et positive (éventuellement infinie).*
2. *$h = 0$ si et seulement si μ est invariante.*

Démonstration. — La définition de l'entropie se lit plus précisément

$$h = \int_{\{J>0\}} -\log J(\omega, x) \mathrm{d}\nu(\omega) \mathrm{d}\mu(x) + \infty (\nu \times \mu)(\{J = 0\}).$$

Remarquons d'abord que

$$\int J(\omega, x) \mathrm{d}\nu(\omega) \mathrm{d}\mu(x) \leqslant \int (f_\omega^{-1})_* \mu \,\mathrm{d}\nu(\omega) \leqslant 1.$$

Comme la fonction $-\log$ est convexe et décroissante, par l'inégalité de Jensen on en déduit que

$$0 = -\log 1 \leqslant -\log \left(\int_{\{J>0\}} J(\omega, x) \mathrm{d}\nu(\omega) \mathrm{d}\mu(x) \right) \leqslant \int_{\{J>0\}} -\log J(\omega, x) \mathrm{d}\nu(\omega) \mathrm{d}\mu(x),$$

ce qui montre que h est bien définie et $h \geqslant 0$. En outre, si $h = 0$ alors $J > 0$ p.p. et comme $-\log$ est strictement convexe, le cas d'égalité de l'inégalité de Jensen impose que $J(\omega, x)$ est p.p. égale à une constante, et donc nécessairement $J(\omega, x) = 1$ p.p. Ainsi $(f_\omega^{-1})_* \mu = \mu$ pour tout ω, et donc $(f_\omega)_* \mu = \mu$. $\square$

5.3. — Pour la suite nous aurons besoin d'une version générale du théorème de densité de Lebesgue, et de l'inégalité maximale associée.

Proposition. — *Soient m_1 et m_2 deux mesures de probabilité sur le cercle. Alors:*

1. *pour m_2-presque tout $x \in \mathbb{S}^1$ on a* $\dfrac{dm_1}{dm_2}(x) = \lim\limits_{I \ni x, \operatorname{diam}(I) \to 0} \dfrac{m_1(I)}{m_2(I)}$.
2. *Si on pose* $M^*(x) = \sup\limits_{I \ni x} \dfrac{m_1(I)}{m_2(I)}$, *on a*

$$\int \log^+ M^*(x) dm_2(x) \leqslant C_1, \tag{4}$$

où C_1 est une constante universelle ne dépendant que de la dimension (en fait $C_1 = 2$ convient).

L'ingrédient principal de la preuve est le théorème de recouvrement de Besicovich, voir par exemple le chapitre 2 de [**9**] pour les détails.

5.4. — Posons maintenant pour $\varepsilon > 0$

$$J_\varepsilon(\omega, x) = \sup \left\{ \frac{\mu(f_\omega(I))}{\mu(I)}, \ I \ni x, \ \mu(I) \leqslant \varepsilon \right\}$$

et

$$h_\varepsilon = \int -\log J_\varepsilon(\omega, x) \mathrm{d}\nu(\omega) \mathrm{d}\mu(x).$$

Lemme technique. — *Quand $\varepsilon \to 0$ on a $J_\varepsilon \to J$ p.p. et $h_\varepsilon \to h$.*

Démonstration. — La première affirmation découle directement du premier point de la proposition précédente. Pour la deuxième, il faut intervertir limite et intégrale. Pour cela, écrivons

$$h_\varepsilon = -\int \log^+ J_\varepsilon(\omega, x) \mathrm{d}\nu(\omega) \mathrm{d}\mu(x) + \int \log^- J_\varepsilon(\omega, x) \mathrm{d}\nu(\omega) \mathrm{d}\mu(x).$$

L'inégalité maximale (4) implique que la première intégrale vérifie les hypothèses du théorème de convergence dominée. Pour la deuxième, on remarque que J_ε décroît quand ε décroît vers 0, donc $\log^- J_\varepsilon$ est positif et croissant, ainsi les hypothèses du théorème de convergence monotone sont satisfaites. □

5.5. — Pour conclure la démonstration du théorème, supposons donc que $h > 0$ et montrons que $\chi(\mu) < 0$ (la preuve montre plus précisément que $\chi(\mu) \leqslant -h$, en particulier l'entropie est finie). Posons $\varepsilon > 0$ assez petit pour que $h_\varepsilon > 0$ et fixons $0 < h' < h_\varepsilon$. Par le théorème de Birkhoff appliqué au système dynamique produit sur $\Sigma \times \mathbb{S}^1$ et à la fonction $(\underline{\omega}, x) \mapsto -\log J_\varepsilon(\omega_0, x)$ on a pour $\nu^{\mathbb{N}} \times \mu$ presque tout $(\underline{\omega}, x)$ la convergence

$$\frac{1}{n} \sum_{k=0}^{n-1} (-\log J_\varepsilon)(\omega_k, f_{\underline{\omega}}^k(x)) \xrightarrow[n \to \infty]{} h_\varepsilon.$$

Si $(\underline{\omega}, x)$ vérifie cette convergence, il existe donc une constante C_0 telle que

$$\prod_{k=0}^{n-1} J_\varepsilon(\omega_k, f^k_{\underline{\omega}}(x)) \leqslant C_0 e^{-nh'}.$$

Sans perte de généralité on peut supposer $C_0 > 1$.

Affirmation. — Si I est un intervalle assez petit centré en x, on a $\mu(f^n_{\underline{\omega}}(I)) \leqslant C_0\mu(I)e^{-nh'}$.

En effet, supposons que $\mu(I) < \varepsilon/C_0$ et montrons cette propriété par récurrence sur n. Si $n = 0$ c'est évident. Si maintenant la propriété a été démontrée pour $k = 0, \ldots, n-1$, alors pour tout $0 \leqslant k \leqslant n-1$ on a $\mu(f^k_{\underline{\omega}}(I)) \leqslant C_0 e^{-kh'}\mu(I) \leqslant \varepsilon$, et par définition de J_ε on a donc

$$\frac{\mu\left(f^{k+1}_{\underline{\omega}}(I)\right)}{\mu\left(f^k_{\underline{\omega}}(I)\right)} = \frac{\mu\left(f_{\omega_k}\left(f^k_{\underline{\omega}}(I)\right)\right)}{\mu\left(f^k_{\underline{\omega}}(I)\right)} \leqslant J_\varepsilon(\omega_k, f^k_{\underline{\omega}}(x)),$$

où dans la dernière égalité on a utilisé le fait que $f^k_{\underline{\omega}}(x) \in f^k_{\underline{\omega}}(I)$. Ainsi

$$\mu(f^n_{\underline{\omega}}(I)) = \mu(I) \prod_{k=0}^{n-1} \frac{\mu\left(f^{k+1}_{\underline{\omega}}(I)\right)}{\mu\left(f^k_{\underline{\omega}}(I)\right)} \leqslant \mu(I) \prod_{k=0}^{n-1} J_\varepsilon(\omega_k, f^k_{\underline{\omega}}(x)) \leqslant C_0\mu(I)e^{-nh'},$$

et notre affirmation est démontrée.

5.6. — À ce stade nous avons démontré que pour $(\nu^N \times \mu)$-presque tout $(\underline{\omega}, x)$, la mesure (selon μ) de $f^n_{\underline{\omega}}(I)$ décroît exponentiellement comme $e^{-nh'}$. Pour conclure il faut estimer la *longueur* de cet intervalle. La clé va être à nouveau l'inégalité maximale de Lebesgue. En effet dans (4) prenons $m_1 = \mathrm{Leb}$ et $m_2 = \mu$. Alors $\int \log^+ M^*(x)\mathrm{d}\mu(x) \leqslant 2$, où $M^*(x) = \sup_{I \ni x} \frac{\mathrm{Leb}(I)}{\mu(I)}$. Ceci implique que la fonction $(\underline{\omega}, x) \mapsto \log^+ M^*(x)$ est $(\nu^{\mathbb{N}} \times \mu)$-intégrable, et donc il découle du théorème de Birkhoff que $\frac{1}{n}\log^+ M^*(f^k_{\underline{\omega}}(x))$ converge p.s. vers 0 (voir p.2.4 pour un argument direct basé sur Borel-Cantelli). Or par définition

$$\mathrm{Leb}(f^n_{\underline{\omega}}(I)) \leqslant M^*(f^k_{\underline{\omega}}(x))\, \mu(f^n_{\underline{\omega}}(I))$$

et on conclut que pour presque tout $\underline{\omega}$, si $I \ni x$ est un intervalle assez petit, on a

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathrm{Leb}(f^n_{\underline{\omega}}(I)) \leqslant -h'.$$

Ceci implique que $\chi(\underline{\omega}, x) \leqslant -h'$ en deux étapes, par l'argument de distortion déjà rencontré plusieurs fois (cf la Proposition 2.1): en effet si $y \in I$, comme x et y sont exponentiellement proches sous itération, on déduit que

$$\frac{1}{D} \leqslant \frac{\left|(f^n_{\underline{\omega}})'(y)\right|}{\left|(f^n_{\underline{\omega}})'(x)\right|} \leqslant D.$$

Ainsi la longueur de $f^n_{\underline{\omega}}(I)$ est comparable à $\left|(f^n_{\underline{\omega}})'(x)\right|$, et le théorème est démontré. $\square$

6. Généralisations

Il est illusoire d'imaginer obtenir une compréhension aussi précise des systèmes dynamiques aléatoires sur des variétés de dimension supérieure. Comme on l'a remarqué au chapitre précédent, même dans le cas des actions par automorphismes linéaires sur les tores, la description des mesures stationnaires est un problème difficile et profond.

En revanche, le principe d'invariance admet une version générale, due à Baxendale et Crauel, et s'inspirant de travaux de Furstenberg et Ledrappier.

6.1. *Théorème. — Soit M une variété compacte et ν une mesure à support fini sur* $\mathrm{Diff}^{1+\alpha}(M)$*, avec $\alpha > 0$. Alors si μ est une mesure stationnaire ergodique on a :*

1. *soit μ est ν-invariante;*
2. *soit μ admet un exposant de Lyapunov strictement négatif.*

Comme pour le Théorème 3.1, la démonstration repose sur l'introduction d'une entropie mesurant la non-invariance de μ, et qui vérifie une inégalité dans l'esprit de l'inégalité de Ruelle.

◇

Bibliographie

[1] Yves Benoist and Jean-François Quint. *Random walks on reductive groups*, volume 62 of *Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge. A Series of Modern Surveys in Mathematics [Results in Mathematics and Related Areas. 3rd Series. A Series of Modern Surveys in Mathematics].* Springer, Cham, 2016.

[2] V. I. Bogachev. *Measure theory. Vol. I, II.* Springer-Verlag, Berlin, 2007.

[3] Philippe Bougerol and Jean Lacroix. *Products of random matrices with applications to Schrödinger operators*, volume 8 of *Progress in Probability and Statistics.* Birkhäuser Boston, Inc., Boston, MA, 1985.

[4] Michael Brin and Garrett Stuck. *Introduction to dynamical systems.* Cambridge University Press, Cambridge, 2002.

[5] Yves Coudène. *Théorie ergodique et systèmes dynamiques.* Savoirs Actuels (Les Ulis). [Current Scholarship (Les Ulis)]. EDP Sciences, Les Ulis; CNRS Éditions, Paris, 2012.

[6] Anatole Katok and Boris Hasselblatt. *Introduction to the modern theory of dynamical systems*, volume 54 of *Encyclopedia of Mathematics and its Applications.* Cambridge University Press, Cambridge, 1995. With a supplementary chapter by Katok and Leonardo Mendoza.

[7] V. A. Kaĭmanovich and A. M. Vershik. Random walks on discrete groups: boundary and entropy. *Ann. Probab.*, 11(3):457–490, 1983.

[8] Yuri Kifer. *Ergodic theory of random transformations*, volume 10 of *Progress in Probability and Statistics.* Birkhäuser Boston, Inc., Boston, MA, 1986.

[9] Pertti Mattila. *Geometry of sets and measures in Euclidean spaces*, volume 44 of *Cambridge Studies in Advanced Mathematics.* Cambridge University Press, Cambridge, 1995. Fractals and rectifiability.

[10] William Parry. *Topics in ergodic theory*, volume 75 of *Cambridge Tracts in Mathematics.* Cambridge University Press, Cambridge-New York, 1981.

[11] Robert R. Phelps. *Lectures on Choquet's theorem*, volume 1757 of *Lecture Notes in Mathematics.* Springer-Verlag, Berlin, second edition, 2001.

[12] Feliks Przytycki and Mariusz Urbański. *Conformal fractals: ergodic theory methods*, volume 371 of *London Mathematical Society Lecture Note Series.* Cambridge University Press, Cambridge, 2010.

[13] Marcelo Viana. *Lectures on Lyapunov exponents*, volume 145 of *Cambridge Studies in Advanced Mathematics.* Cambridge University Press, Cambridge, 2014.

[14] Peter Walters. *An introduction to ergodic theory*, volume 79 of *Graduate Texts in Mathematics.* Springer-Verlag, New York-Berlin, 1982.

April 5, 2023

Sorbonne Universités, Laboratoire de Probabilités, Statistique et Modélisation (LPSM, UMR 8001), 4 place Jussieu, 75005 Paris, France • *E-mail :* `romain.dujardin@upmc.fr`

Romain Dujardin