

Théorie ergodique et systèmes dynamiques

Yves Coudène, 15 juin 2018

Version 2

Introduction	2
Théorème ergodique en moyenne	6
Théorème ergodique presque partout	12
Mélange	18
L'argument de Hopf	24
Dynamique topologique	30
Non-errance	36
Conjugaison	42
Linéarisation	48
Un attracteur étrange	54
Entropie	60
Entropie et théorie de l'information	66
Calculs d'entropie	72
Espaces de Lebesgue et isomorphisme	78
Spectre des systèmes dynamiques	84
Décomposition ergodique	90
Annexes :	
Convergence faible dans un espace de Hilbert	96
Calcul différentiel	98
Topologie et mesure	99
Partitions mesurables et σ -algèbres	103
Références	106

Introduction

*Il semble que la perfection soit atteinte
non quand il n'y a plus rien à ajouter,
mais quand il n'y a plus rien à retrancher.*
A. de Saint-Exupéry (1900-1944)

Génèse du livre

Ces notes sont issues d'un cours de Master que j'ai donné à l'université de Rennes 1 pendant la période 2005-2008. Il s'agissait d'un cours d'introduction à la théorie ergodique et aux systèmes dynamiques ; le but était de présenter quelques idées générales qui sont à la base de ces deux théories, avant que les étudiants ne se spécialisent en suivant des cours spécialisés.

Le cours était composé de douze séances de deux heures chacune ; il m'a paru judicieux de focaliser chaque séance sur un concept particulier, et de faire en sorte que les différentes séances soient largement indépendantes entre elles. De fait, le matériel présenté est à l'intersection de théories mathématiques très diverses, et l'auditoire intéressé par le sujet est souvent composé d'étudiants et de chercheurs d'horizons très différents : probabilistes, dynamiciens, géomètres, physiciens, etc.

Ce livre reflète l'organisation du cours, tel qu'il a été enseigné. Chaque chapitre est conçu de façon à occuper deux heures de temps ; il commence par une présentation informelle des concepts et des problèmes qu'on cherche à résoudre. Viennent ensuite les définitions rigoureuses et les démonstrations, qu'on a cherché à illustrer par des exemples simples et pertinents. Les figures forgent l'intuition du lecteur tandis que les exercices lui permettent de tester sa compréhension du sujet. J'ai rajouté quelques commentaires à la fin de chaque chapitre, afin de remettre le matériel étudié dans son contexte historique, présenter quelques problèmes actuels, et orienter le lecteur dans la littérature, en fonction de ses intérêts propres. Ces commentaires sont plutôt destinés à une seconde lecture et supposent une certaine maîtrise des concepts présentés dans ce livre.

Pour ce qui est du contenu, j'ai voulu insister sur les idées plus que sur la théorie, sur les exemples plus que sur la technique. Il existe plusieurs livres présentant les théories générales avec un grand luxe de détail, aussi bien dans le domaine des systèmes dynamiques que dans celui de la théorie ergodique. Ces notes n'ont pas vocation à les remplacer. En particulier, j'ai donné pour quelques résultats classiques des preuves nouvelles ou inhabituelles, afin d'illustrer certains aspects méconnus du sujet. Ces preuves sont susceptibles d'intéresser même les chercheurs les plus aguerris. Le lecteur est bien sûr invité à consulter les ouvrages de référence pour prendre connaissance des approches plus classiques, qui sont résumées dans les commentaires.

Thèmes abordés

Théorie ergodique et systèmes dynamiques sont deux théories qui vont très bien ensemble. La première apporte à la seconde ses résultats quantitatifs les plus remarquables, tandis que la seconde est une pourvoyeuse infatigable d'exemples prompts à infirmer les conjectures les plus chères à la première. Toutes deux nées au début du vingtième siècle, au moins d'un point de vue mathématique et moderne, sous la houlette d'un des géants du siècle, Henri Poincaré (1854-1912), elles ont connues un développement soutenu jusqu'à aujourd'hui, et les ouvrages qui prétendent rendre compte d'une part non négligeable de ces théories sont, de par leur taille et leur style, prompts à effrayer même les étudiants les plus motivés.

Ce livre s'intéresse aux rapports qu'entretiennent les concepts d'hyperbolicité et de hasard. Il a été écrit dans le but d'être accessible au plus grand nombre,

de susciter l'intérêt pour un domaine très actif des mathématiques, et peut servir d'introduction à la littérature plus avancée. On espère avoir évité l'écueil du discours ennuyeux et technique, et si le lecteur referme ce livre avec le désir d'en savoir plus, alors il aura rempli son rôle.

Les grands problèmes qu'on cherche à comprendre n'ont pas tellement évolué en un siècle. Prenons l'exemple d'une application qui agit sur un certain espace de configurations X . Les points de X représentent les différents états que peut prendre le système au cours de son mouvement. Partant d'une configuration initiale donnée par un point x de X , les itérés de x correspondent aux états successifs que visite le système au cours de son évolution. Ce livre s'intéresse aux questions suivantes :

– *Le système repasse-t-il proche de son état initial au cours de son évolution ?*

C'est ce qu'essaient de formaliser les concepts de récurrence et de non-errance, aussi bien sur le plan quantitatif (mesure) que sur le plan qualitatif (topologie).

– *Est-il possible de construire une représentation du système, dans laquelle l'évolution prend une forme particulièrement simple à décrire ?*

Les notions de conjugaison locale ou globale, d'isomorphisme, de codage et de modèle symbolique, cherchent chacune à leur façon, à mettre le système sous une forme où l'évolution peut être effectivement calculée.

– *Le système peut-il évoluer de manière à se rapprocher d'un état donné à priori, si on le perturbe au cours de son évolution ?*

Ce thème est prépondérant en dynamique hyperbolique, où l'existence d'instabilités locales, modélisées par les variétés stables et instables, conduit à un comportement uniforme du système, stable par perturbation.

– *Dans quelle mesure l'évolution du système peut-elle être prédite à long terme, ou encore, quelle quantité de hasard le système est-il susceptible de simuler ?*

Le concept d'entropie, introduit en 1958 par A. N. Kolmogorov en théorie des systèmes dynamiques, a permis de faire des progrès décisifs sur cette question.

Les quatre premiers chapitres sont consacrés à des résultats de théorie ergodique (récurrence, ergodicité, mélange), illustrés par des exemples de nature algébromo-mécano-probabiliste : flots hamiltoniens, décalages de Bernoulli, automorphismes des tores, flots sur $SL_2(\mathbf{R})$ etc. On a cherché à mettre en valeur le rôle joué par la topologie faible dans les questions d'ergodicité et de mélange ; les propriétés de cette topologie sont rappelées en annexe. Le quatrième chapitre présente l'argument de Hopf, un des arguments fameux de la théorie hyperbolique des systèmes dynamiques, et fait le lien avec les cinq chapitres suivants, qui portent sur des questions de dynamique topologique.

Les chapitres suivants portent sur la dynamique des transformations, d'un point de vue topologique. On introduit les concepts de non-errance, de transitivité et de conjugaison, qu'on illustre par la construction de quelques transformations de type Morse-Smale, et par l'étude de la dynamique de quelques polynômes (exemples de Schroeder). Le théorème de linéarisation de Hartman-Grobman permet d'analyser le comportement du système au voisinage de ses points périodiques hyperboliques ; on l'applique à l'étude d'un système obtenu par perturbation d'un automorphisme du tore (dit dérivé d'Anosov), après l'avoir démontré.

Trois chapitres sont consacrés à l'entropie. On démontre le théorème de Kolmogorov-Sinaï sur les partitions génératrices. Comme application, on calcule l'entropie des applications dilatantes (formule de Rokhlin) et de quelques applications de l'intervalle. Un chapitre est consacré à l'interprétation de l'entropie en théorie de l'information.

Les notions d'espace de Lebesgue et de décomposition ergodique sont étudiées dans les deux derniers chapitres. Ces notions importantes sont rarement traitées en détail dans la littérature. L'objectif avoué de ces chapitres est de présenter le théorème de décomposition de manière claire, concise et complète. Pour ce faire, on s'est inspiré de l'argument de Hopf, en construisant les composantes ergodiques de manière "géométrique".

A qui s'adresse ce livre

La plus grande partie de ce livre peut être abordée par un étudiant de master, qui a suivi un cours de théorie de la mesure, et possède le vocabulaire de la théorie des espaces de Hilbert. Certains exemples nécessitent une certaine familiarité avec les notions de flots et de variétés différentielles.

Le chercheur qui désire se familiariser avec la problématique à l'interface des systèmes dynamiques et de la théorie ergodique peut aussi tirer profit de ce livre, par le biais des commentaires situés en fin de chapitre. Ceux-ci donnent un bref aperçu des problèmes et des méthodes qui ont marqué la théorie, et présentent quelques questions ouvertes dans le domaine.

Ce livre comporte quatre annexes, qui résument quelques résultats qui ne font pas forcément partie du cursus de licence ou de maîtrise. On fait usage dès le premier chapitre de la topologie faible dans le cadre des espaces de Hilbert. Cette notion est détaillée dans la première annexe. La seconde annexe présente le théorème de Liouville et sa preuve. Ce théorème est utilisé dans l'exemple du premier chapitre ayant trait à la mécanique classique.

Certains aspects de la théorie des espaces métriques et de la théorie de la mesure ne sont pas forcément abordés en licence sous leur forme la plus générale : séparabilité, support, régularité, densité des fonctions lipschitziennes dans les espaces L^p . Ils sont détaillés dans la troisième annexe. Si le lecteur n'est pas familier avec ces résultats, il est sans doute préférable de les admettre en première lecture. Ils deviennent plus ou moins évidents dès l'instant où l'on travaille sur des ouverts de $\mathbf{R}^n$ avec des mesures du type $f(x) dx$; c'est avec ce genre d'espace en tête que le lecteur est invité à commencer sa lecture.

La dernière annexe sur les partitions et les σ -algèbres n'est pas utilisée dans le texte. Elle traite d'un résultat de théorie de la mesure destiné à éclairer les deux derniers chapitres, et peut être entièrement omise.

Remerciements

Je tiens à remercier les personnes qui, par leurs commentaires et leurs relectures, ont contribué à l'amélioration de ce manuscrit. Il va sans dire que ce texte a beaucoup bénéficié des remarques et des questions des étudiants du Master de Rennes 1.

Théorème ergodique en moyenne

The most useful piece of advice I would give to a mathematics student is always to suspect an impressive sounding theorem if it does not have a special case which is both simple and non trivial.

M. F. Atiyah

La théorie ergodique est l'étude du comportement à long terme des systèmes préservant une certaine forme d'énergie.

D'un point de vue mathématique, un système physique peut être modélisé par la donnée d'un espace X , d'une transformation $T : X \rightarrow X$ et d'une mesure μ définie sur X , invariante par T : pour tout ensemble mesurable $A \subset X$, $\mu(T^{-1}(A)) = \mu(A)$. Le quadruplet formé de l'espace X , de la mesure μ , de la tribu des ensembles mesurables relativement à μ et de la transformation mesurable T qui préserve μ forme ce qu'on appelle un *système dynamique mesuré*.

L'espace X est composé de l'ensemble de tous les états que peut prendre le système au cours de son évolution. La transformation T décrit son évolution au cours du temps ; $T(x)$ est l'état dans lequel se trouve le système au temps 1 s'il se trouvait dans l'état x au temps 0. Les itérés successifs $T^2(x)$, $T^3(x)$,... donnent l'état du système aux temps 2, 3, ... Enfin, la mesure μ correspond à n'importe quelle quantité extensive, définie sur l'espace X , et préservée au cours du mouvement.

L'exemple de base vient de la mécanique classique. Il est donné par un point matériel se déplaçant sous l'action d'un potentiel indépendant du temps. L'ensemble X est l'espace (x, v) des positions-vitesses, aussi appelé espace des phases. La transformation T associe à la condition initiale (x, v) les valeurs de position et de vitesse après un laps de temps donné, par exemple 1 seconde, 1 jour ou 1 année, selon les échelles de temps étudiées. Enfin, la mesure μ est le volume standard $dx dv$ défini sur l'espace X . Son invariance se déduit de la préservation de l'énergie.

On cherche à déterminer le comportement de la suite des itérés $T^n = T \circ T \circ \dots \circ T$. La remarque suivante, due à B. O. Koopman (1931), est cruciale pour la suite. Si on fait opérer par composition la transformation T sur l'espace $L^2(X, \mu)$ des fonctions de carré intégrable, l'application U obtenue est une isométrie linéaire : si $f \in L^2$ et $Uf = f \circ T$, alors $\|Uf\| = \|f\|$. Ceci découle de l'invariance de μ par T . On peut donc appliquer les techniques d'analyse hilbertienne pour étudier le comportement "en moyenne" de la suite $f \circ T^n$, c'est-à-dire son comportement en norme L^2 .

En passant à l'action sur l'espace L^2 , on a remplacé un problème à priori non linéaire, disons en dimension finie, par un problème linéaire en dimension infinie. A-t-on vraiment gagné au change ? Il se trouve que les espaces de Hilbert possèdent un certain nombre de propriétés réminiscentes de la dimension finie. La plus utile est la compacité faible de la boule unité. Montrer une convergence faible revient donc à identifier la limite par le biais d'une propriété qui la caractérise de manière unique, tâche qui s'avère en général plus simple que celle de montrer la convergence.

Ces méthodes hilbertiennes permettent d'obtenir la convergence des moyennes $\frac{1}{n} \sum_{k=1}^n U^k$, pour toute application linéaire U satisfaisant : $\forall f, \|Uf\| \leq \|f\|$. Ce résultat, initialement obtenu par J. Von Neuman (1932) dans un contexte un peu différent par des méthodes de calcul fonctionnel, illustre un fait fréquemment utilisé en analyse, comme quoi "moyenner tend à régulariser".

Voici une conséquence du théorème ergodique : si l'espace X est de mesure finie, alors presque toute trajectoire revient arbitrairement près de son état initial. C'est l'une des rares conclusions générales qu'on puisse faire sur le caractère du mouvement en mécanique classique. Antérieur au théorème ergodique, ce résultat, démontré par H. Poincaré en 1899, est souvent considéré comme le premier résultat mathématique de la théorie ergodique, et marque la naissance de cette discipline.

Théorème ergodique en moyenne

Théorème ergodique de Von Neumann

Soit H un espace de Hilbert et $U : H \rightarrow H$ une application linéaire satisfaisant :
 $\forall f \in H, \|Uf\| \leq \|f\|$. Posons $S_n(f) = \sum_{k=0}^{n-1} U^k f$, $Inv = \{f \in H \mid Uf = f\}$.
 Notons $P : H \rightarrow H$ le projecteur orthogonal sur le sous-espace Inv des vecteurs U -invariants. Alors :

$$\frac{1}{n} S_n(f) \longrightarrow Pf \quad \text{en norme.}$$

Preuve

Montrons que tout élément $g \in H$ invariant par U est invariant par l'adjoint U^* :

$$\|g - U^*g\|^2 = \|g\|^2 + \|U^*g\|^2 - 2\langle g, U^*g \rangle \leq 2\|g\|^2 - 2\langle Ug, g \rangle = 2\langle g - Ug, g \rangle$$

De même, tout $g \in H$ invariant par U^* est invariant par U , par un calcul similaire.

Comme $\frac{1}{n} S_n(f) = f$ si $f \in Inv$, on veut montrer que $\frac{1}{n} S_n(f) \longrightarrow 0$ pour $f \in Inv^\perp$.

On a l'égalité : $\|\frac{1}{n} S_n(f)\|^2 = \langle f, \frac{1}{n} S_n^* \frac{1}{n} S_n(f) \rangle$. Il s'agit donc de vérifier, pour tout $f \in Inv^\perp$, que la suite $\frac{1}{n} S_n^* \frac{1}{n} S_n(f)$ converge faiblement vers 0, ou encore que les valeurs d'adhérence de cette suite sont toutes nulles. Comme elles sont dans $Inv^\perp$, il suffit de montrer qu'elles sont invariantes par U ou par U^* , ce qui découle de la majoration suivante :

$$\|(I - U^*) \frac{1}{n} S_n^* \frac{1}{n} S_n(f)\| \leq \frac{1}{n} \|(I - U^{*n})\| \|\frac{1}{n} S_n(f)\| \leq \frac{2}{n} \|f\| \xrightarrow{n \rightarrow \infty} 0$$

Théorème ergodique L^2

Soit $(X, \mathcal{T}, \mu)$ un espace mesuré, $T : X \rightarrow X$ une application mesurable qui préserve μ , et $f \in L^2(X)$. Alors,

$$\frac{1}{n} \sum_{k=0}^{n-1} f \circ T^k \xrightarrow[n \rightarrow \infty]{L^2} Pf$$

où P est le projecteur orthogonal sur le sous-espace $\{f \in L^2 \mid f \circ T = f\}$; cf fig.1.

Preuve

Il suffit d'appliquer le théorème précédent à l'isométrie de L^2 donnée par $Uf = f \circ T$.

Propriétés du projecteur P

- $\forall f \in L^2, \forall g \in L^2$ tel que $g \circ T = g$, $\int Pf \cdot g \, d\mu = \int f \cdot g \, d\mu$
- $\forall f \in L^2, \forall A \subset X$ tel que $T^{-1}A = A$ et $\mu(A) < \infty$, $\int_A Pf \, d\mu = \int_A f \, d\mu$
- $\forall f \in L^2$ tel que $f \geq 0$, $Pf \geq 0$.

De plus, si $\mu(X) < \infty$,

- $\forall f \in L^2$, $\int Pf \, d\mu = \int f \, d\mu$
- $\forall f \in L^2$ tel que $f \geq 0$, pp $x \in X$, $f(x) > 0$ implique $Pf(x) > 0$.

Preuve

On a l'égalité : $\int Pf \cdot g \, d\mu = \langle Pf, g \rangle = \langle f, Pg \rangle = \int f \cdot g \, d\mu$. On applique cette égalité à la fonction $g = \mathbf{1}_A$ et on prend $A = X$. Montrons maintenant les inégalités. Pour tout $N > 0$, on a la majoration $\mu(\{x \mid Pf(x) < -1/N\}) \leq N^2 \int |Pf|^2 \, d\mu < \infty$. On a donc :

$$-\frac{1}{N} \mu(\{x \mid Pf(x) < -1/N\}) \geq \int_{\{x \mid Pf(x) < -\frac{1}{N}\}} Pf \, d\mu \geq \int_{\{x \mid Pf(x) < -\frac{1}{N}\}} f \, d\mu \geq 0$$

ce qui implique $\mu(\{x \mid Pf(x) < -1/N\}) = 0$, et donc $\mu(\{x \mid Pf(x) < 0\}) = 0$.

Enfin, $\int_{\{x \mid Pf(x)=0\}} f \, d\mu = \int_{\{x \mid Pf(x)=0\}} Pf \, d\mu = 0$ si la mesure de l'ensemble $\{x \mid Pf(x) = 0\}$ est finie. f est donc nulle sur $\{x \mid Pf(x) = 0\}$ si elle est positive.

Théorème de récurrence de Poincaré

Soit $(X, \mathcal{T}, \mu)$ un espace mesuré et $T : X \rightarrow X$ une application mesurable qui préserve μ . On suppose que $\mu(X) < +\infty$. Soit $B \subset X$ un ensemble mesurable. Alors pour presque tout $x \in B$, il existe une infinité de $n \in \mathbf{N}$ tels que $T^n(x) \in B$.

Preuve

Si la trajectoire de x ne passe qu'un nombre fini de fois dans B , on a :

$$\frac{1}{n} \sum_{k=0}^{n-1} \mathbf{1}_B(T^k(x)) \rightarrow 0.$$

Mais le théorème ergodique montre qu'il existe une sous-suite n_i , telle que pour pp $x \in X$, cette somme converge vers $P\mathbf{1}_B(x)$, quantité qui est strictement positive pour pp $x \in B$. Le théorème est illustré par la figure 2.

Application à la mécanique classique

Considérons un point matériel soumis à un champ de force indépendant du temps. Si l'espace est clos, et si l'énergie est conservée au cours du mouvement, on peut montrer qu'il existe une mesure finie invariante dans l'espace des positions-vitesse. D'après le théorème de Poincaré, le système va sûrement revenir dans un état proche de son état initial.

Soit $V : \mathbf{R}^n \rightarrow \mathbf{R}$ une fonction C^2 . L'énergie associée au potentiel V est donnée par :

$$\forall (x, v) \in \mathbf{R}^n \times \mathbf{R}^n, E(x, v) = \frac{1}{2}mv^2 + V(x).$$

Supposons qu'il existe $E_0 \in \mathbf{R}$ tel que la surface d'énergie $E^{-1}(E_0)$ est compacte et $E^{-1}(E_0) \cap \{(x, 0) \mid \nabla V(x) = 0\} = \emptyset$. Alors :

– Pour tout $(x_0, v_0) \in E^{-1}(E_0)$, l'équation différentielle :

$$m \frac{d}{dt} \begin{pmatrix} x \\ v \end{pmatrix} = \begin{pmatrix} mv \\ -\nabla V(x) \end{pmatrix}$$

admet une unique solution $\varphi_t(x_0, v_0)$ satisfaisant $\varphi_0(x_0, v_0) = (x_0, v_0)$, et définie pour tout temps.

– L'énergie E est constante le long des trajectoires du flot φ_t .

– Notons vol_{2n-1} le volume riemannien porté par la variété $E^{-1}(E_0)$. La mesure borélienne $d\mu = \|\nabla E\|^{-1} d\text{vol}_{2n-1}$ est une mesure finie, invariante par les transformations $(x, v) \mapsto \varphi_t(x, v)$, pour tout $t \in \mathbf{R}$. Son support est égal à $E^{-1}(E_0)$.

Preuve de l'invariance de μ

Comme la divergence du vecteur $\begin{pmatrix} x \\ v \end{pmatrix} \mapsto \begin{pmatrix} -\nabla V(x)/m \\ v \end{pmatrix}$ est nulle, la forme volume $dx_1 \wedge \dots \wedge dx_n \wedge dv_1 \wedge \dots \wedge dv_n$ est invariante par φ_t ; cf fig. 3 et annexe. Notons par ω la forme volume sur $E^{-1}(E_0)$ associée au volume riemannien ; elle satisfait la relation : $\|\nabla E\|^{-1} \omega \wedge dE = dx_1 \wedge \dots \wedge dx_n \wedge dv_1 \wedge \dots \wedge dv_n$. Pour établir cette égalité, il suffit d'évaluer chacun des deux termes sur une base de $\mathbf{R}^n \times \mathbf{R}^n$ de la forme $(\frac{\partial \psi}{\partial y_1}, \dots, \frac{\partial \psi}{\partial y_{2n-1}}, \nabla E)$, où $\psi(y_1 \dots y_n)$ est un système de coordonnées sur $E^{-1}(E_0)$. L'invariance de $\|\nabla E\|^{-1} \omega$ se déduit alors du calcul suivant :

$$\varphi_t^* \left(\frac{\omega}{\|\nabla E\|} \right) \wedge dE = \varphi_t^* \left(\frac{\omega}{\|\nabla E\|} \right) \wedge \varphi_t^* dE = \varphi_t^* \left(\frac{\omega \wedge dE}{\|\nabla E\|} \right) = \frac{\omega}{\|\nabla E\|} \wedge dE$$

Corollaire

Soit $B \subset E^{-1}(E_0)$ un ensemble mesurable, relativement à vol_{2n-1} . Alors pour presque tout point de B , la trajectoire associée repasse une infinité de fois dans B .

Il suffit d'appliquer le théorème de récurrence à l'application $(x, v) \mapsto \varphi_1(x, v)$. La récurrence des trajectoires sera étudiée plus en détail dans le chapitre consacré à la dynamique topologique.

figure 1 : Projection sur le sous-espace invariant

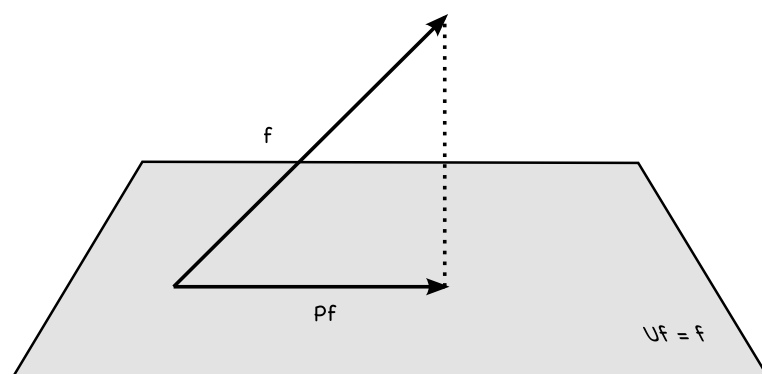


figure 2 : Récurrence

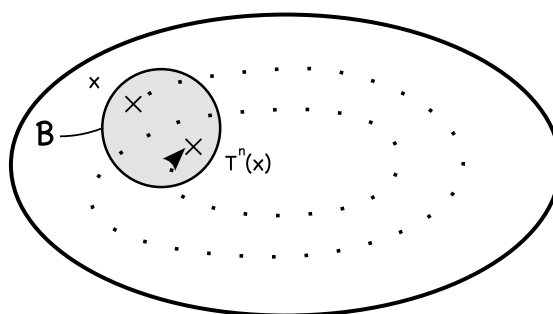
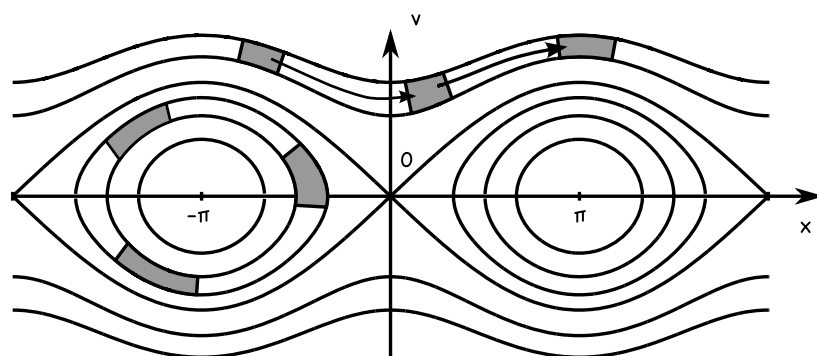


figure 3 : Conservation des aires



Pendule pesant : $\frac{1}{2}mv^2 + \cos(x) = C^{\text{ste}}$

Exercices

exercice 1 :

Soit H un espace de Hilbert, U une isométrie inversible de H et $f \in H$. Montrez que la suite $\frac{1}{2n+1} \sum_{k=-n}^n U^k f$ converge en norme. A quoi est égale sa limite ?

exercice 2 :

Soit H un espace de Hilbert, U une isométrie de H et $f \in H$. Montrez que la suite $U^n f$ converge en norme si et seulement si $Uf = f$.

exercice 3 :

Soit H un espace de Hilbert, U une isométrie de H et $f \in H$. Montrez l'identité :

$$f - \frac{1}{n} S_n(f) = g_n - U g_n, \text{ avec } g_n = \frac{1}{n} \sum_{k=0}^{n-1} S_k(f).$$

Un *cobord* est un élément de H de la forme $g - U g$, avec $g \in H$. Montrez que les cobords sont denses dans l'orthogonal des fonctions U -invariantes.

exercice 4 :

Une contraction U définie sur un espace de Hilbert H est une application linéaire continue de norme inférieure ou égale à un : $\|U\| \leq 1$. Soit $Inv = \{f \in H \mid Uf = f\}$ et soit P le projecteur orthogonal sur Inv .

– Montrez que $P = P^2 = P^* = PU = UP = PU^* = U^*P$.

– On pose $L = \frac{1}{2}(Id + U)$. Montrez que $L^n f \rightarrow Pf$ en norme.

(Indic : montrez l'inégalité $\|L^n(1 - U)\| \leq C_n^{n/2}/2^{n-1}$)

exercice 5 :

Soit H un espace de Hilbert, U une isométrie de H , $\theta \in \mathbf{R}$ et $P_\theta : H \rightarrow H$ le projecteur orthogonal sur le sous-espace $\{f \in H \mid Uf = e^{i\theta} f\}$. Montrez que :

$$\frac{1}{n} \sum_{k=0}^{n-1} e^{-ik\theta} U^k f \xrightarrow[n \rightarrow \infty]{L^2} P_\theta f$$

exercice 6 :

Soit $(X, \mathcal{T}, \mu)$ un espace probabilisé, $T : X \rightarrow X$ une application mesurable qui préserve la mesure μ . Soit P le projecteur orthogonal de $L^2(X)$ sur le sous-espace vectoriel des fonctions T -invariantes.

– Montrez que pour tout $A \subset X$ mesurable, on a : $\|P\mathbf{1}_A\|_2 \geq \mu(A)$.

– Que vaut la limite de la suite : $\frac{1}{n} \sum_{k=0}^{n-1} \mu(A \cap T^{-k}A)$?

– En déduire l'inégalité : $\overline{\lim} \mu(A \cap T^{-n}A) \geq \mu(A)^2$.

exercice 7 :

Soit $(X, \mathcal{T}, \mu)$ un espace probabilisé, $T : X \rightarrow X$ une application mesurable qui préserve la mesure μ et A, B deux sous-ensembles mesurables de X . Que vaut la limite :

$$\frac{1}{n} \sum_{k=0}^{n-1} \mu(A \cap T^{-k}B) \xrightarrow[n \rightarrow \infty]{} ?$$

Montrez l'égalité : $\langle P\mathbf{1}_A, P\mathbf{1}_B \rangle = \mu(A)\mu(B) + \langle P\mathbf{1}_A - \mu(A), P\mathbf{1}_B - \mu(B) \rangle$.

exercice 8 :

Soit $(X, \mathcal{T}, \mu)$ un espace mesuré tel que $\mu(X) < \infty$ et $T : X \rightarrow X$ une application mesurable qui préserve la mesure μ . Soit $A \subset X$ un sous-ensemble mesurable ; montrez que $T^{-1}A \subset A$ implique $\mu(A \setminus T^{-1}A) = 0$.

exercice 9 :

Soit $(X, \mathcal{T}, \mu)$ un espace mesuré tel que $\mu(X) < \infty$ et $T : X \rightarrow X$ une application mesurable qui préserve la mesure μ . Soit $f : X \rightarrow]0, +\infty[$ une fonction mesurable. Montrez que $\sum_{k=0}^{\infty} f(T^k(x)) = \infty$ pour presque tout $x \in X$.

Commentaires

La monographie de Krengel [Kr95] contient une présentation détaillée des théorèmes ergodiques.

La preuve du théorème ergodique en moyenne que nous avons présentée est due à R. Mañé. Il existe d'autres démonstrations :

– La preuve la plus populaire est due à F. Riesz [RN65] et date des années 40. Elle consiste à vérifier le théorème pour les cobords $g - Ug$, puis à montrer la densité des cobords par un calcul direct : supposons U unitaire ; il faut montrer que tout vecteur $f \in H$ orthogonal aux cobords est U -invariant :
 $\forall g \in H, \langle g - Ug, f \rangle = 0$; $\langle g, f \rangle = \langle Ug, f \rangle = \langle g, U^{-1}f \rangle$; par conséquent, $U^{-1}f = f$.

– Dans son livre [RN65], F. Riesz donne une preuve du théorème ergodique basée sur un argument de convexité : soit C un sous-ensemble convexe dans un espace de Hilbert, et μ la borne inférieure des normes des éléments de C . Alors toute suite de C dont la norme converge vers μ est en fait convergente. Ceci se démontre à l'aide de l'inégalité du parallélogramme.

– La preuve originale de Von Neuman (1931) faisait appel au calcul fonctionnel pour les opérateurs unitaires. cf p. ex. Riesz Nagy [RN65] ou Dunford Schwartz [DS88]. Cette preuve peut se résumer comme suit : Soit $U \in \mathcal{L}(H)$ un opérateur unitaire. On peut construire un morphisme d'algèbre de l'ensemble des fonctions $f : S^1 \rightarrow \mathbb{C}$ boréliennes bornées, dans $\mathcal{L}(H)$, qui envoie 1 sur l'identité, et $z \mapsto z$ sur U . L'image de $g : S^1 \rightarrow \mathbb{C}$ est noté $g(U)$. Ce morphisme vérifie de plus : si g_n est une suite uniformément bornée, et si $g_n \rightarrow g$ simplement, alors $\forall f \in H, g_n(U)f \rightarrow g(U)f$ en norme. Pour obtenir le théorème ergodique, Il suffit de prendre $g_n(z) = \frac{1}{n} \sum_{k=0}^{n-1} z^k$ et de remarquer que $\forall z \in S^1, g_n(z) \rightarrow 1_{\{0\}}(z)$.

– Enfin, on peut donner une preuve valable dans n'importe quel Banach réflexif (par exemple dans L^p , $1 < p < \infty$) en utilisant la compacité faible et un lemme de convexité à la Banach-Saks (cf Krengel [Kr95] ch 2). La limite P est identifiée au projecteur d'image Inv et de noyau $Im(id - T)$. Curieusement, cette généralisation est non triviale, même en dimension finie :

Soit Q une matrice $n \times n$ stochastique ($\forall i, j, Q_{i,j} \geq 0$ et $\forall i, \sum_j Q_{i,j} = 1$), qu'on identifie à une contraction de $\mathbf{R}^n$ muni de la norme uniforme. Alors $\frac{1}{n} \sum_{i=1}^n Q^k \rightarrow P$, P projecteur défini précédemment.

Voici une preuve directe de ce résultat, dans l'esprit de F. Riesz : l'espace des matrices stochastiques est convexe compact, et contient les Q^k . Il suffit donc de montrer que P est la seule valeur d'adhérence possible pour $\frac{1}{n} \sum_{i=1}^n Q^k$. Soit P_1 une telle valeur d'adhérence ; un calcul direct montre que $P_1 x = x$ si $x \in \ker(id - Q)$ et $P_1 x = 0$ si $x \in Im(id - Q)$. Ceci montre que les sous-espaces $\ker(Q - id)$ et $Im(Q - id)$ sont en somme directe et que P_1 est le projecteur attendu. Le livre de J. G. Kemeny, J. L. Snell et A. W. Knapp intitulé "Denumerable Markov chains" (ch 6.1) présente une preuve différente de ce résultat.

Il existe un procédé pour généraliser un résultat portant sur des opérateurs unitaires à des contractions arbitraires. Il repose sur le fait suivant (Halmos 1950 ; cf Riesz-Nagy App. §4) :

Soit T une application linéaire définie sur un espace de Hilbert H et satisfaisant $\|T\| \leq 1$. Alors il existe un espace de Hilbert H_1 contenant H et un opérateur unitaire $U : H_1 \rightarrow H_1$ tels que $T^n f = PU^n f$, $T^{*n} f = PU^{-n} f$ pour tout $f \in H$; on a noté P la projection orthogonale de H_1 sur H .

Le théorème ergodique énoncé plus haut est de peu d'utilité lorsque la mesure μ est infinie et que la transformation est ergodique, car dans ce cas il n'y a pas de fonction invariante L^2 non nulle. Lorsque la transformation n'est pas ergodique, il peut arriver que les composantes ergodiques soient finies, auquel cas la limite peut être non nulle ; c'est le cas par exemple pour une rotation définie sur $\mathbf{R}^2$.

Il n'est pas nécessaire que U soit linéaire pour obtenir la convergence faible dans le théorème ergodique en moyenne (théorème ergodique non linéaire de Baillon).

La convergence L^2 de $\frac{1}{n} \sum e^{-ik\theta} f \circ T^k$ est en fait uniforme en θ (théorème de Wiener-Wintner). Des généralisations de ce résultat apparaissent dans les travaux de J. Bourgain (1990).

Il existe des versions topologiques du théorème de récurrence de Poincaré. Si X est un espace métrique, presque tout point appartenant au support de la mesure est topologiquement récurrent : non seulement la trajectoire issue du point revient dans B , mais elle admet une sous-suite, dans B , qui converge vers son point de départ. Ceci sera démontré plus loin. Dans son livre *Measure and Category*, ch.17, J. Oxtoby donne une version abstraite du théorème de récurrence qui unifie les deux aspects, topologique et mesurable.

On peut généraliser le théorème de récurrence de Poincaré dans plusieurs directions ; par exemple, le théorème ergodique de Von Neuman montre que presque tout point de B revient dans B avec une fréquence positive. Une autre généralisation, "proved in many books, never applied" (Krengel), est due à Khintchine : Pour tout $\varepsilon > 0$, on peut trouver $L > 0$ tel que tout intervalle de longueur L contient un entier n satisfaisant : $\mu(B \cap T^{-n}B) \geq (1 - \varepsilon) \mu^2(B)$.

Théorème ergodique presque partout

Le second, de diviser chacune des difficultés que j'examinerois, en autant de parcelles qu'il se pourroit, et qu'il seroit requis pour les mieux résoudre.
R. Descartes (1596-1650)

Considérons un système dynamique, modélisé par la donnée d'un certain espace d'états X , d'une transformation $T : X \rightarrow X$ décrivant l'évolution du système au cours du temps, et d'une mesure finie μ représentant une quantité extensive conservée au cours du mouvement. On cherche à étudier la suite $\{T^n(x)\}_{n \in \mathbb{N}}$, qui représente la succession des états que le système adopte au cours du temps. Cette suite constitue la *trajectoire* du point x , ou encore son *orbite*.

Intéressons nous au comportement asymptotique de cette suite. Pour cela, considérons une certaine quantité observable $f : X \rightarrow \mathbf{R}$ et étudions son évolution au cours du temps. Les quantités : $S_n(f)(x) = \sum_{k=0}^{n-1} f(T^k(x))$ sont appelées sommes de Birkhoff de la fonction f et les moyennes :

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

sont les *moyennes de Birkhoff* de f . G. D. Birkhoff montre en 1932 que la suite des moyennes $\frac{1}{n} S_n(f)(x)$ converge pour presque tout $x \in X$, dès que la fonction f est intégrable. Lorsque f est la fonction indicatrice d'un certain ensemble $A \subset X$, ces moyennes correspondent aux fréquences de passage, entre les temps 0 et $n-1$, des itérés de x dans l'ensemble A . Ces fréquences convergent, et la limite est le temps moyen passé par x dans A au cours de son déplacement.

Sans doute l'idée la plus naturelle pour attaquer un problème, est de chercher à le subdiviser en plusieurs sous-problèmes, qui seront avec un peu de chance plus faciles à traiter que le problème initial. Pour étudier la dynamique d'une transformation, on peut chercher à "casser" l'espace X en plusieurs morceaux disjoints, chacun de mesure non nulle, afin de restreindre la transformation à chacun de ces morceaux ; cf fig.2. Si cela n'est pas possible, le système est dit *ergodique*.

Lorsqu'un système est ergodique, il est possible de calculer explicitement la limite des moyennes de Birkhoff $\frac{1}{n} S_n(f)$. Cette limite ne dépend pas de x et s'obtient en moyennant f sur X relativement à la mesure considérée. On peut donc dire que pour un système ergodique,

"Les moyennes temporelles coïncident avec les moyennes spatiales."

Le théorème ergodique de Birkhoff permet donc de passer d'une propriété de nature qualitative : *pas d'ensembles invariants non triviaux*, à un énoncé quantitatif : *la fréquence de passage dans un ensemble quelconque est proportionnelle à la taille de cet ensemble*. En particulier, les trajectoires visitent tout l'espace au cours du mouvement, si le système est ergodique ; cf fig.3.

Les exemples les plus simples de systèmes ergodiques proviennent de la théorie des probabilités. Considérons une épreuve aléatoire, comme le lancer d'un dé ou le tirage d'une boule dans une urne. Soit Ω l'ensemble des résultats pouvant être obtenus à l'issue de cette épreuve et P la mesure de probabilité sur Ω associée à ces résultats. La répétition de cette épreuve, de manière indépendante et un nombre indéfini de fois, peut se modéliser en considérant l'espace des suites de résultats $\Omega^{\mathbf{N}}$ muni de la probabilité produit $P^{\otimes \mathbf{N}}$ et de la transformation "décalage", qui consiste à supprimer le premier élément de la suite et à décaler les autres éléments d'un cran vers la gauche. Dans ce contexte, le théorème ergodique, couplé à l'ergodicité du décalage, redonne la loi forte des grands nombres, dont la première démonstration dans ce cadre général est due à A. N. Kolmogorov (1933).

Théorème ergodique de Birkhoff (1932)

Soit $(X, \mathcal{T}, \mu)$ un espace mesuré tel que $\mu(X) < \infty$, $T : X \rightarrow X$ une application mesurable qui préserve la mesure μ et $f : X \rightarrow \mathbf{R}$ une fonction intégrable. Notons $\mathcal{I} = \{A \subset X \mid A \text{ mesurable et } T^{-1}A = A\}$. Alors,

$$pp \ x \in X, \quad \frac{1}{n} \sum_{k=0}^{n-1} f(T^k(x)) \xrightarrow{n \rightarrow \infty} E(f \mid \mathcal{I})(x)$$

Preuve

Posons $\bar{f}(x) = \overline{\lim} \frac{1}{n} \sum_{k=0}^{n-1} f(T^k(x))$, $\underline{f}(x) = \underline{\lim} \frac{1}{n} \sum_{k=0}^{n-1} f(T^k(x))$.

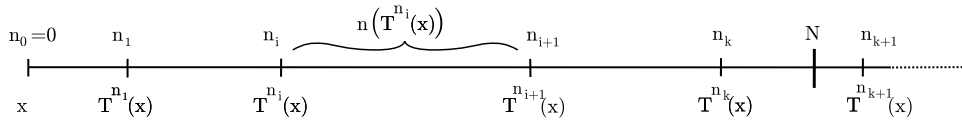
Pour montrer le théorème, il suffit d'obtenir l'encadrement $\int \bar{f} d\mu \leq \int f d\mu \leq \int \underline{f} d\mu$ car alors $\bar{f} - \underline{f}$ est une fonction positive d'intégrale nulle, donc nulle presque partout. On montre $\int \bar{f} d\mu \leq \int f d\mu$, l'autre inégalité s'obtient en changeant le signe de f .

Supposons pour commencer f bornée par une constante $M > 0$. Il en va alors de même pour $\bar{f}$. Rappelons que la limite supérieure d'une suite est sa plus grande valeur d'adhérence. Pour tout $\varepsilon > 0$, il existe donc une infinité de $n \in \mathbf{N}^*$ qui satisfont l'inégalité $\frac{1}{n} \sum_{k=0}^{n-1} f(T^k(x)) \geq \bar{f}(x) - \varepsilon$. Soit $n(x)$ le plus petit de ces entiers :

$$\bar{f}(x) \leq \frac{1}{n(x)} \sum_{k=0}^{n(x)-1} f(T^k(x)) + \varepsilon$$

L'intersection des ensembles $\{x \in X \mid n(x) > R\}$, $R \in \mathbf{N}$, est vide. Choisissons R de telle sorte que $A = \{x \in X \mid n(x) > R\}$ soit de mesure plus petite que ε .

Définissons par récurrence une suite n_i , dépendant de x , comme suit ; cf fig.1 :



- Si $T^{n_i}(x) \notin A$, on pose $n_{i+1} = n_i + n(T^{n_i}(x))$ et on utilise la majoration

$$n(T^{n_i}(x)) \bar{f}(T^{n_i}x) \leq \sum_{j=0}^{n(T^{n_i}(x))-1} f(T^j(T^{n_i}x)) + n(T^{n_i}(x)) \varepsilon.$$

- Si $T^{n_i}(x) \in A$, on pose $n_{i+1} = n_i + 1$ et on utilise la majoration $\bar{f}(x) \leq M$.

Soit $\tilde{f} = f \mathbf{1}_{A^c} + M \mathbf{1}_A$. Comme $\bar{f}$ est T -invariante, ces majorations impliquent

$$\forall x \in X, \quad (n_{i+1} - n_i) \bar{f}(x) \leq \sum_{j=n_i}^{n_{i+1}-1} \tilde{f}(T^j(x)) + (n_{i+1} - n_i) \varepsilon.$$

Soit $N \in \mathbf{N}^*$ et k l'entier (dépendant de x) satisfaisant $n_k \leq N < n_{k+1}$. Remarquons que $0 \leq N - n_k \leq R$ et sommons les inégalités précédentes :

$$N \bar{f}(x) = \sum_{i=0}^k (n_{i+1} - n_i) \bar{f}(x) + (N - n_k) \bar{f}(x) \leq \sum_{j=0}^N \tilde{f}(T^j(x)) + N \varepsilon + 2RM.$$

On peut alors intégrer : $\int \bar{f} d\mu \leq \int \tilde{f} d\mu + (\varepsilon + \frac{2RM}{N}) \mu(X)$, $\int \tilde{f} d\mu \leq \int f d\mu + 2M\varepsilon$.

Il reste à faire tendre N vers l'infini puis ε vers 0 pour obtenir la majoration voulue.

Lorsque f n'est pas bornée, on procède comme suit. L'ensemble $\{\bar{f} = -\infty\}$ reste négligeable car $\bar{f}$ est minorée par la fonction intégrable $-\underline{\lim} \frac{1}{n} \sum_{k=0}^{n-1} |f \circ T^k|$. On fixe une constante $M > 0$ et on choisit A tel que $\int_A |f| + M d\mu < \varepsilon$. On reprend le calcul précédent avec $\bar{f}$ remplacé par $\min(\bar{f}, M)$ et $\tilde{f}$ par $f + (|f| + M) \mathbf{1}_A$. De là, on obtient l'inégalité $\int \min(\bar{f}, M) d\mu \leq \int f d\mu$. Il suffit de faire tendre M vers l'infini pour conclure.

Identification de la limite

Commençons par le cas f bornée et montrons que la limite presque partout, notée $\bar{f}$, vérifie les propriétés qui caractérisent l'espérance conditionnelle. En premier lieu, $\bar{f}$ est invariante par T donc mesurable relativement à $\mathcal{I}$. Ensuite, par convergence dominée, $\int \bar{f} d\mu = \int f d\mu$. Soit A un ensemble invariant mesurable. Comme $\mathbf{1}_A = \mathbf{1}_A \circ T$, on a l'égalité $\mathbf{1}_A \bar{f} = \overline{\mathbf{1}_A f}$, ce qui entraîne :

$$\int_A \bar{f} d\mu = \int \overline{\mathbf{1}_A f} d\mu = \int \mathbf{1}_A f d\mu = \int_A f d\mu.$$

Pour f non bornée, on approche f par une fonction g bornée et on remarque que :

$$\left\| \frac{1}{n} \sum_{k=0}^{n-1} (f - g) \circ T^k \right\|_1 \leq \|f - g\|_1, \quad \|E(f - g | \mathcal{I})\|_1 \leq \|f - g\|_1.$$

Remarques

- La convergence a aussi lieu en norme L^1 . Pour f bornée, c'est une conséquence du théorème de convergence dominée ; pour f intégrable, on raisonne comme ci-dessus.
- L'ensemble $\{x \in X \mid \frac{1}{n} \sum_{k=0}^{n-1} f(T^k(x)) \text{ CV} \}$ dépend de f mais pas de μ .
- Une fonction f est $\mathcal{I}$ -mesurable si et seulement si elle est invariante par T ; en effet, les égalités $f(T(x)) = f(x)$ et $T^{-1}f^{-1}(\{f(x)\}) = f^{-1}(\{f(x)\})$ sont équivalentes.

Définition

Soit $(X, \mathcal{T}, \mu)$ un espace mesuré, $T : X \rightarrow X$ une application mesurable qui préserve la mesure μ . La transformation T est dite *ergodique relativement à la mesure μ* si les seuls ensembles mesurables invariants sont de mesure nulle ou de complémentaire de mesure nulle :

$$T^{-1}A = A \quad \text{implique} \quad \mu(A) = 0 \quad \text{ou} \quad \mu(A^c) = 0$$

Proposition

Une transformation T est ergodique si et seulement si les fonctions mesurables invariantes par T sont constantes presque partout. Lorsque $\mu(X)$ est fini non nul, ceci implique l'égalité : $\forall f \in L^1, \quad E(f | \mathcal{I}) = \frac{1}{\mu(X)} \int_X f d\mu$.

Preuve

Soit g une fonction invariante. Posons $C = \sup \{t \mid \mu(g^{-1}([-\infty, t])) = 0\}$. Par ergodicité, nous avons l'égalité : $C = \inf \{t \mid \mu(g^{-1}([t, \infty])) = 0\}$. La constante C est finie dès que $\mu(X) \neq 0$, auquel cas les ensembles $g^{-1}([-\infty, C])$ et $g^{-1}(]C, \infty])$ sont négligeables, et la fonction g est constante égale à C pp. La fonction $E(f | \mathcal{I})$ est donc constante et son intégrale vaut $\int f d\mu$. Ceci donne l'égalité recherchée.

Exemple

Soit $(\Omega, \mathcal{T}, \mu)$ un espace probabilisé. On définit sur l'espace produit $(\Omega^{\mathbf{N}}, \mathcal{T}^{\otimes \mathbf{N}}, \mu^{\otimes \mathbf{N}})$ une transformation T en posant $T(\{x_i\}) = \{x_{i+1}\}$. Alors T est ergodique.

Preuve

Soit f une fonction intégrable invariante. Pour tout $\varepsilon > 0$, on peut trouver $g \in L^1$ qui ne dépend que d'un nombre n fini de coordonnées, et telle que $\|g - f\|_1 < \varepsilon/4$.

$$\|g - g \circ T^n\|_1 \leq \|g - f\|_1 + \|f - f \circ T^n\|_1 + \|f \circ T^n - g \circ T^n\|_1 \leq \varepsilon/2$$

Calculons la norme de $g - g \circ T^n$ explicitement :

$$\begin{aligned} \|g - g \circ T^n\|_1 &= \int |g(x_0, \dots, x_{n-1}) - g(x_n, \dots, x_{2n-1})| d\mu(x_0) \dots d\mu(x_{2n-1}) \\ &= \int |g(x_0, \dots, x_{n-1}) - g(y_0, \dots, y_{n-1})| d\mu(x_0) \dots d\mu(x_{n-1}) d\mu(y_0) \dots d\mu(y_{n-1}) \\ &= \int |g(x) - g(y)| d\mu^{\otimes \mathbf{N}}(x) d\mu^{\otimes \mathbf{N}}(y) \end{aligned}$$

Puis, en utilisant à nouveau le fait que g et f sont proches en norme L^1 :

$$\int |f(x) - f(y)| d\mu^{\otimes \mathbf{N}}(x) d\mu^{\otimes \mathbf{N}}(y) \leq \int |g(x) - g(y)| d\mu^{\otimes \mathbf{N}}(x) d\mu^{\otimes \mathbf{N}}(y) + 2\|f - g\| \leq \varepsilon$$

Ceci montre que $f(x) = f(y)$ pour pp $(x, y) \in \Omega^{\mathbf{N}} \times \Omega^{\mathbf{N}}$; f est donc constante pp.

figure 1 : Preuve du théorème ergodique

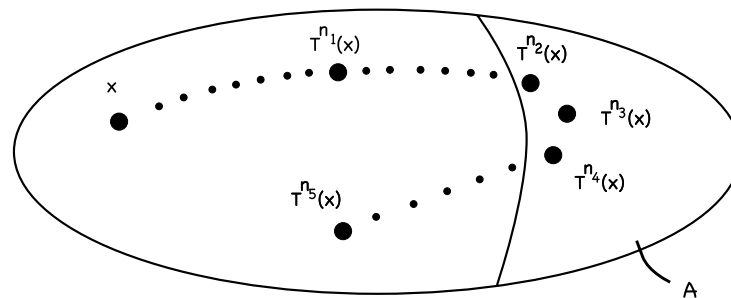


figure 2 : Un système qui n'est pas ergodique

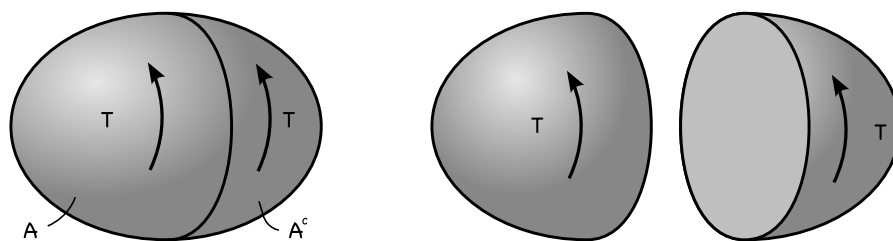
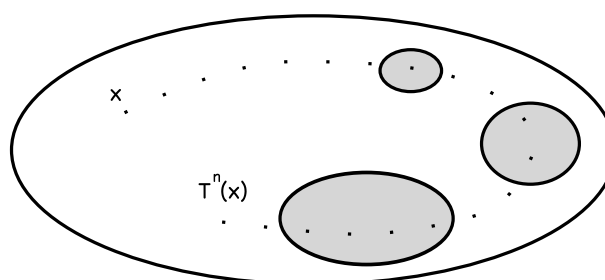


figure 3 : Ergodicité



Le temps passé dans un ensemble est proportionnel
à la mesure de cet ensemble.

Exercices

exercice 1 :

Vérifiez que l'application $n(x)$ qui intervient dans la preuve du théorème ergodique est mesurable.

exercice 2 :

Soit $(X, \mathcal{T}, \mu)$ un espace mesuré, $\mu(X) < \infty$, $T : X \rightarrow X$ une application mesurable qui préserve la mesure μ et $f : X \rightarrow \mathbf{R}$ intégrable. Montrez que pour μ -pp $x \in X$, $\frac{1}{n}f(T^n(x)) \rightarrow 0$.

exercice 3 :

Soit $(X, \mathcal{T}, \mu)$ un espace probabilisé, $T : X \rightarrow X$ une application mesurable qui préserve la mesure μ , et A un sous-ensemble mesurable de X . Montrez que :

$$\text{pp } x \in X, \quad \frac{1}{n} \text{Card} \{k \in \{0, 1, \dots, n-1\} \mid T^k(x) \in A\} \xrightarrow{n \rightarrow \infty} \mu(A).$$

exercice 4 :

Soit $(X, \mathcal{T}, \mu)$ un espace mesuré et $T : X \rightarrow X$ une application mesurable qui préserve la mesure μ . Montrez que T est ergodique si et seulement si pour tous les ensembles mesurables $A, B \subset X$ de mesure positive, presque tout point de A a une orbite qui passe une infinité de fois dans B .

exercice 5 :

Soit α un nombre réel irrationnel ; montrez que la transformation T de $\mathbf{R}/\mathbf{Z}$ dans $\mathbf{R}/\mathbf{Z}$ donnée par $T(x) = x + \alpha \bmod 1$ préserve la mesure de Lebesgue et est ergodique. Pour cela, on pourra regarder ce qu'implique l'invariance de f par T sur ses coefficients de Fourier.

Quelle condition faut-il sur $\alpha \in \mathbf{R}$ pour que la transformation de $(\mathbf{R}/\mathbf{Z})^n$ donnée par $T(x) = x + \alpha \bmod 1$ soit ergodique relativement à la mesure de Lebesgue ?

exercice 6 :

Soit α un nombre réel irrationnel ; montrez que la transformation T définie de $\mathbf{R}/\mathbf{Z} \times \mathbf{R}/\mathbf{Z}$ dans $\mathbf{R}/\mathbf{Z} \times \mathbf{R}/\mathbf{Z}$ par la formule $T(x, y) = (x + \alpha, x + y)$ préserve la mesure de Lebesgue et est ergodique. Là encore, on pourra regarder ce qu'implique l'invariance de f par T sur ses coefficients de Fourier.

exercice 7 :

Démontrez la loi des grands nombres :

Soit $(\Omega, \mathcal{T}, P)$ un espace probabilisé, $X_i : \Omega \rightarrow \mathbf{R}$ une suite de fonctions intégrables, telles que pour tout $i \in \mathbf{N}$, $X_{i*}P = X_{0*}P$, et qui sont indépendantes :

$$\forall n \geq 0, \quad \forall f_1 \dots f_n \text{ bornées, } \int f_1(X_0(\omega)) \dots f_n(X_n(\omega)) dP(\omega) = \prod \int f_i(X_0) dP.$$

Alors la suite $\frac{1}{n} \sum_{k=0}^{n-1} X_k$ converge vers $\int X_0 dP$ presque partout.

On pourra considérer l'application $\varphi : \Omega \rightarrow \mathbf{R}^{\mathbf{N}}$ donnée par $\varphi(\omega) = \{X_i(\omega)\}$, et étudier le décalage sur l'ensemble $\mathbf{R}^{\mathbf{N}}$ muni de la mesure φ_*P .

exercice 8 :

Donnez un exemple de transformation ergodique T telle que $T \circ T$ n'est pas ergodique.

exercice 9 :

Soit T une transformation ergodique sur un espace mesuré $(X, \mathcal{T}, \mu)$. On note $P : L^2(X) \rightarrow L^2(X)$ le projecteur orthogonal sur les fonctions T -invariantes. On suppose que $\mu(X) = \infty$. Montrez que : $\forall f \in L^2(X), Pf = 0$.

exercice 10 * :

Construisez une transformation de la droite réelle qui préserve la mesure de Lebesgue, et qui est ergodique.

Commentaires

La preuve du théorème ergodique présentée plus haut est due à Y. Katznelson et B. Weiss (1982). Il existe d'autres démonstrations:

- La démonstration originale du théorème ergodique presque partout, due à G. Birkhoff (1931), fait appel à une *inégalité maximale*. Cette inégalité a été ensuite généralisée et simplifiée par Wiener (1939), Yosida, Kakutani (1939), Pitt (1942), Riesz (1945), Hopf (1954) ... La version suivante, dû à A. Garsia (1965), admet une preuve élémentaire :

Soit $U: L^1 \rightarrow L^1$ linéaire tel que $\|Uf\| \leq \|f\|$. Notons $E_n = \{x \mid \max_{0 \leq m \leq n} S_m f > 0\}$. Alors $\int_{E_n} f \geq 0$.

- E. Bishop donne en 1966 une preuve du théorème ergodique inspirée de la théorie des martingales, et qui repose sur des inégalités "upcrossing".

- T. Kamae donne en 1982 une démonstration basée sur l'analyse non-standard. La preuve de Y. Katznelson et B. Weiss présentée plus haut est inspirée de la démonstration de Kamae.

- P. Shields donne en 1987 une nouvelle preuve du théorème ergodique, qui ne repose pas sur une inégalité maximale. J. Bourgain propose en 1988 une démonstration basée sur des inégalités variationnelles. Plus récemment, M. Keane et K. Petersen (2006) ont proposés des preuves "élémentaires" du théorème ergodique, dans la lignée de Katznelson et Weiss.

Il existe des versions du théorème ergodique pour les contractions de L^1 . L'énoncé le plus général est sans doute dû à Chacon ; cf Krengel[Kr95] Ch 4 Th 1.11. Il se démontre à l'aide d'un "schéma de remplissage".

Si $\mu(X) = \infty$, on a encore convergence presque partout dans le théorème ergodique, mais la limite n'est plus forcément donnée par une espérance conditionnelle. En particulier, cette limite est nulle si T est ergodique. Lorsque la mesure est infinie, mais que la transformation T est récurrente, E. Hopf donne une version "quotient" du théorème ergodique : pour $f, g \in L^1$ positives, le quotient $S_n f / S_n g$ converge vers $\int f / \int g$. Cet énoncé est étendu aux contractions positives par R. Chacon et D. Ornstein (1960). Là encore, il peut être démontré en passant par une inégalité maximale. Il peut aussi se déduire du théorème en mesure finie par des techniques d'induction (R. Zweimüller 2004).

Pour démontrer le théorème ergodique, on peut se restreindre au cas d'un décalage sur $\mathbf{R}^{\mathbf{N}}$, avec pour observable la projection sur la première coordonnée. Le cas général s'en déduit en factorisant le système par le biais du morphisme $\varphi: X \rightarrow \mathbf{R}^{\mathbf{N}}$ donné par $x \mapsto \{f(T^i(x))\}_{i \in \mathbf{N}}$. Cette remarque est utilisée dans la preuve donnée par Kamae.

Le théorème ergodique s'intéresse aux moyennes des puissances d'un opérateur. Il existe également des résultats sur la convergence presque partout des puissances elles-mêmes. Le théorème suivant, dû à Rota (1962) et Stein (1961), s'applique par exemple à l'opérateur auto-adjoint $Tf = 1/2(f \circ T + f \circ T^{-1})$ et donne une version pondérée du théorème ergodique :

*Soit $T: L^1 \rightarrow L^1$ tel que $\|T\|_1 \leq 1$, $\|Tf\|_\infty \leq \|f\|_\infty$ pour f bornée, $Tf \geq 0$ pour $f \geq 0$, $T1 = 1$ et $T^*1 = 1$. Alors $T^n T^{*n} f$ converge presque partout si $f \in L^p$, $1 < p < \infty$.*

La convergence n'a pas forcément lieu pour tout $f \in L^1$ (D. Ornstein, 1968).

Le comportement des sommes ergodiques du point de vue topologique est différent de son comportement en mesure. L'ensemble $\{(x_n) \in \{0, 1\}^{\mathbf{N}} \mid \frac{1}{n} \sum x_i \text{ converge}\}$ est de mesure totale pour toute mesure de probabilité définie sur $\{0, 1\}^{\mathbf{N}}$, invariante par le décalage. Pourtant, l'ensemble des suites $(x_i)_{i \in \mathbf{N}}$, pour lesquelles tout réel de $[0, 1]$ est valeur d'adhérence des moyennes $\frac{1}{n} \sum x_i$, est un G_δ -dense de $\{0, 1\}^{\mathbf{N}}$.

Mélange

Pour apprendre quelque chose aux gens, il faut mélanger ce qu'ils connaissent avec ce qu'ils ignorent.
P. Picasso (1881-1973)

Considérons un potentiel V défini sur $\mathbf{R}^3$ et intéressons nous au mouvement d'un point matériel sous l'action du champ de force engendré par ce potentiel. Soit $(x, v) \in \mathbf{R}^3 \times \mathbf{R}^3$ la position et la vitesse initiale du point matériel. On note par $T(x, v)$ la position du point au temps 1. L'énergie initiale du système est donnée par la formule : $E(x, v) = \frac{1}{2}mv^2 + V(x)$, elle est préservée au cours du mouvement. Lorsque la surface d'énergie $E(x, v) = E_0$ est bornée, on peut restreindre la mesure de Lebesgue $dx dv$ à cette surface de façon à obtenir une mesure de probabilité, qu'on note μ .

Cherchons à étudier la propagation d'un gaz ou d'un liquide sous l'action du potentiel V . La distribution initiale du gaz peut être représentée par une mesure de probabilité de la forme $d\nu = h d\mu$, avec h une fonction positive définie sur la surface d'énergie considérée. Si A est un sous-ensemble de cette surface d'énergie, $\nu(A)$ représente la quantité de gaz ou de liquide présent dans A . On peut aussi l'interpréter comme la probabilité qu'une particule soit dans la région A à l'instant initial.

Comment modéliser l'évolution du gaz ? Une première approche, très naïve, consiste à négliger les interactions au sein du gaz et considérer que chaque molécule se déplace conformément aux lois classiques du mouvement. La distribution du gaz à l'instant 1 est alors donné par la mesure $T_*\nu$, qui est définie par $T_*\nu(A) = \nu(T^{-1}A)$ pour tout $A \subset X$ mesurable.

La suite $T_*^n\nu$ représente l'évolution du gaz au cours du temps. Si cette suite converge vers la mesure μ , on dit que la transformation est *mélangeante* relativement à μ : toute distribution initiale de gaz de la forme $h d\mu$ finit par se répartir de manière uniforme sur la surface d'énergie, suivant la loi μ .

La propriété de mélange est plus forte que l'ergodicité. Elle exclut un comportement limite périodique (e.g. $T^n = id$ pour un certain $n \geq 2$), alors qu'un tel comportement est possible pour une transformation ergodique. L'ergodicité de la mesure μ est en fait équivalente à la convergence des moyennes $\frac{1}{n} \sum T_*^k\nu$ vers μ , pour toute mesure de probabilité ν de la forme $h d\mu$.

Les décalages sur les espaces produits sont mélangeants relativement aux mesures produits. Pour ces systèmes, il est d'usage de déduire l'ergodicité du mélange, car les preuves sont du même ordre de difficulté. Une deuxième famille d'applications mélangeantes est donnée par les *automorphismes hyperboliques* des tores. Ces applications sont obtenues en considérant des matrices de déterminant un, à coefficients entiers, sans valeurs propres de module 1. L'action d'une telle matrice sur l'espace quotient $\mathbf{T}^n = \mathbf{R}^n/\mathbf{Z}^n$ préserve la mesure de Lebesgue, et donne une application mélangeante relativement à cette mesure. Ce résultat peut se démontrer à l'aide des séries de Fourier.

Mélange

Définition

Soit $(X, \mathcal{T}, \mu)$ un espace probabilisé et $T : X \rightarrow X$ une application mesurable qui préserve la mesure μ . La transformation T est *mélangeante* relativement à la mesure μ si elle vérifie :

$$\forall A, B \subset X \text{ mesurables}, \quad \mu(A \cap T^{-n}B) \xrightarrow{n \rightarrow \infty} \mu(A) \mu(B). \quad \text{cf fig.1.}$$

Théorème

Une transformation mélangeante est ergodique.

Preuve

Soit $A \subset X$ un ensemble invariant ; comme $T^{-n}A = A$, on doit avoir $A \cap T^{-n}A = A$ et le mélange implique $\mu(A) = \mu(A)^2$, c'est-à-dire $\mu(A) = 0$ ou 1.

Critère

Soit $\mathcal{D}$ un sous-ensemble de L^2 qui engendre un sous-espace vectoriel dense dans L^2 . La transformation T est mélangeante si et seulement si pour tout $f, g \in \mathcal{D}$,

$$\int f \circ T^n g \, d\mu \xrightarrow{n \rightarrow \infty} \int f \, d\mu \int g \, d\mu.$$

Preuve

L'expression $(f, g) \mapsto \int f \circ T^n g \, d\mu$ est bilinéaire ; la convergence a donc lieu pour tout $f, g \in \text{Vect}(\mathcal{D})$. Soit $f, g \in L^2$ et $f', g' \in \text{Vect}(\mathcal{D})$ proche de f et g .

$$\begin{aligned} \int f \circ T^n g \, d\mu - \int f' \circ T^n g' \, d\mu &= \langle f \circ T^n, g \rangle - \langle f' \circ T^n, g' \rangle \\ &= \langle (f - f') \circ T^n, g \rangle - \langle f' \circ T^n, g - g' \rangle \\ &= \leq \| (f - f') \| \|g\| + \|f'\| \|g - g'\| \end{aligned}$$

La quantité $\int f' \circ T^n g' \, d\mu$ est proche de $\int f' \, d\mu \int g' \, d\mu$ lorsque n est grand, c'est à dire proche de $\int f \, d\mu \int g \, d\mu$. On a donc la convergence pour tout $f, g \in L^2$. On termine en remarquant que $\int f \circ T^n g \, d\mu = \mu(A \cap T^{-n}B)$ si $f = \mathbf{1}_B$ et $g = \mathbf{1}_A$.

En particulier, T est mélangeante si et seulement si la suite $f \circ T^n$ converge faiblement vers une constante, pour tout $f \in L^2$. Lorsque X est un espace métrique et que T et μ sont boréliennes, le mélange peut s'exprimer à l'aide de la *convergence étroite* ; la transformation T est mélangeante si et seulement si

$$\text{pour tout } g \in L^2 \text{ tel que } \int g \, d\mu = 1, \quad T_*^n(g \, d\mu) \xrightarrow{n \rightarrow \infty} d\mu \text{ étroitement.}$$

C'est une conséquence de la densité des fonctions continues bornées dans L^2 .

Exemple de la multiplication par 2

On considère la transformation de $[0, 1[$ dans $[0, 1[$ donnée par :

$$\begin{aligned} T(x) &= 2x & \text{si } x \in [0, \frac{1}{2}[\\ &= 2x - 1 & \text{si } x \in [\frac{1}{2}, 1[\end{aligned}$$

Montrons qu'elle préserve la mesure de Lebesgue et qu'elle est mélangeante.

L'image réciproque d'un intervalle $[a, b]$ par T est une union disjointe de 2 intervalles de longueur $\frac{(b-a)}{2}$; cf fig.2. La transformation conserve donc la mesure de Lebesgue.

Pour démontrer le mélange, on peut se restreindre au cas où A est de la forme $[k/2^n, k+1/2^n[$, $n \in \mathbf{N}$, $0 \leq k \leq 2^n - 1$ car ces intervalles engendrent la tribu des boréliens. L'ensemble $T^{-N}[k/2^n, (k+1)/2^n[$ est composé des 2^N intervalles suivants :

$$\left[\frac{(k+i \cdot 2^n)}{2^{n+N}}, \frac{k+1+i \cdot 2^n}{2^{n+N}} \right[$$

(i entier). Si $n+N > n'$, l'intersection de ces intervalles avec $B = [k'/2^{n'}, k'+1/2^{n'}[$ est constituée de $2^{N-n'}$ intervalles de longueur 2^{-n-N} , ce qui donne la relation recherchée : $\mu(B \cap T^{-n}A) = \mu(A) \mu(B)$.

Exemple du décalage de Bernoulli

L'exemple suivant vient de la théorie des probabilités. Il entraîne l'ergodicité du décalage et permet de retrouver la loi des grands nombres pour des variables aléatoires indépendantes de même loi.

Théorème

Soit $(\Omega, \mathcal{T}, \mu)$ un espace probabilisé. On définit sur l'espace produit $(\Omega^{\mathbf{N}}, \mathcal{T}^{\otimes \mathbf{N}}, \mu^{\otimes \mathbf{N}})$ une transformation σ en posant $\sigma(\{x_i\}) = \{x_{i+1}\}$. Alors σ est mélangeante relativement à $\mu^{\otimes \mathbf{N}}$.

Preuve

Les fonctions de $L^2(X)$ qui ne dépendent que d'un nombre fini de coordonnées sont denses dans L^2 . Soit donc f, g dépendant de j coordonnées et supposons $n > j$.

$$\begin{aligned} \int g \circ f \circ \sigma^n d\mu^{\otimes \mathbf{N}} &= \int g(x_0, \dots, x_j) f(x_n, \dots, x_{j+n}) d\mu^{\otimes \mathbf{N}} \\ &= \int g(x_0, \dots, x_j) f(x_n, \dots, x_{j+n}) d\mu(x_0) \dots d\mu(x_{j+n}) \\ &= \int g(x_0, \dots, x_j) d\mu(x_0) \dots d\mu(x_j) \int f(x_n, \dots, x_{j+n}) d\mu(x_n) \dots d\mu(x_{j+n}) \\ &= \int f d\mu^{\otimes \mathbf{N}} \int g d\mu^{\otimes \mathbf{N}} \end{aligned}$$

Ceci démontre le mélange du décalage σ .

Lorsque Ω est fini, le système dynamique constitué par la transformation σ sur $\Omega^{\mathbf{N}}$ et la mesure $\mu^{\otimes \mathbf{N}}$ est appelé *système de Bernoulli*. Numérotions les éléments de Ω : $\Omega = \{x_1, \dots, x_k\}$ et posons $p_i = \mu(\{x_i\})$. Le système est complètement déterminé par les paramètres $p_1, \dots, p_k$.

Exemple des endomorphismes des tores

Théorème

Soit A une matrice $n \times n$ à coefficients entiers de déterminant non nul. Cette matrice induit une application sur le quotient $\mathbf{T}^n = (\mathbf{R}/\mathbf{Z})^n$ qui préserve la mesure de Lebesgue. Cette application est mélangeante si A n'a aucune valeur propre qui est racine de l'unité.

Les applications de $\mathbf{T}^n$ obtenues à partir de matrices de déterminant 1, sans valeurs propres de module 1, sont appelées *automorphismes hyperboliques de $\mathbf{T}^n$* .

Preuve

On peut démontrer l'invariance de la mesure par le biais des séries de Fourier. Soit $k \in \mathbf{Z}^n$ et $k \cdot x$ la quantité $\sum k_i x_i$. Posons $e_k(x) = e^{ik \cdot x}$; ces fonctions forment une base hilbertienne de $L^2(\mathbf{T}^n)$. Soit $f \in L^2$ et c_k ses coefficients de Fourier.

$$\int_{\mathbf{T}^n} f(Ax) dx = \sum_{k \in \mathbf{Z}^n} c_k \int e^{i k \cdot Ax} dx = \sum_{k \in \mathbf{Z}^n} c_k \int e^{i({}^t A k) \cdot x} dx = c_0 = \int_{\mathbf{T}^n} f(x) dx.$$

Soit $k, l \in \mathbf{Z}^n$. Montrons à présent le mélange de la transformation :

$$\int_{\mathbf{T}^n} e_k(x) e_l(A^n x) dx = \int_{\mathbf{T}^n} e^{ik \cdot x} e^{il \cdot A^n x} dx = \int_{\mathbf{T}^n} e^{i(k + {}^t A^n l) \cdot x} dx = 0 \quad \text{si } {}^t A^n l \neq -k.$$

Si cette quantité ne tend pas vers 0 quand n tend vers l'infini, on peut trouver des entiers distincts n_1, n_2 tels que ${}^t A^{n_1} l = -k = {}^t A^{n_2} l$. On aurait donc ${}^t A^{n_2 - n_1} l = l$. Comme A n'a pas de valeur propre racine de l'unité, $l = 0$ et $k = 0$.

Remarque

L'application induite par la matrice $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ sur le tore $\mathbf{T}^2$ est parfois appelée *application du chat d'Arnold*, en référence à une figure qui se trouve dans le livre d'Arnold, Avez [A67], qui montre l'effet de cette transformation sur l'image d'un chat, cf fig.3.

figure 1 : Mélange

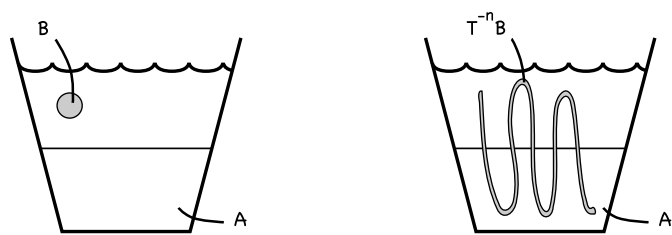


figure 2 : Multiplication par 2

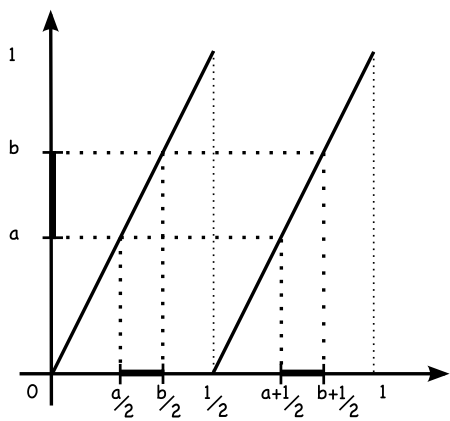
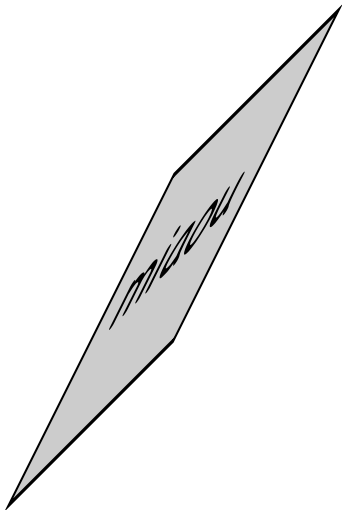


figure 3 : Automorphisme du tore

on applique la matrice
 $\begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$
au carré unité.

Voici le résultat
modulo 1.



Exercices

exercice 1 :

Montrez qu'une rotation sur le cercle S^1 n'est pas mélangeante relativement à la mesure de Lebesgue ; on pourra se servir des exponentielles complexes.

exercice 2 :

Montrez que si T est une transformation mélangeante, $T \circ T$ est aussi mélangeante.

exercice 3 :

Soit $(X, \mathcal{T}, \mu)$ un espace probabilisé et $T : X \rightarrow X$ une application mesurable qui préserve la mesure μ . Montrez que T est mélangeante si et seulement si pour tout $A \subset X$ mesurable, $\mu(A \cap T^{-n}A)$ converge vers $\mu(A)^2$. (*Indic : On peut s'intéresser au sous-espace engendré par les fonctions $\mathbf{1}_A \circ T^n$, $n \in \mathbf{N}$*)

exercice 4 :

Montrez que l'application $F : [0, 1] \rightarrow [0, 1]$ donnée par :

$$\begin{aligned} F(x) &= 2x && \text{si } x \in [0, 1/2] \\ &= 2 - 2x && \text{si } x \in]1/2, 1] \end{aligned}$$

préserve la mesure de Lebesgue et est mélangeante.

exercice 5 :

Soit $(X, \mathcal{T}, \mu)$ un espace probabilisé, $T : X \rightarrow X$ mesurable qui préserve μ .

On suppose que T est mélangeante. Soit k_i une suite d'entiers strictement croissante ; montrez que pour tout $f \in L^2$,

$$\frac{1}{n} \sum_{i=1}^n f \circ T^{k_i} \xrightarrow[n \rightarrow \infty]{L^2} \int f d\mu$$

Réciproquement, montrez que si cette convergence a lieu pour tout $f \in L^2$ et toute suite k_i d'entiers strictement croissante, alors T est mélangeante. On pourra utiliser la caractérisation du mélange par le biais de la topologie faible ainsi que le lemme de Banach-Saks.

exercice 6 :

Soit $(X, \mathcal{T}, \mu)$ un espace probabilisé, $T : X \rightarrow X$ une transformation mesurable qui préserve μ . On suppose T mélangeante. Montrez que pour tout $A \subset X$ de mesure non nulle, pour toute suite $n_i \rightarrow \infty$, $\bigcup_{i \in \mathbf{N}} T^{-n_i} A = X \text{ mod } 0$.

exercice 7 :

Montrez qu'un homéomorphisme de $[0, 1]$ n'est jamais topologiquement mélangeant.

exercice 8 :

Peut-on trouver un exemple de matrice 3×3 à coefficients entiers, de déterminant 1, qui n'est pas hyperbolique, mais dont l'action sur le tore $\mathbf{T}^3$ est mélangeante ?

exercice 9 :

Donnez un exemple de matrice 4×4 à coefficients entiers, de déterminant un, qui n'est pas hyperbolique, et telle que l'application induite sur le tore $\mathbf{T}^4$ est mélangeante.

exercice 10 :

Soit $(X, \mathcal{T}, \mu)$ un espace probabilisé, $T : X \rightarrow X$ mesurable qui préserve μ . Montrez que si T est mélangeante, il n'existe pas de fonction $f : X \rightarrow \mathbf{C}$ mesurable non constante et de nombre complexe λ de module 1 qui satisfont,

$$\text{pour presque tout } x \in X, \quad f(T(x)) = \lambda f(x).$$

Commentaires

Étant donné une mesure de probabilité arbitraire μ , Il est possible que la suite $\frac{1}{n} \sum T_*^k(hd\mu)$ converge pour toute fonction h positive bornée, vers une limite qui est singulière relativement aux mesures $hd\mu$. L'exemple le plus simple est donné par $T(x) = \frac{1}{2}x$ sur $[0, 1]$ et $\mu = \lambda_{[0,1]}$. On a alors $T_*^n(h\lambda_{[0,1]}) \rightarrow \delta_0$.

Dans quelle mesure le mélange est-il une propriété typique des systèmes dynamiques ? Pour répondre à cette question, définissons une distance sur l'ensemble des transformations boréliennes inversibles de $[0, 1]^d$ ou $\mathbf{T}^d$ préservant la mesure de Lebesgue. Soit I_n une suite de rectangles séparant les points.

$$d(T, T') = \sum \frac{1}{2^n} \left(\lambda(T(I_n)\Delta T'(I_n)) + \lambda(T^{-1}(I_n)\Delta T'^{-1}(I_n)) \right)$$

Relativement à cette topologie, l'ensemble des transformations mélangeantes est maigre (i.e. union dénombrable de fermés d'intérieur vide). Ce résultat est encore vrai pour la topologie C^0 sur l'espace des homéomorphismes, mais faux si on considère l'espace des difféomorphismes C^2 , définis sur $\mathbf{T}^d$ et préservant la mesure de Lebesgue. En topologie C^2 , tout difféomorphisme proche d'un automorphisme hyperbolique est mélangeant.

La notion de mélange est plus difficile à définir lorsque la mesure est infinie. U. Krengel et L. Sucheston démontrent en 1969 que sur un espace mesuré σ -fini, infini, il n'existe pas de transformation inversible préservant la mesure et satisfaisant :

$$\forall g \in L^\infty, \forall f \in L^1 \text{ telle que } \int f d\mu = 0, \quad \int f \circ T^n g d\mu \xrightarrow{n \rightarrow \infty} 0.$$

Le mélange des endomorphismes hyperboliques des tores peut se démontrer de plusieurs façons.

– La preuve donnée plus haut, à l'aide des séries de Fourier, se généralise aux automorphismes transitifs des groupes compacts abéliens. De manière générale, les techniques issues de l'analyse harmonique fonctionnent bien dans un cadre algébrique.

– Il est possible de coder ces applications à l'aide d'un système symbolique ; la preuve du mélange procède alors comme avec un décalage. Le codage le plus simple est donné par la décomposition en base dix. Cette décomposition conjugue la multiplication par dix sur $\mathbf{R}/\mathbf{Z}$ au décalage sur l'alphabet $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$.

– Le mélange peut être déduit de la densité, dans le tore, de la projection des sous-espaces stables de la matrice, densité qui peut s'obtenir en utilisant l'ergodicité des translations irrationnelles sur le tore. Il suffit ensuite de calculer explicitement l'image de rectangles dont les faces sont parallèles aux directions propres de la matrice.

– Une autre méthode consiste à montrer qu'une valeur d'adhérence de la suite $f \circ T^n$ est constante le long des sous-espaces stables et instables de la matrice. Cet argument se généralise à des systèmes de nature géométrique.

Une transformation préservant une mesure de probabilité est dite mélangeante d'ordre 3 si elle vérifie la propriété suivante :

$$\forall A, B, C \subset X \text{ mesurables, } \mu(A \cap T^{-n_1} B \cap T^{-n_1-n_2} C) \xrightarrow{n_1, n_2 \rightarrow \infty} \mu(A) \mu(B) \mu(C).$$

Existe-t-il des transformations mélangeantes qui ne sont pas mélangeantes d'ordre 3 ? Cette question, posée par V. Rokhlin en 1949, est toujours ouverte à l'heure actuelle. B. Host (1991) démontre qu'une transformation mélangeante dont le spectre est singulier est mélangeante de tout ordre.

L'argument de Hopf

The author has had complaints about too much detail missing in the presentation of the material in the latter paper. This has been rectified in the present paper.

E. Hopf (1902-1983)

Le théorème ergodique est démontré par G. D. Birkhoff en 1932. A cette époque, on connaît déjà des exemples de systèmes ergodiques. Ils sont issus de la théorie des probabilités et modélisent des phénomènes aléatoires comme le lancer d'un dé ou le tirage de boules dans une urne. On n'est donc pas étonné de voir surgir l'ergodicité dans ce contexte.

La question qui se pose alors est de savoir si cette ergodicité se rencontre en mécanique classique. L'attention se porte sur les flots géodésiques en courbure négative. Jacques Hadamard avait en effet démontré dès 1898 que ces flots sont instables du point de vue topologique.

Expliquons brièvement comment sont définis ces systèmes dynamiques : considérons une surface, dont la forme est en “selle de cheval” au voisinage de chacun de ses points. La variété d'équation $\{(x, y, z) \in \mathbf{T}^3 \mid \cos(x) + \cos(y) + \cos(z) = 0\}$ est un exemple de surface plongée dans le tore $\mathbf{T}^3$ pour laquelle la courbure est négative hormis en huit points ; cf fig.2. Cette surface intervient dans l'étude d'un système physique composé de trois pendules doubles reliés par leurs extrémités. Le flot géodésique agit sur l'ensemble des vecteurs de norme un tangents à la surface, en translatant ces vecteurs le long des géodésiques. Il préserve le volume canonique défini sur l'ensemble des vecteurs unitaires.

Gustav Hedlund est le premier à donner un exemple de surface à courbure négative, pour laquelle le flot géodésique est ergodique relativement au volume. Poursuivant les travaux de Jacques Hadamard, il montre en 1934 que, sur certaines surfaces, le flot géodésique est semi-conjugué à un système symbolique, ce qui lui permet de se ramener à une situation bien connue.

En 1936, Eberhard Hopf propose un argument de nature géométrique, qui lui permet de démontrer l'ergodicité du flot géodésique sur toutes les surfaces de volume fini et de courbure négative. Cet argument s'avère délicat à mettre en oeuvre en dimension supérieure. Mais il semble s'appliquer à une classe plus large de systèmes dynamiques : s'il existe suffisamment de directions dilatées et contractées par la transformation, alors il y a bon espoir de montrer l'ergodicité du système à l'aide de l'argument de Hopf.

Le flot géodésique sur les surfaces de courbure négative constante peut se décrire en termes algébriques. L'espace sur lequel est défini le système s'identifie à $PSL_2(\mathbf{R})$, quotient de l'ensemble des matrices 2×2 de déterminant 1 par le sous-groupe $\{id, -id\}$. Le flot est donné par la famille de transformations suivante :

$$\forall t \in \mathbf{R}, \quad \varphi_t\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix}\right) = \begin{pmatrix} e^t & 0 \\ 0 & e^{-t} \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

Ces transformations préservent la mesure $d\mu = da db dc / |a|$. On peut démontrer que toutes les surfaces orientables complètes connexes de volume fini, à courbure constante négative, s'identifient à un quotient de $PSL_2(\mathbf{R})$, par le biais d'un isomorphisme qui envoie le volume sur la mesure μ et le flot géodésique sur le flot $\{\varphi_t\}_{t \in \mathbf{R}}$.

Un flot $\{\varphi_t\}_{t \in \mathbf{R}}$ est *ergodique* relativement à une mesure invariante μ si les seuls ensembles mesurables invariants par toutes les transformations φ_t , $t \in \mathbf{R}$, sont de mesure nulle ou de complémentaire de mesure nulle. On va démontrer que le flot $\{\varphi_t\}_{t \in \mathbf{R}}$ défini plus haut est ergodique sur tous les quotients de $PSL_2(\mathbf{R})$ de mesure finie.

L'argument de Hopf

Soit X un espace métrique, $T : X \rightarrow X$ une application et $x \in X$. La feuille stable de x associée à l'application T est définie par :

$$W^{ss}(x) = \{ y \in X \mid d(T^n(x), T^n(y)) \xrightarrow{n \rightarrow \infty} 0 \}$$

Les feuilles stables partitionnent l'espace X . Si T est bijective, on peut aussi définir la feuille instable $W^{su}(x)$ de x : il s'agit de la feuille stable de x associée à T^{-1} .

Soit μ une mesure invariante par T . Une fonction mesurable $f : X \rightarrow \mathbf{R}$ est dite W^{ss} -invariante si, après restriction à un ensemble X_0 satisfaisant $\mu(X_0^c) = 0$, elle est constante sur les feuilles stables : $\forall x, y \in X_0, y \in W^{ss}(x)$ entraîne $f(x) = f(y)$.

Théorème

Soit X un espace métrique, μ une mesure borélienne finie sur X , $T : X \rightarrow X$ une application mesurable qui préserve μ . Soit $f \in L^2(X)$; alors les valeurs d'adhérence faibles de la suite $f \circ T^n$ sont W^{ss} -invariantes. Si de plus T est inversible, ces valeurs d'adhérence sont aussi W^{su} -invariantes.

Preuve

Soient n_i et g tels que $f \circ T^{n_i} \rightharpoonup g$. Supposons d'abord f lipschitzienne bornée. Le lemme de Banach-Saks (cf annexes) donne des sous-suites m_ℓ, n_{i_k} telles que :

$$\Psi_\ell(x) := \frac{1}{m_\ell} \sum_{k=1}^{m_\ell} f \circ T^{n_{i_k}}(x) \xrightarrow{\ell \rightarrow \infty} g(x) \quad p.p.$$

$$\text{Si } y \in W^{ss}(x), \quad |\Psi_\ell(x) - \Psi_\ell(y)| \leq C \frac{1}{m_\ell} \sum_{k=1}^{m_\ell} d(T^{n_{i_k}}(x), T^{n_{i_k}}(y)) \xrightarrow{\ell \rightarrow \infty} 0.$$

Par conséquent, la fonction g est W^{ss} -invariante.

Soit $f \in L^2$. Pour tout $\varepsilon > 0$, on peut trouver f' lipschitzienne telle que $\|f - f'\| < \varepsilon$. Quitte à extraire, on peut supposer que $f' \circ T^{n_k}$ converge faiblement vers une fonction g' , qui est W^{ss} -invariante. On a donc : $(f - f') \circ T^{n_k} \rightharpoonup g - g'$ ce qui implique :

$$\|g - g'\| \leq \liminf \|(f - f') \circ T^{n_k}\| \leq \|f - f'\| < \varepsilon.$$

On peut donc trouver une suite de fonctions W^{ss} -invariantes qui converge vers g en norme L^2 et, après extraction, presque partout. La fonction g est W^{ss} -invariante.

Passons au cas T inversible. Soit I le sous-espace des fonctions W^{su} -invariantes. Montrons que si f appartient à $I^\perp$, alors $f \circ T^n$ converge faiblement vers zéro. Soit g une limite faible de $f \circ T^{n_i}$. Appliquons ce qui précède à T^{-1} ; on peut trouver une sous-suite n_{i_k} et une fonction $g_0 \in I$ telle que $g \circ T^{-n_{i_k}} \rightharpoonup g_0$. On obtient :

$$\langle g, g \rangle = \lim_{k \rightarrow +\infty} \langle f \circ T^{n_{i_k}}, g \rangle = \lim_{k \rightarrow +\infty} \langle f, g \circ T^{-n_{i_k}} \rangle = \langle f, g_0 \rangle = 0.$$

Toute fonction $f \in L^2$ peut s'écrire comme une somme $f = f_1 + f_2$ avec $f_1 \in I$ et $f_2 \in I^\perp$. La suite $f_2 \circ T^n$ tend faiblement vers 0. Les valeurs d'adhérence de $f \circ T^n$ sont donc aussi des valeurs d'adhérence de la suite $f_1 \circ T^n$, qui appartient à I .

Si f est une fonction invariante, $f \circ T^n = f$ et on obtient le corollaire suivant :

L'argument de Hopf

Soit X un espace métrique, μ une mesure borélienne finie, $T : X \rightarrow X$ une application mesurable qui préserve μ . Alors toute fonction $f \in L^2(X)$ invariante par T est W^{ss} -invariante. Si de plus T est inversible, f est aussi W^{su} -invariante.

En termes ensemblistes, cela revient à dire que tout ensemble mesurable invariant par T coïncide, à un ensemble négligeable près, avec une union de feuilles stables.

Appliquons cet argument aux automorphismes hyperboliques du tore $\mathbf{T}^n$. Considérons une matrice sans valeurs propres de module un. Notons E_s la projection sur le tore du sous-espace vectoriel associé aux valeurs propres de module inférieur à un. Soit E_u la projection sur $\mathbf{T}^n$ du sous-espace associé aux valeurs propres de module supérieur à un ; cf fig. 1. Les feuilles stables et instables de l'application induite par la matrice sur le tore sont données par $W^{ss}(x) = x + E_s$, $W^{su}(x) = x + E_u$.

On se place dans un système de coordonnées dirigé selon E_s et E_u , ce qui donne une carte $(x, y) \in U$ définie dans un voisinage U d'un point quelconque du tore. Dans cette carte, les feuilles stables sont horizontales, les feuilles instables verticales, et la mesure de Lebesgue prend la forme $dx dy$. Soit $f \in L^2$ une fonction invariante par l'automorphisme. D'après l'argument de Hopf, elle est invariante par W^{ss} et W^{su} . Dans le système de coordonnées (x, y) , f ne dépend donc (presque) pas de x et de y . Le lemme suivant montre qu'elle est constante presque partout sur U .

Lemme

Soit $(X, \mathcal{T}, \mu)$, $(Y, \mathcal{S}, \nu)$ deux espaces probabilisés et $f : X \times Y \rightarrow \mathbf{R}$ une fonction L^2 . On suppose qu'il existe $\varphi_1 : X \rightarrow \mathbf{R}$ et $\varphi_2 : Y \rightarrow \mathbf{R}$ deux fonctions mesurables, $Z \subset X \times Y$ un sous-ensemble de $\mu \otimes \nu$ -mesure totale, tels que :

$$\forall (x, y) \in Z, \quad f(x, y) = \varphi_1(x), \quad f(x, y) = \varphi_2(y).$$

Alors f est constante presque partout.

Preuve du lemme

D'après le théorème de Fubini, il existe $Y_0 \subset Y$ de mesure totale et $x_0 \in X$ tels que $\{x_0\} \times Y_0 \subset Z$. Pour tout $(x, y) \in Z \cap (X \times Y_0)$, le point (x_0, y) est dans Z , ce qui implique : $\varphi_1(x_0) = \varphi_2(y) = f(x, y)$. Le lemme est démontré.

On en déduit que f est localement presque constante. Le lemme suivant (cf ex. 2) montre que f est constante presque partout, i.e. l'automorphisme est ergodique.

Lemme

Soit X un espace métrique, μ une mesure dont le support est connexe et f une fonction localement presque constante. Alors f est constante pp sur $\text{supp } \mu$.

Le point clef dans cette preuve est l'existence d'un système de coordonnées dans lequel les feuilles stables et instables s'identifient aux horizontales et aux verticales, et tel que la mesure invariante est équivalente à une mesure produit. La mesure est dite *absolument continue* relativement aux feuilletages stables et instables.

Flots sur les quotients de $PSL_2(\mathbf{R})$

Soit $X = PSL_2(\mathbf{R})$ l'ensemble des matrices 2×2 de déterminant 1, au signe près. Cet espace est homéomorphe à $\mathbf{R}^2 \times \mathbf{S}^1$. Soit d une distance sur X invariante à droite ; par exemple, $d(A, B) = \log(\|AB^{-1}\| \|BA^{-1}\|)$, pour une norme $\|\cdot\|$ bien choisie. On définit trois familles de transformations en posant, pour $t \in \mathbf{R}$ et $M \in X$:

$$\varphi_t(M) = \begin{pmatrix} e^t & 0 \\ 0 & e^{-t} \end{pmatrix} M, \quad h_t^{su}(M) = \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix} M, \quad h_t^{ss}(M) = \begin{pmatrix} 1 & 0 \\ t & 1 \end{pmatrix} M.$$

L'égalité $\varphi_t \circ h_s^{ss} = h_{se^{-2t}}^{ss} \circ \varphi_t$ montre que pour tout $s \in \mathbf{R}$, les points $h_s^{ss}(M)$ appartiennent à la feuille stable de φ_t passant par M . Un calcul similaire montre que $h_s^{su}(M)$ appartient à la feuille instable de φ_t . Soit $u' = e^{2t} u (1 - e^{2t} su)^{-1}$; dans les coordonnées $(t, s, u) \mapsto h_u^{su} h_s^{ss} \varphi_t(id)$, les variétés $W^{su}(M)$ sont des droites verticales, et les $W^{ss}(\varphi_t(M))$, $t \in \mathbf{R}$, sont des plans horizontaux ; cf ex. 8 et fig. 3.

On vérifie que la mesure $d\mu(\begin{pmatrix} a & b \\ c & d \end{pmatrix}) = \frac{da db dc}{|a|}$ sur X est invariante par multiplication à gauche et à droite. L'espace X est de mesure infinie, mais admet des quotients à droite, de volume fini. Considérons un de ces quotients X/Γ ; les flots, la distance et la mesure passent au quotient. On peut donc appliquer l'argument de Hopf sur X/Γ : toute fonction L^2 invariante par les transformations φ_t , $t \in \mathbf{R}$, est constante presque partout. Le flot $\{\varphi_t\}_{t \in \mathbf{R}}$ est ergodique relativement à μ .

figure 1 : Feuilles stables et instables de l'automorphisme du tore $\begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$

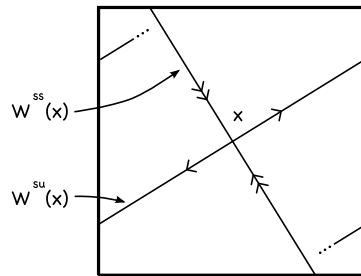


figure 2 : Une variété à courbure négative

Surface plongée dans le tore $\mathbf{T}^3$
 d'équation :
 $\{ (x,y,z) \mid \cos(x)+\cos(y)+\cos(z) = 0 \}$
 La courbure s'annule en huit points.

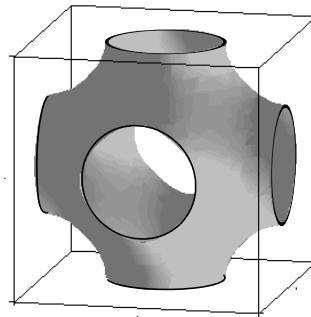
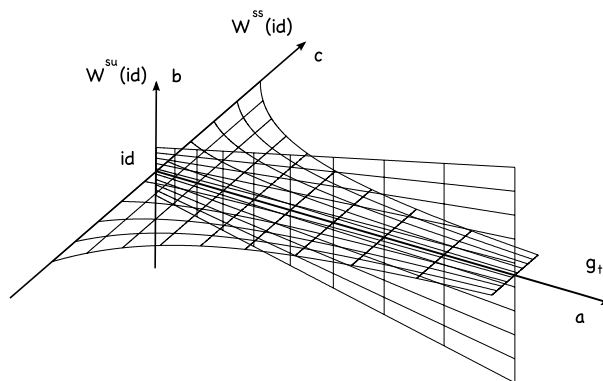


figure 3 : Flot géodésique en courbure négative constante

Orbite du flot g_t dans les
 coordonnées $(a,b,c) \rightarrow \begin{pmatrix} a & b \\ c & d \end{pmatrix}$
 $g_t h_s^{ss}(1,0,0) = (e^t, 0, se^{-t})$
 $g_t h_s^{su}(1,0,0) = (e^t, se^t, 0)$



Exercices

Dans la suite, X est un espace métrique, μ est une mesure borélienne finie définie sur X et T est une transformation mesurable qui préserve μ .

exercice 1 :

Montrez que toute fonction f mesurable invariante par T est W^{ss} -invariante. Montrez qu'une fonction mesurable $f : X \rightarrow \mathbf{R}$ est W^{ss} -invariante si et seulement si elle coïncide presque partout avec une fonction constante sur toutes les feuilles stables.

exercice 2 :

Soit f une fonction presque localement constante (tout point admet un voisinage sur lequel f est constante presque partout). On pose, pour tout $x \in \text{supp}(\mu)$,

$$\bar{f}(x) = \lim_{r \rightarrow 0} \frac{1}{\mu(B(x, r))} \int_{B(x, r)} f d\mu.$$

- Montrez que $\bar{f}$ est localement constante.
- Montrez que $\mu\{x \in \text{supp } \mu, f(x) \neq \bar{f}(x)\} = 0$ (Rq : $\text{supp } \mu$ est séparable, cf annexes).
- En déduire que f est constante μ -pp sur le support de μ si celui-ci est connexe.

exercice 3 :

L'enveloppe convexe d'un ensemble $A \subset L^2(X)$ est notée par $\text{Conv}(A)$; il s'agit du plus petit convexe contenant A . Soit $f \in L^2(X)$. Montrez que tous les éléments du convexe suivant sont W^{ss} -invariants :

$$\bigcap_{N \in \mathbf{N}} \overline{\text{Conv}(\{f \circ T^n \mid n \geq N\})}$$

exercice 4 :

Soit $f_1, f_2 \in L^2(X)$. Supposons qu'il y ait deux suites n_1, n_2 tendant vers l'infini, telles que le produit $f_1 \circ T^{n_1} f_2 \circ T^{n_1+n_2}$ converge faiblement. Montrez que la limite est W^{ss} -invariante. Généralisez au cas d'un nombre fini de fonctions f_k .

exercice 5 :

Soient $(X, \mathcal{T}, \mu)$ et $(Y, \mathcal{S}, \nu)$ deux espaces mesurés σ -finis. Soit $B \subset X \times Y$ un ensemble $\mu \otimes \nu$ -mesurable. Pour $(x, y) \in X \times Y$, on pose $B_x = \{y \in Y \mid (x, y) \in B\}$ et $B^y = \{x \in X \mid (x, y) \in B\}$. Supposons que :

- pour μ -presque tout $x \in X$, $\nu(B_x) = 0$ ou $\nu(B_x^c) = 0$;
- pour ν -presque tout $y \in Y$, $\mu(B^y) = 0$ ou $\mu(B^y)^c = 0$.

Montrez que $\mu \otimes \nu(B) = 0$ ou $\mu \otimes \nu(B^c) = 0$.

exercice 6 :

Donnez un exemple de norme sur $\mathbf{R}^2$ pour laquelle les seules isométries linéaires préservant l'orientation sont $\pm id$. Soit $\|\cdot\|$ la norme induite sur l'ensemble des matrices 2×2 par cette norme. Montrez que l'expression :

$$d(A, B) = \log(\|AB^{-1}\|) + \log(\|BA^{-1}\|)$$

définit une distance sur $PSL_2(\mathbf{R})$, invariante par multiplication à droite.

Indic : prendre la norme L^1 dans le premier quadrant et la norme L^2 dans le second.

exercice 7 :

Posons $u' = e^{2t}u(1 - e^{2t}su)^{-1}$, $s' = s(1 - e^{2t}su)$, $t' = t - \ln(1 - e^{2t}su)$. Montrez que :

$$h_{u'}^{su} \circ h_s^{ss} \circ g_t = h_{s'}^{ss} \circ g_{t'} \circ h_u^{su}$$

Montrez que la transformation $(s, t, u) \mapsto h_{u'}^{su} \circ h_s^{ss} \circ g_t(M)$ est un difféomorphisme d'un voisinage de l'origine de $\mathbf{R}^3$ sur un voisinage de $M \in PSL_2(\mathbf{R})$. Soit M' un point de ce voisinage ; vérifiez que dans ce système de coordonnées, les variétés instables $W^{su}(M')$ correspondent aux droites verticales et que les variétés stables $\bigcup_{t \in \mathbf{R}} W^{ss}(\varphi_t(M'))$ correspondent aux plans horizontaux.

Commentaires

Le terme de feuilletage employé au sujet de la partition de X donnée par les feuilles $W^{ss}(x)$, $x \in X$, est bien sûr impropre ; en toute généralité, il ne faut pas s'attendre à ce que cette partition forme un feuilletage au sens géométrique du terme. Le terme de “distributions stables” est parfois utilisé dans la littérature mais il entre en conflit avec le concept de distribution qui provient de l'analyse.

L'argument de Hopf est encore valide si on considère le feuilletage “moyen” :

$$W_{moy}^{ss}(x) = \{y \in X \mid \frac{1}{n} \sum_{k=1}^n d(T^k(x), T^k(y)) \xrightarrow{n \rightarrow \infty} 0\}$$

Les fonctions invariantes par T sont W_{moy}^{ss} -invariantes. Plus généralement, les fonctions propres de $f \circ T^n$, $f \in L^2$, sont W_{moy}^{ss} -invariantes. Par contre les valeurs d'adhérence de $f \circ T^n$ ne sont en général pas invariantes par ce feuilletage.

Les théorèmes précédents se généralisent en partie au cas de la mesure infinie. Si on peut trouver une famille dénombrable d'ouverts U_i de mesure finie telle que $\mu(X \setminus \bigcup U_i) = 0$, alors les valeurs d'adhérence de $f \circ T^n$, $f \in L^2$, sont W^{ss} -invariantes. La preuve est inchangée, la condition sur la mesure assurant la densité dans $L^2(X)$ des fonctions Lipschitziennes L^2 bornées.

Cependant, en mesure infinie, il ne suffit pas de montrer que les fonctions L^2 invariantes par T sont constantes presque partout, pour obtenir l'ergodicité. Par exemple, pour une translation sur $\mathbf{R}$, il n'y a pas d'ensembles invariants de mesure de Lebesgue finie non nulle, et donc pas de fonction L^2 invariante ; mais il y a de nombreux ensembles invariants de mesure infinie dont le complémentaire est de mesure infinie, et beaucoup de fonctions invariantes bornées.

En mesure infinie, l'argument de Hopf reste vrai si on suppose la mesure conservatrice, c'est-à-dire si tout ensemble de mesure non nulle a une intersection de mesure non nulle avec un de ses itérés. La preuve est basée sur le théorème ergodique “quotient”, démontré par E. Hopf en 1937 dans le but d'étendre l'argument au cas de la mesure infinie.

L'argument original de E. Hopf faisait appel au théorème ergodique de Birkhoff, plutôt qu'au lemme de Banach-Saks, et n'utilisait pas de topologie faible. De ce point de vue, il ne permettait pas d'attaquer la question du mélange fort de la transformation.

Le mélange faible reste accessible par l'argument original, en utilisant les sommes pondérées $\sum e^{ik\theta} f \circ T^k$ en lieu et place des sommes de Birkhoff. A partir des années 60, la question du mélange est étudiée par des techniques entropiques. La tribu des ensembles invariants est remplacée par la tribu de Pinsker, composée des ensembles appartenant à une partition d'entropie nulle, et l'argument de Hopf est démontré par ordre de généralisation croissante, dans une série de travaux qui débute avec D.V. Anosov et Y. Sinai en 1967 et se termine avec F. Ledrappier et L. S. Young en 1984. Sous leur forme la plus générale, ces résultats montrent l'équivalence entre la tribu de Pinsker et la tribu des ensembles mesurables union de feuilles stables “rapides” $W_{vite}^{ss}(x) = \{y \in X \mid \lim \frac{1}{n} \log d(T^n(x), T^n(y)) < 0\}$, pour tout difféomorphisme C^2 sur une variété compacte. Le lien avec le mélange se fait par le biais de la remarque suivante : les valeurs d'adhérence des suites de la forme $f \circ T^n$ sont mesurables par rapport à la tribu de Pinsker.

Il est en général difficile de démontrer que la mesure de Lebesgue est absolument continue relativement au feuilletage stable. Pour un flot géodésique sur une variété compacte à courbure strictement négative, l'absolue continuité du volume est démontrée par D. V. Anosov en 1963. Le cas de la courbure négative ou nulle n'est pas si bien compris, la question de l'ergodicité du volume reste ouverte à ce jour.

Il existe quelques exemples de surfaces à courbure négative ou nulle pour lesquelles on sait démontrer l'absolue continuité du volume relativement aux feuilletages stables et instables. Par exemple, s'il existe un point de courbure négative sur chaque géodésique, alors le flot géodésique est Anosov (Eberlein, 1973) ce qui entraîne l'absolue continuité, et donc l'ergodicité. C'est le cas pour la surface d'équation $\{(x, y, z) \in \mathbf{T}^3 \mid \cos(x) + \cos(y) + \cos(z) = 0\}$ car sa courbure ne s'annule qu'en un nombre fini de points.

La surface $\{(x, y, z) \in \mathbf{T}^3 \mid \cos(x) + \cos(y) + \cos(z) = 0\}$ intervient dans les travaux de T. J. Hunt et R. S. Mackay (2003), qui donnent un exemple de système mécanique qui se ramène à l'étude du flot géodésique sur cette surface. Le principe de Maupertuis constitue une autre motivation physique à l'étude des flots géodésiques. Ce principe affirme qu'à haute énergie, un système hamiltonien se comporte comme un flot géodésique associé à une certaine métrique sur l'espace des phases. Cette métrique est cependant rarement à courbure négative.

L'identification des surfaces orientables complètes connexes et de courbure -1 à des quotients de $PSL_2(\mathbf{R})$ découle du théorème de Hadamard : ce théorème affirme que l'application exponentielle, définie de l'espace tangent en un point de la surface sur la surface elle-même, est un revêtement ; la métrique se calcule explicitement dans ces coordonnées exponentielles : $ds^2 = dr^2 + sh^2(r) d\theta^2$. Ceci permet d'identifier le revêtement universel de la surface au demi-plan de Poincaré $\mathbf{H} = \{z \in \mathbf{C} \mid \operatorname{Re}(z) > 0\}$ muni de la métrique $\frac{|dz|^2}{\operatorname{Re}(z)}$. Pour cette métrique, les isométries préservant l'orientation sont les homographies $z \mapsto \frac{az+b}{cz+d}$, $a, b, c, d \in \mathbf{R}$, $ad - bc = 1$. Enfin, deux vecteurs unitaires de $T\mathbf{H}$ se déduisent l'un de l'autre par le biais d'une unique homographie. D'un point de vue algébrique, l'isomorphisme entre $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in PSL_2(\mathbf{R})$ et $(x + iy, \theta) \in T^1\mathbf{H}$ est donné par la décomposition d'Iwasawa de la matrice.

Il existe des méthodes algébriques pour construire des quotients de $PSL_2(\mathbf{R})$ de volume fini. On peut par exemple quotienter par $PSL_2(K)$, où K est une algèbre de quaternions sur un corps de nombres. Ces constructions sont décrites par S. Katok [Ka92]. Voici un exemple : soit $a, b \in \mathbf{N}$ deux nombres premiers avec a qui n'est pas un carré modulo b . Le quotient de $PSL_2(\mathbf{R})$ par le groupe suivant est compact :

$$\left\{ \begin{pmatrix} x_0 + x_1 \sqrt{a} & x_2 + x_3 \sqrt{a} \\ b(x_2 - x_3 \sqrt{a}) & x_0 - x_1 \sqrt{a} \end{pmatrix} \mid x_0, x_1, x_2, x_3 \in \mathbf{Z}, x_0^2 - a x_1^2 - b x_2^2 + a b x_3^2 = 1 \right\}.$$

Les constructions les plus générales sont de nature géométrique et passent par l'identification de $PSL_2(\mathbf{R})$ au fibré unitaire du demi-plan de Poincaré $\mathbf{H}$.

Dynamique topologique

Toutes les pensées vraiment sages ont déjà été pensées des milliers de fois ; mais pour les faire vraiment nôtres, nous devons les penser encore, honnêtement, jusqu'à ce qu'elles prennent racine dans notre expérience personnelle.

J. W. von Goethe (1749-1832)

Un système dynamique topologique est donné par un espace topologique X et une application $T : X \rightarrow X$. On s'intéressera au cas où X est un espace métrique, et où T vérifie certaines conditions de compatibilité avec la topologie, par exemple T continue ou borélienne. La suite $x, T(x), T(T(x)), \dots$ des itérés d'un point $x \in X$ constitue la *trajectoire*, ou encore l'*orbite* du point x .

Voici plusieurs exemples de systèmes dynamiques topologiques :

- *les systèmes issus de la physique* : la mécanique du point matériel peut être décrite à l'aide de la transformation qui associe aux conditions initiales $(x_0, v_0) \in \mathbf{R}^3 \times \mathbf{R}^3$ la position et la vitesse (x, v) du point au temps 1.
- *les systèmes issus de l'algorithmique* : lorsqu'une équation ne peut pas être résolue de manière explicite, on peut chercher à approcher ses solutions à l'aide d'une suite définie par récurrence $x_{n+1} = T(x_n)$.
- *les systèmes issus de la théorie des probabilités* : la répétition d'une épreuve de manière indépendante se décrit à l'aide d'un décalage défini sur un espace produit.
- *les systèmes issus de la géométrie* : l'étude d'une équation d'évolution (flot géodésique, flot de Ricci...) donne des informations sur la structure de l'espace sous-jacent.
- *les systèmes issus de l'arithmétique* : un des premiers exemples est donné par Gauss, qui remarque que le calcul du développement en fractions continues peut se faire à l'aide d'une transformation qui préserve une mesure naturelle.
- *les systèmes issus de la théorie des groupes* : on peut regarder des actions linéaires sur des quotients de groupes de matrices, ou même faire agir un élément du groupe par translation sur le quotient. Le système obtenu possède en général une dynamique non triviale.

Soit x un point de X . Que peut-on dire du comportement de la suite $\{T^n(x)\}_{n \in \mathbf{N}}$ en général ? Il est possible que cette suite soit convergente. C'est le comportement souhaité lorsque le système doit permettre de calculer les solutions d'une équation. A l'opposé, la trajectoire peut être dense dans l'espace X . C'est presque toujours le cas lorsque la transformation est ergodique relativement à une mesure finie de support total.

En général, ces deux comportements peuvent coexister au sein d'un même système. Afin de les distinguer, on introduit la notion d'*ensemble non-errant*. Les points qui n'appartiennent pas à cet ensemble, sont ceux qui possèdent un voisinage distinct de tous ses itérés. Leur trajectoire ne peut pas être dense ; elle ne peut pas non plus revenir proche de son point de départ. La trajectoire d'un point arbitraire ne peut pas s'accumuler sur un point errant. Elle doit donc partir à l'infini, ou terminer dans l'ensemble non-errant.

On va faire usage dans ce chapitre du théorème de Baire. Ce résultat joue un rôle important en dynamique topologique. Il est valide dans tout espace *topologiquement complet*, c'est-à-dire dans tout espace topologique homéomorphe à un espace métrique complet, et s'énonce comme suit : dans un tel espace, toute intersection dénombrable d'ouverts denses est dense. Les espaces métriques complets ou localement compacts sont des exemples d'espaces topologiquement complets, et c'est souvent dans ce cadre qu'est énoncé le théorème de Baire.

Dynamique topologique

Définition

Soit X un espace métrique et $T : X \rightarrow X$ une application. Cette application est dite transitive si pour tout ouverts non vides $U, V \subset X$, on peut trouver une suite $n_i \rightarrow \infty$ telle que $T^{-n_i}U \cap V$ est non vide ; cf fig.1.

T est topologiquement mélangeante si pour tout ouverts non vides $U, V \subset X$, on peut trouver $N \in \mathbf{N}$ tel que pour tout $n \geq N$, $T^{-n}U \cap V$ est non vide.

Une application topologiquement mélangeante est transitive. Le lien avec l'ergodicité et le mélange en mesure découle des définitions :

Théorème

Soit X un espace métrique ; une transformation borélienne de X qui préserve une mesure borélienne finie ergodique de support total est transitive. Si la mesure est mélangeante, T est topologiquement mélangeante.

Si T est ergodique, on peut montrer l'existence d'un point dont l'orbite est dense :

Définition

Soit X un espace métrique et $T : X \rightarrow X$ une application. L'ensemble ω -limite de x est l'ensemble de toutes les valeurs d'adhérence de la suite $\{T^n x\}_{n \in \mathbf{N}}$:

$$\omega(x) = \{y \in X \mid \exists n_i \rightarrow \infty, T^{n_i}x \rightarrow y\} = \bigcap_{n \in \mathbf{N}} \overline{\{T^k x \mid k \geq n\}}$$

Proposition

Soit X un espace métrique, μ une mesure borélienne finie et $T : X \rightarrow X$ une application borélienne qui préserve μ . On suppose μ ergodique. Alors, pour presque tout $x \in X$, $\text{supp } \mu \subset \omega(x)$.

Preuve

Soit $\{x_i\}_{i \in \mathbf{N}}$ une partie dénombrable de $\text{supp } \mu$, dense dans $\text{supp } \mu$, et $r \in \mathbf{Q}$ positif. Les ensembles $B(x_i, r)$ sont de mesure positive. Par ergodicité, pour presque tout $x \in X$, pour tout $i \in \mathbf{N}$ et $r \in \mathbf{Q}$ positif, on peut trouver une infinité de n tels que $T^n(x) \in B(x_i, r)$. Pour ces x , on a $x_i \in \omega(x)$, $\forall i \in \mathbf{N}$, et donc $\text{supp } \mu \subset \omega(x)$.

Corollaire

Soit X un espace métrique et T une transformation borélienne de X qui préserve une mesure borélienne finie ergodique, de support total. Alors on peut trouver un point $x \in X$ tel que $\omega(x) = X$; cf fig.2.

Dans un espace topologique, un ensemble est dense si et seulement si il rencontre tous les ouverts. L'existence d'un point x tel que $\omega(x) = X$ implique donc la transitivité. Pour établir une réciproque, considérons l'égalité suivante :

$$\{x \in X \mid \omega(x) = X\} = \bigcap_{U \text{ ouvert}} \bigcap_{N \in \mathbf{N}} \bigcup_{n > N} T^{-n}U$$

On peut restreindre la première intersection aux ouverts appartenant à une base de la topologie ; si X est à base dénombrable, l'ensemble $\{x \in X \mid \omega(x) = X\}$ peut donc s'écrire comme une intersection dénombrable d'ouverts.

Théorème

Soit X un espace séparable topologiquement complet, $T : X \rightarrow X$ une application continue. Alors T est transitive si et seulement s'il existe $x \in X$ tel que $\omega(x) = X$.

Preuve

Si T est transitive, les ensembles $\bigcup_{n > N} T^{-n}U$ sont denses car ils rencontrent tous les ouverts. L'ensemble $\{x \in X \mid \omega(x) = X\}$ est une intersection dénombrable d'ouverts denses ; il est donc non vide par le théorème de Baire.

Définition de l'ensemble récurrent

Un point $x \in X$ est récurrent si $x \in \omega(x)$, c'est-à-dire si on peut trouver une suite $n_i \rightarrow \infty$ telle que $T^{n_i}x \rightarrow x$. L'ensemble des points récurrents sera noté $\mathcal{R}$.

On vient de voir que si T est continue, l'ensemble des $x \in X$ tels que $\omega(x) = X$ est une intersection d'ouverts. Il en va de même pour les points récurrents. Ceci provient de l'égalité suivante, valide dans tout espace métrique :

$$\{x \in X \mid x \in \omega(x)\} = \bigcap_{k \in \mathbf{N}^*} \bigcap_{N \in \mathbf{N}} \bigcup_{n \geq N} \{x \in X \mid d(x, T^n x) < \frac{1}{k}\}.$$

Le résultat suivant est à rapprocher du théorème de récurrence de Poincaré.

Théorème

Soit X un espace métrique, μ une mesure borélienne finie et $T : X \rightarrow X$ une application borélienne qui préserve μ . Alors, presque tout point du support de μ est récurrent.

Preuve

Soit $\{x_i\}_{i \in \mathbf{N}}$ une partie dénombrable de $\text{supp } \mu$, dense dans $\text{supp } \mu$ (cf annexe) ; soit $\mathcal{D}$ l'ensemble des boules centrées en les x_i et dont le rayon est rationnel. Pour tout $U \in \mathcal{D}$, l'ensemble $\{x \in X \mid T^k x \in U \text{ pour une infinité de } k \text{ si } x \in U\}$ est de mesure totale dans X , d'après le théorème de récurrence de Poincaré. L'ensemble suivant est donc de mesure totale dans X :

$$\{x \in X \mid \forall U \in \mathcal{D}, x \in U \text{ implique } T^k x \in U \text{ pour une infinité de } k\}$$

On termine en remarquant que pour tout $x \in \text{supp } \mu$, l'ensemble des $U \in \mathcal{D}$ tels que $x \in U$ forme une base de voisinages de x .

Définition de l'ensemble non-errant

Un point x est non-errant si pour tout ouvert U contenant x , on peut trouver une suite $n_i \rightarrow \infty$ telle que $T^{-n_i}(U) \cap U$ est non vide. L'ensemble des points non-errants est noté Ω .

L'ensemble non-errant contient tous les ensembles ω -limites : $\forall x \in X, \omega(x) \subset \Omega$. L'ensemble non-errant est fermé ; les points récurrents sont non-errants, on a donc :

Corollaire

Soit X un espace métrique, μ une mesure borélienne finie et $T : X \rightarrow X$ une application borélienne qui préserve μ . Alors $\text{supp } \mu \subset \overline{\mathcal{R}} \subset \Omega$.

Pour les transformations continues, on a le critère suivant :

Proposition

Soit T une application continue définie sur un espace métrique. Un point x est errant ($x \notin \Omega$) si et seulement si on peut trouver un ouvert U contenant x tel que $T^{-n}U \cap U$ est vide pour tout $n \geq 1$; cf fig.3.

Preuve

Si x est périodique de période $p > 0$, x est non errant car $x \in T^{-nk}U \cap U, \forall k \geq 0$. Soit $x \notin \Omega$. Il existe un ouvert U contenant x et $N \geq 1$ tel que pour tout $n > N$, $T^{-n}U \cap U = \emptyset$. Soit $r = \frac{1}{2} \min\{d(T^n x, x) \mid n \leq N\}$ et V l'intersection de U avec tous les $T^{-n}B(T^n x, r)$, $0 \leq n \leq N$. L'ensemble $T^{-i}V \cap V$ est vide pour tout $i \geq 1$, car inclus dans $T^{-i}U \cap U$ si $i > N$, et inclus dans $T^{-i}(B(x, r) \cap B(T^i x, r))$ si $i \leq N$.

Enfin, si T est une transformation inversible de X , l'ensemble non-errant de T^{-1} coïncide avec celui de T . Qui plus est, T^{-1} est transitif (resp. topologiquement mélangeant) si et seulement si T est transitif (resp. topologiquement mélangeant).

figure 1 : Transitivité



figure 2 : Transitivité des automorphismes hyperboliques du tore

On itère un point x du tore T^2 pris au hasard à l'aide de la transformation $\begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$

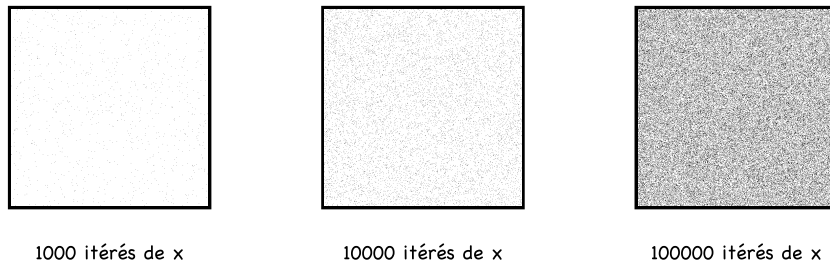
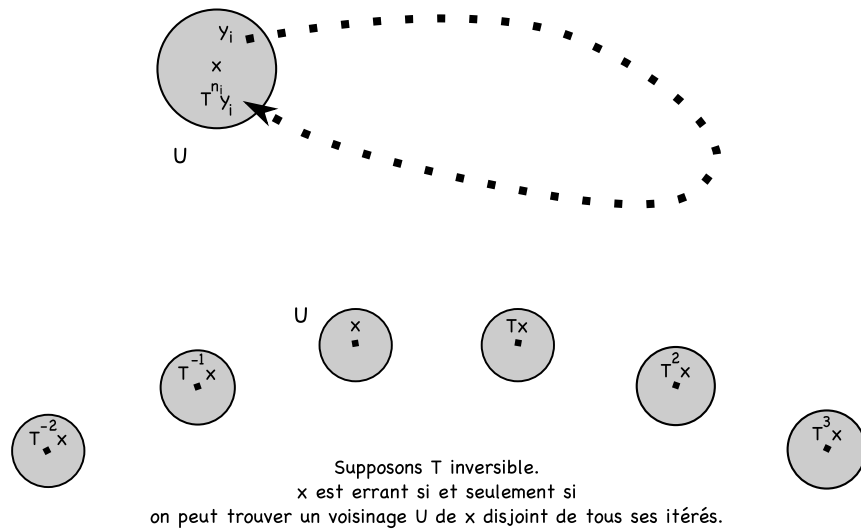


figure 3 : Ensemble non-errant



Exercices

exercice 1 :

Soit X un espace métrique, $T : X \rightarrow X$ une application continue et $x \in X$. Montrez que $T(\omega(x)) \subset \omega(x)$. Montrez que $T(\mathcal{R}) \subset \mathcal{R}$. Montrez que $T(\Omega) \subset \Omega$.

On suppose que X est compact. Montrez que $T(\omega(x)) = \omega(x)$.

exercice 2 :

Soit X un espace métrique sans point isolé, $T : X \rightarrow X$ une application et $x \in X$. Montrez que si $\bigcup_{n \in \mathbf{N}} T^n x$ est dense dans X , alors $\omega(x) = X$ et T est transitive.

exercice 3 :

Soit X un espace métrique, $T : X \rightarrow X$ une application. Soit $x, y \in X$. Montrez que si $d(T^n x, T^n y) \xrightarrow{n \rightarrow \infty} 0$, alors $\omega(x) = \omega(y)$. Montrez que $\omega(Tx) = \omega(x)$.

exercice 4 :

Donnez un exemple d'homéomorphisme de $\mathbf{R}^d$ dont l'ensemble non-errant est vide.

exercice 5 :

Soit X un espace métrique, μ une mesure borélienne finie sur X , $T : X \rightarrow X$ une application borélienne. On suppose qu'il existe $x \in X$ et $n_i \rightarrow \infty$ tels que :

$$\frac{1}{n_i} \sum_{k=0}^{n_i-1} \delta_{T^k x} \xrightarrow{i \rightarrow \infty} \mu \quad \text{étroitement}$$

Montrez que $\text{supp } \mu \subset \omega(x)$.

exercice 6 :

Construire une application holomorphe bijective du disque unité fermé dans lui-même qui possède un unique point fixe. Est-elle transitive, topologiquement mélangeante ? Quelles sont les probabilités boréliennes invariantes par cette application ? Sont-elles ergodiques ?

exercice 7 :

Soit X un espace métrique compact et $T : X \rightarrow X$ une application continue surjective. Soit U un ouvert de X , différent de X , tel que $T(U) \subset U$. Montrez que l'ensemble $\bigcup_{n \in \mathbf{N}} T^{-n}(U)$ n'est pas égal à X .

exercice 8 :

On considère le décalage $\sigma(\{x_i\}_{i \in \mathbf{N}}) = \{x_{i+1}\}_{i \in \mathbf{N}}$ défini sur l'espace $X = \{0, 1\}^{\mathbf{N}}$. Cet espace est muni de la topologie produit. On considère l'ensemble suivant : $F = \{\{x_i\} \mid \forall k \in \mathbf{N}, x_k = 1 \text{ implique } x_{k+1} = 0\}$. Trouvez des points $x \in X$ tels que $\omega(x) = F$. Même question avec F réduit au point $\{0\}_{i \in \mathbf{N}}$.

exercice 9 :

Soit X un espace métrique topologiquement complet, $T : X \rightarrow X$ une application continue. Soit x un point dont l'orbite est fermée. Montrez que x est pré-périodique ou $\omega(x)$ est vide.

exercice 10 :

Soit X un espace métrique sans point isolé, $T : X \rightarrow X$ un homéomorphisme. On suppose qu'il existe $x \in X$ tel que $\bigcup_{n \in \mathbf{Z}} T^n x$ est dense dans X . Montrez que T est transitive.

exercice 11 * :

Donnez un exemple de système dynamique continu transitif mais pas ergodique.

Commentaires

On vient de voir que pour une transformation préservant une mesure finie de support total, presque tous les points sont récurrents. Stricto sensu, c'est ce résultat qui devrait s'appeler théorème de récurrence de Poincaré. Voici comment il est énoncé dans le mémoire de 1890, "Sur le problème des trois corps et les équations de la dynamique" :

"Si on laisse de côté certaines trajectoires exceptionnelles, dont la réalisation est infiniment peu probable, on peut démontrer que le système repassera une infinité de fois aussi près que l'on voudra de sa position initiale."

L'usage réserve cependant le terme de Théorème de Récurrence de Poincaré à l'énoncé purement mesurable démontré précédemment.

Il existe d'autres concepts de récurrence. Un point x est dit récurrent par chaîne si pour tout $\varepsilon > 0$, on peut trouver $n \in \mathbf{N}$ et $x_1, x_2, \dots, x_n$ avec $x_1 = x_n = x$, tels que $d(T(x_i), x_{i+1}) < \varepsilon$ pour tout $i < n$. En d'autres termes, quitte à s'autoriser une petite erreur à chaque itération, on peut revenir arbitrairement près du point de départ. Un point non-errant est récurrent par chaîne, et l'ensemble des points récurrents par chaîne est fermé, invariant par T . Le résultat suivant, dû à C. Conley (1978), est une première étape dans la décomposition de X en ensembles invariants. *Supposons X métrique compact, f continue. Il existe une fonction continue $\psi : X \rightarrow \mathbf{R}$ constante sur les orbites des points récurrents par chaîne, et strictement décroissante sur les autres orbites ; de plus $\psi(\mathbf{R})$ est d'intérieur vide.*

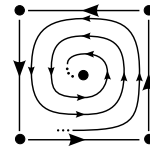
On peut caractériser la transitivité en terme de fonctions invariantes (Keynes Robertson, 1968) : *Soit $T : X \rightarrow X$ un homéomorphisme défini sur un espace métrique sans point isolé. Alors T est transitive si et seulement si toute fonction invariante par T , dont l'ensemble des points de continuité est dense, est constante sur un G_δ -dense.*

On a utilisé le théorème de Baire, pour montrer que si tout ouvert invariant non vide est dense, alors la transformation admet un G_δ -dense de points qui ont tous une orbite dense. Ici, le théorème de Baire joue un rôle analogue au théorème ergodique dans le cadre mesurable : l'absence d'ensembles invariants non-triviaux donne un "gros" ensemble de points dont l'orbite occupe beaucoup d'espace.

Le théorème de Baire est valide dans tous les espaces *topologiquement complets*, c'est-à-dire dans tout espace topologique homéomorphe à un espace métrique complet. Un sous-ensemble d'un espace topologiquement complet est topologiquement complet si et seulement si c'est une intersection dénombrable d'ouverts (Théorème de Mazurkiewicz, Alexandrov). Les espaces métriques localement complets (i.e. ceux dont chaque point admet un voisinage complet), *a fortiori* les espaces métriques localement compacts, sont topologiquement complets ; pour le démontrer, il suffit de vérifier qu'un tel espace est l'intersection de tous ses $1/k$ -voisinages dans sa complétion. La boule unité ouverte d'un espace de Hilbert est un exemple d'espace topologiquement complet qui n'est ni localement compact, ni complet pour sa métrique naturelle.

En lien avec la théorie de la mesure, il est intéressant de travailler avec des espaces séparables, c'est-à-dire avec des espaces qui possèdent une partie dénombrable dense. Un espace *polonais* est un espace topologique homéomorphe à un espace métrique séparable complet. Un tel espace possède une base dénombrable d'ouverts, le théorème de Baire est vérifié, et les mesures finies boréliennes sont automatiquement intérieurement régulières (cf annexe). Les espaces polonais fournissent donc un cadre de travail naturel pour les questions de dynamique topologique, si on souhaite sortir du cadre des espaces métriques compacts.

Soit T une application continue et Ω son ensemble non-errant. L'ensemble non-errant de la restriction de T à Ω n'est pas forcément égal à Ω . Contrairement à la récurrence, il ne suffit pas d'examiner la trajectoire d'un point x pour déterminer s'il est errant ou non. En effet, les trajectoires assurant le retour au voisinage de x peuvent s'éloigner de l'orbite de x . On a dessiné un exemple ci-contre. La transformation est définie sur $[0, \pi]^2$. Elle possède cinq points fixes. Le point au centre du carré est attractif et tous les points à l'intérieur du carré convergent vers ce point. Les points sur le bord du carré convergent vers l'un des quatre sommets. L'ensemble non-errant de la transformation est constitué de tous les points sur le bord du carré. Après restriction au carré, l'ensemble non-errant ne contient plus que les quatre sommets. Une telle transformation peut être construite à partir du flot d'équation : $x' = \sin(x) (\cos(y) + 1/10 \cos(x))$, $y' = \sin(y) (1/10 \cos(y) - \cos(x))$.



Une transformation transitive étant donnée, il est en général difficile de déterminer si un point particulier a une orbite dense. Par exemple, la transformation de l'intervalle $[0, 1]$ donnée par $x \mapsto 4x(1-x)$ est transitive ; peut-on exhiber un nombre rationnel explicite dans $[0, 1]$ qui admette une orbite dense sous l'action de cette transformation ? Cette question revient à trouver un rationnel r tel que la suite $\{\frac{2^n}{\pi} \arcsin(r)\}_{n \geq 0}$ est dense dans $\mathbf{R}/\mathbf{Z}$. La question est ouverte, même si numériquement, il semble que la plupart des rationnels ont effectivement une orbite dense.

Non-errance

Et il s'efforçait, poussant ce rocher des mains et des pieds jusqu'au faite d'une montagne. Et quand il était près d'atteindre ce faite, alors la force lui manquait, et l'immense rocher roulait jusqu'au bas. Et il recommençait de nouveau, et la sueur coulait de ses membres, et la poussière s'élevait au-dessus de sa tête.

Homère

Considérons un système dynamique donné par un espace métrique localement compact X et par une application continue $T : X \rightarrow X$. Un point est errant s'il admet un voisinage disjoint de tous ses itérés.

Un point peut être non-errant et pourtant avoir une trajectoire qui part à l'infini. Voici un exemple : considérons un système dynamique défini sur un ensemble compact, préservant une mesure de probabilité et possédant un point fixe, par exemple un automorphisme d'un tore plongé dans $\mathbf{R}^3$. Envoyons le point fixe à l'infini, en s'arrangeant pour que l'aire de la surface obtenue reste finie. Tous les points sont non-errants car le système préserve une mesure finie de support total, mais les points qui convergeaient vers l'origine partent maintenant à l'infini.

Lorsque tous les points sont non-errants, ce comportement est cependant exceptionnel. On va montrer que dans cette situation, les points récurrents forment un G_δ -dense de points sur X : la plupart des trajectoires reviennent arbitrairement proches de leur position initiale.

A l'opposé, que peut-on dire si l'ensemble non-errant est fini ? Dans ce cas, le système n'a qu'un nombre fini de points périodiques et toutes les trajectoires partent à l'infini ou convergent vers une de ces orbites périodiques. Cette situation se rencontre en mécanique lorsqu'on prend en compte les forces de frottement. L'énergie n'est pas conservée et le système finit par atteindre une position d'équilibre stable. D'un certain point de vue, c'est le comportement asymptotique le plus simple qu'on puisse observer pour un système dynamique.

Ce genre de dynamique existe sur toutes les variétés compactes. Voici comment construire des exemples en dimension 2 : toute surface compacte orientable est homéomorphe à une sphère, un tore ou un "beignet". Elle peut donc être découpée en pantalons et en disques. Il suffit donc de construire des transformations sur chacun des morceaux, de façon à ce qu'elles se correspondent par le biais des recolllements.

Lorsque l'ensemble non-errant est fini, on peut représenter la dynamique à l'aide d'un graphe ; les sommets du graphe sont les points de l'ensemble non-errant et deux sommets sont connectés s'il existe une trajectoire reliant les deux points. Une telle représentation s'avère utile dans les problèmes de classification.

On va utiliser le théorème de Baire dans ce chapitre. Rappelons qu'un espace *topologiquement complet* est un espace topologique homéomorphe à un espace métrique complet ; pour saisir l'intérêt de cette notion, on peut penser à l'intervalle $]0, 1[$, qui n'est pas complet pour sa distance naturelle, mais qui est topologiquement complet. Dans un tel espace, le théorème de Baire est satisfait : toute intersection dénombrable d'ouverts denses est dense.

Non-errance

Soit X un espace métrique, $T : X \rightarrow X$ une application continue. On rappelle la définition de l'ensemble non-errant :

$$\Omega = \{x \in X \mid \forall U \text{ ouvert contenant } x, \exists n_i \rightarrow \infty \text{ tq } T^{-n_i}U \cap U \neq \emptyset\}$$

Cet ensemble est fermé, satisfait $T(\Omega) \subset \Omega$, et contient tous les ensembles limites $\omega(x) = \{y \mid \exists n_i \rightarrow \infty \text{ tq } T^{n_i}(x) \rightarrow y\}$. La feuille stable d'un point $x \in X$ est définie par :

$$W^{ss}(x) = \{y \in X \mid d(T^n(x), T^n(y)) \xrightarrow{n \rightarrow \infty} 0\}$$

Les résultats qui suivent décrivent la dynamique de T dans les cas extrêmes Ω fini et $\Omega = X$.

Théorème

Soit X un espace métrique localement compact, $T : X \rightarrow X$ une application continue. On suppose que l'ensemble non-errant Ω de T est fini. Soit $\mathcal{P}$ l'ensemble des points périodiques de T . Alors tous les points de X ont une orbite qui est attirée par un point périodique, ou qui part à l'infini :

$$X = \bigcup_{x \in \mathcal{P}} W^{ss}(x) \cup \{x \mid \omega(x) = \emptyset\}$$

Preuve

Soit $x_0 \in X$. Considérons un point dans $\omega(x_0) \subset \Omega$; comme son orbite est contenue dans Ω qui est fini, un de ses itérés p est périodique. Notons l sa période et remplaçons p par une de ses images de telle sorte qu'il appartienne à l'adhérence de $\{T^{nl}(x_0)\}_{n \in \mathbb{N}}$. Soit $\varepsilon_0 > 0$ tel que $\{x \in X \mid d(p, x) \leq \varepsilon_0\}$ soit compact et disjoint de $\Omega \setminus \{p\}$. Pour tout $\delta < \varepsilon_0$ suffisamment petit, on a l'inclusion $T^l(B(p, \delta)) \subset B(p, \varepsilon_0)$. cf fig 1.

Comme la couronne $\{x \mid \delta \leq d(p, x) \leq \varepsilon_0\}$ est compacte, disjointe de Ω , elle ne contient qu'un nombre fini d'itérés de x_0 . Par conséquent, on peut trouver un $N \in \mathbb{N}$ tel que pour tout $n \geq N$, $T^{nl}(x_0)$ est dans $B(p, \delta) \cup B(p, \varepsilon_0)^c$. Soit $n \geq N$; si $T^{nl}(x_0)$ est dans $B(p, \delta)$, $T^l(T^{nl}(x_0))$ est dans $B(p, \varepsilon_0)$ et donc $T^{(n+1)l}(x_0)$ est dans $B(p, \delta)$. Par récurrence, $T^{nl}(x_0)$ est dans $B(p, \delta)$ pour tout n suffisamment grand. Ceci montre que $T^{nl}(x_0)$ converge vers p et x_0 est dans $W^{ss}(p)$.

Remarque

Plus généralement, si $\omega(x)$ rencontre une partie compacte invariante K de Ω qui est isolée dans Ω (ie $d(K, \Omega \setminus K) > 0$), alors $\omega(x)$ est contenu dans K .

Théorème

Soit X un espace métrique topologiquement complet, $T : X \rightarrow X$ une application continue. On suppose que l'ensemble non-errant de T est égal à X . Alors les points récurrents sont denses dans X .

Preuve

Posons

$$U_{k,N} = \bigcup_{n \geq N} \{x \in X \mid d(x, T^n x) < \frac{1}{k}\}.$$

L'ensemble des points récurrents coïncide avec l'intersection des $U_{k,N}$ pour tout $k, N \in \mathbb{N}^*$. Si $\Omega = X$, alors les ouverts $U_{k,N}$ sont denses dans X ; il suffit d'appliquer le théorème de Baire pour conclure.

Remarque

En général, les points récurrents sont denses dans l'intérieur de Ω (si cet intérieur est non vide), mais pas forcément dans Ω . De plus l'ensemble non-errant de la transformation T restreinte à Ω n'est pas toujours égal à Ω .

Exemples

– Dynamique Nord-Sud

Soit $x \in \mathbf{R}^n$ et $\lambda \in \mathbf{R}$. On considère l'application $x \mapsto \lambda x$. Cette application peut être étendue à la sphère $\mathbf{S}^n$ à l'aide de la projection stéréographique définie de $\{(x, t) \in \mathbf{R}^n \times \mathbf{R} \mid |x|^2 + t^2 = 1\}$ sur $\mathbf{R}^n$ et donnée par $(x, t) \mapsto \frac{2x}{1-t}$. L'application obtenue est un homéomorphisme de $\mathbf{S}^n$. Si $|\lambda| \neq 1$, son ensemble non-errant est composé d'un point fixe attractif et d'un point fixe répulsif. cf fig. 2.

– Un point-selle sur $\mathbf{S}^2$

On considère l'application de $\mathbf{R}^2$ dans $\mathbf{R}^2$ donnée par $(x, y) \mapsto (\frac{1}{2}x + \frac{x}{1+x^2}, \frac{1}{2}y)$. L'ensemble non-errant se compose des points $(-1, 0)$ et $(1, 0)$ qui sont attractifs, et du point $(0, 0)$, qui est un “point-selle”. Leurs feuilles stables sont données par :

$$W^{ss}(\pm 1, 0) = \{(x, y) \mid \pm x > 0\}, \quad W^{ss}(0, 0) = \{(x, y) \mid x = 0\}$$

Cette application peut se relever à la sphère $\mathbf{S}^2$ à l'aide de la projection stéréographique, auquel cas le pôle nord devient un point fixe répulsif. cf fig. 3.

L'application $(x, y) \mapsto (\frac{1}{2}(|x + \frac{1}{2}| - |x - \frac{1}{2}| + x), \frac{1}{2}y)$ a une dynamique similaire ; elle est linéaire au voisinage des points fixes.

– Encore la sphère

Soit A une matrice $n \times n$ à coefficients réels qui possède n valeurs propres réelles distinctes, et soit $f : \mathbf{S}^{n-1} \rightarrow \mathbf{S}^{n-1}$ l'application donnée par $f(x) = \frac{Ax}{\|Ax\|}$. L'ensemble non-errant est composé de n points qui peuvent être fixes ou périodiques de période deux, selon le signe des valeurs propres.

– Recollement au niveau des points attractifs

Considérons deux homéomorphismes f_1, f_2 de la sphère $\mathbf{S}^2$. On suppose que f_1 possède un point fixe p_1 attractif et qu'il existe une carte locale U_1 autour de ce point dans laquelle f_1 est de la forme $x \mapsto \frac{1}{2}x$. On suppose également que f_2 possède un point fixe répulsif de la forme $x \mapsto 2x$ dans une carte U_2 bien choisie.

On peut alors construire une nouvelle surface en supprimant des disques autour de p_1 et p_2 et en raccordant les deux sphères à l'aide d'un cylindre joignant les deux trous. Pour cela, on se place dans des système de coordonnées polaires (r_1, θ_1) autour de p_1 et (r_2, θ_2) autour de p_2 . On supprime les deux disques $\{|r_1| \leq 1/4\}$, $\{|r_2| \leq 1/4\}$ et on raccorde les anneaux $\{1/4 < |r_1| < 1\}$, $\{1/4 < |r_2| < 1\}$ à l'aide de l'application $(r_2, \theta_2) = (4r_1^{-1}, \theta_1)$. Les applications f_1 et f_2 se correspondent par le biais de ce recollement et donnent un homéomorphisme f de notre nouvelle surface. L'ensemble non-errant de f est égal à l'union des ensembles non-errants de f_1 et f_2 privés de p_1 et p_2 . cf fig. 4.

On peut aussi recoller deux points fixes p_1, p_2 d'un même homéomorphisme f_1 . Si le bassin d'attraction de p_1 est disjoint du bassin de répulsion de p_2 , l'ensemble non-errant obtenu est égal à celui de f_1 , privé des deux points p_1 et p_2 .

Par contre, si les deux bassins ne sont pas disjoints, il peut arriver que l'ensemble non-errant soit plus gros que celui de l'application initiale. C'est le cas par exemple lorsqu'on recolle le pôle Nord au pôle Sud dans une dynamique Nord-Sud ; l'application obtenue est une translation sur le tore, tous les points sont non-errants.

Graphe associé à la dynamique

Pour représenter la dynamique, on peut construire un graphe orienté comme suit : Les sommets sont les points de Ω . On relie deux sommets par une arête orientée si on peut trouver un point avec des itérés négatifs qui tendent vers le premier point, et des itérés positifs qui tendent vers le second point. En d'autres termes, $x, y \in \Omega$ sont reliés si $W^{su}(x) \cap W^{ss}(y) \neq \emptyset$. La présence de cycles dans ce diagramme peut indiquer la présence de points récurrents pour la transformation. cf fig. 5.

figure 1 : Ensemble non-errant fini



figure 2 : Dynamique Nord-Sud

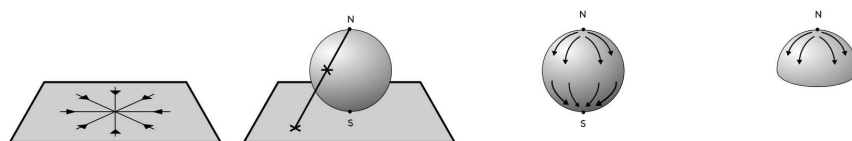


figure 3 : Un point Selle sur la sphère

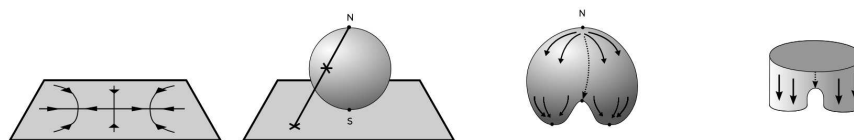


figure 4 : Recollement

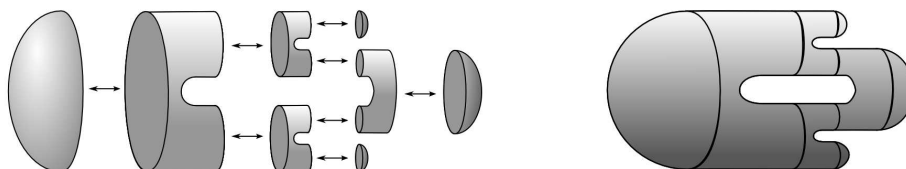
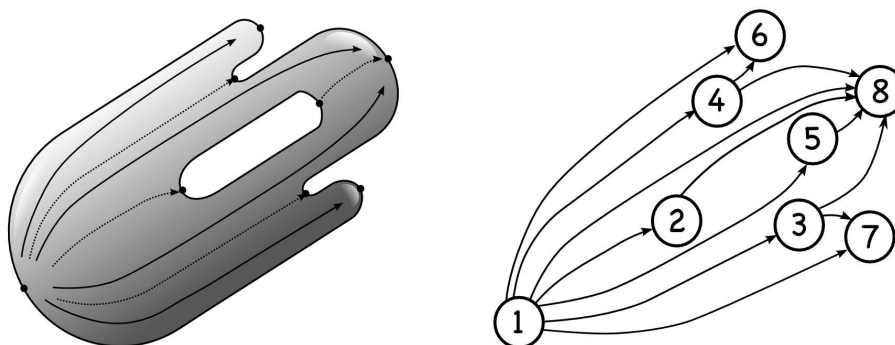


figure 5 : Graphe associé à la dynamique



Exercices

exercice 1 :

Soit $v \in \mathbf{R}^d$. On considère une translation $T(x) = x + v$ définie sur le tore $\mathbf{T}^d$. A quoi est égal son ensemble non-errant ?

exercice 2 :

Soit X un espace métrique et $T : X \rightarrow X$ un homéomorphisme dont l'ensemble non-errant est fini. Montrez que cet ensemble est composé de points périodiques.

exercice 3 :

Soit T un homéomorphisme défini sur un espace compact X . Soit $x \in X$. On définit la feuille instable de x par : $W^{su}(x) = \{ y \in X \mid d(T^{-n}(x), T^{-n}(y)) \xrightarrow{n \rightarrow \infty} 0 \}$.

Supposons que l'ensemble non-errant Ω de T est fini. Montrez que $X = \bigcup_{x \in \Omega} W^{su}(x)$.

exercice 4 :

Soit X un espace localement compact, $T : X \rightarrow X$ continue et $K \subset \Omega$ un compact isolé dans Ω , tel que $T(K) \subset K$. Soit $W^{ss}(K) = \{ y \in X \mid d(T^n(y), K) \xrightarrow{n \rightarrow \infty} 0 \}$. On considère $x \in X$; montrez que si $\omega(x) \cap K \neq \emptyset$, alors $x \in W^{ss}(K)$.

exercice 5 :

Donnez un exemple d'application continue définie sur $\mathbf{S}^d$, dont l'ensemble non errant est réduit à un point. Que dire de la dynamique d'une telle application ?

exercice 6 :

Soit T une application continue définie sur un espace métrique X . Montrez que l'ensemble non-errant de T^2 est inclus dans celui de T . Donnez un exemple où l'inclusion réciproque est fautive.

exercice 7 :

Étudiez la dynamique de l'application $T : S^1 \rightarrow S^1$ donnée par la formule :

$$T(x) = x + \frac{1}{20}(1 - \cos(4\pi x)).$$

Étudiez ensuite la dynamique de l'application de $\mathbf{T}^2$ définie par $(x, y) \mapsto (Tx, Ty)$.

exercice 8 :

On considère $a, b \in \mathbf{C}$ satisfaisant $|a|^2 - |b|^2 = 1$. Montrez que l'homographie $z \mapsto \frac{az+b}{bz+a}$ envoie le disque unité fermé $\{z \in \mathbf{C} \mid |z| \leq 1\}$ dans lui-même. Calculez les points fixes de cette homographie et déterminez son ensemble non-errant.

exercice 9 :

Construisez des homéomorphismes de la sphère $\mathbf{S}^2$ avec un ensemble non-errant qui possède un, trois, sept éléments.

exercice 10 :

Donnez un exemple de difféomorphisme défini sur la surface de genre 2 (ie le beignet à deux trous), dont l'ensemble non-errant est fini.

exercice 11 :

Soit F un sous-ensemble fermé de $[0, 1]$. Construisez un homéomorphisme de $[0, 1]$ dont l'ensemble des points fixes est égal à F . Construisez un homéomorphisme de $[0, 1]$ dont l'ensemble non-errant est égal à F .

exercice 12 * :

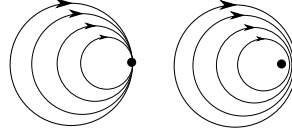
Soit X un espace métrique compact, $T : X \rightarrow X$ une application continue. Soit $x \in X$ tel que $\omega(x)$ ne contienne que des points fixes de T . Montrez que

$$d(T^{n+1}(x), T^n(x)) \xrightarrow{n \rightarrow \infty} 0.$$

En déduire que $\omega(x)$ est connexe. On suppose de plus que l'ensemble non-errant de T est dénombrable. Montrez que la suite $T^n(x)$ converge.

Commentaires

Lorsqu'on perturbe un homéomorphisme en topologie C^0 , l'ensemble non-errant peut rester proche de sa position initiale, ou au contraire "exploser". Considérons par exemple une homographie $z \mapsto \frac{az+b}{bz+\bar{a}}$, $|a|^2 - |b|^2 = 1$, du disque unité fermé dans lui-même. Si $a + b = 2$, l'ensemble non-errant est restreint au point fixe qui se trouve sur le bord du disque. Si $|a + b| < 2$, l'application est conjuguée à une rotation et tous les points du disque sont non-errants. Dans le cas $a + b = 2$, le graphe associé à la dynamique est composé d'un seul point, et d'une boucle en ce point. C'est l'exemple le plus simple de graphe possédant un cycle. La présence de cycles est souvent révélateur d'une certaine forme d'instabilité dans la dynamique. On peut étudier ces problèmes de stabilité à l'aide de filtrations ; cf Shub [Sh78] ch 2 et 3.



Une variété compacte M étant donnée, on peut toujours construire sur M un homéomorphisme T dont l'ensemble non-errant est fini. Par contre, il n'est pas toujours possible de construire un homéomorphisme T dont l'ensemble non-errant est fini, et dont les points périodiques sont prescrits. Supposons par exemple qu'il existe au voisinage de chacun des points périodiques une carte locale dans laquelle T est linéaire, sans valeur propre de module 1. Soit n_i le nombre de points périodiques pour lesquels l'application linéarisée a un sous-espace instable de dimension i . On a la formule suivante due à Smale :

$$\sum (-1)^i n_i = \chi$$

où χ est la caractéristique d'Euler de M . Pour une surface orientable M de genre g , cela donne : $n_0 - n_1 + n_2 = 2 - 2g$. On voit donc, par exemple, qu'on ne peut pas trouver d'homéomorphisme sur M dont l'ensemble non-errant est composé d'un point attractif ($n_0 = 1$), d'un point répulsif ($n_2 = 1$) et d'un point selle ($n_1 = 1$).

Voici comment on démontre la formule donnée plus haut dans le cas des surfaces : on remarque que les feuilles instables W^{su} des points selles forment les arêtes d'une "triangulation" de la surface, dont les sommets sont les points selles et les points répulsifs de T , et dont les faces sont les bassins d'attraction W^{ss} des points attractifs. La formule de Morse-Smale se déduit alors de la relation d'Euler, qui relie le nombre de sommets, d'arêtes et de faces d'une "triangulation" de la surface à sa caractéristique d'Euler.

Dans les exemples obtenus plus haut en recollant des demi-sphères et des "pantalons", la caractéristique d'Euler de la surface est égale à la différence du nombre de demi-sphères et du nombre de "pantalons". Le nombre de demi-sphères est égal à $n_0 + n_2$ tandis que le nombre de pantalons est égal à n_1 . On obtient bien la formule recherchée.

Un difféomorphisme de type Morse-Smale est un difféomorphisme défini sur une variété compacte connexe, dont l'ensemble non-errant est fini et dont les points périodiques sont hyperboliques ; on demande de plus que les variétés stables et instables des points périodiques s'intersectent de manière transverse, si elles s'intersectent. Les exemples construits plus haut sont des exemples de difféomorphismes Morse-Smale. Ces transformations sont stables, au sens où tout difféomorphisme proche d'une telle application en topologie C^1 lui est C^0 -conjuguée.

Le calcul de l'ensemble non-errant peut s'avérer difficile en pratique : par exemple, on ne sait toujours pas si tous les points sont non-errants pour un difféomorphisme Anosov défini sur une variété connexe. De même l'étude de la dynamique d'un polynôme quadratique $z \mapsto z^2 + c$ passe par une démonstration sophistiquée d'un théorème de non-errance.

Conjugaison

The source of all great mathematics is the special case, the concrete example. It is frequent in mathematics that every instance of a concept of seemingly great generality is in essence the same as a small and concrete special case.

P. R. Halmos (1916-2006)

Soit Y un espace métrique et $T : Y \rightarrow Y$ une application continue. Il n'est en général pas possible de calculer explicitement les itérés $T^n(x)$, afin de décider de la convergence ou de la divergence des orbites individuelles.

On peut par contre chercher un nouveau système de coordonnées dans lequel la transformation a une expression plus simple, permettant le calcul effectif de ses itérés. Soit X un autre espace métrique et $S : X \rightarrow X$ une application continue. On dit que T est *conjuguée* à S , d'un point de vue topologique, si on peut trouver un homéomorphisme $\varphi : X \rightarrow Y$ qui satisfait $\varphi \circ S = T \circ \varphi$. L'application φ représente le changement de coordonnées et S est l'expression de T dans ce nouveau système de coordonnées.

De la relation $\varphi \circ S = T \circ \varphi$, on déduit l'égalité : $T^n = \varphi \circ S^n \circ \varphi^{-1}$ pour tout entier n . La conjugaison φ met donc en bijection les orbites de S et celles de T . Si une trajectoire de S est convergente, alors son image par φ est aussi convergente ; si elle est dense, son image est dense. Si un ensemble ou une mesure est invariant par S , son image est également invariante par T , etc. Les dynamiques de S et T se correspondent, et l'étude de T se ramène à celle de S .

Peut-on, par le biais d'une conjugaison, ramener l'étude d'un système dynamique général à celle d'un petit nombre d'exemples bien choisis ? Voici deux familles de systèmes dynamiques pour lesquelles ce but a pu être atteint :

- les systèmes mécaniques complètement intégrables : problème des deux corps, mouvement géodésique sur un ellipsoïde, oscillateurs harmoniques, déplacement d'un solide dans un fluide idéal, toupie symétrique fixée par son extrémité. Ils peuvent être conjugués à des rotations sur des tores.
- les transformations des variétés riemanniennes compactes dilatant une métrique. Là encore, on peut les ramener à l'étude d'un système algébrique sur un espace homogène. Ces transformations ont un comportement nettement plus désordonné que les systèmes intégrables ; par exemple, elles sont topologiquement mélangeantes, ce qui n'est jamais le cas pour un système complètement intégrable.

Dans la suite, on étudie trois exemples : le premier est donné par une application de l'intervalle, qui s'avère dilatante après un changement judicieux de coordonnées. Le second provient de la physique ; il s'agit du pendule simple soumis à un champ de forces constant. Le troisième s'obtient en itérant une fraction rationnelle sur le plan complexe étendu $\mathbf{C} \cup \{\infty\}$. Dans chacun de ces exemples, la conjugaison repose sur une identité remarquable satisfaite par une fonction bien choisie. De ce point de vue, ces systèmes peuvent paraître plutôt exceptionnels.

Construire des conjugaisons est en général difficile, même si on affaiblit les conditions de régularité sur φ . Le problème de la conjugaison topologique des décalages de type fini est toujours ouvert.

Conjugaison

Définition

Soient X et Y deux espaces métriques, $S : X \rightarrow X$ et $T : Y \rightarrow Y$ deux applications boréliennes (continues, C^k). Une semi-conjugaison entre S et T est une application borélienne (continue, C^k) $\varphi : X \rightarrow Y$ qui satisfait : $\varphi \circ S = T \circ \varphi$. En d'autres termes, le diagramme ci-contre commute. On parle de conjugaison lorsque φ est bijective, d'inverse borélien (continu, C^k).

$$\begin{array}{ccc} X & \xrightarrow{S} & X \\ \varphi \downarrow & & \downarrow \varphi \\ Y & \xrightarrow{T} & Y \end{array}$$

Exemples

– L'application $x \rightarrow 2x \bmod 1$ est semi-conjugée au décalage sur un alphabet à deux symboles par le biais de l'application continue :

$$\begin{aligned} \varphi : \{0, 1\}^{\mathbf{N}} &\rightarrow [0, 1[\\ \{a_i\}_{i \in \mathbf{N}} &\mapsto \sum \frac{a_i}{2^{i+1}} \end{aligned}$$

– Considérons l'application $T : [0, 1[\rightarrow [0, 1[$ définie par :

$$T(y) = 4y(1 - y)$$

Posons $\varphi(x) = \sin^2(\frac{\pi}{2}x)$; élevons au carré la formule trigonométrique suivante : $\sin(2\frac{\pi}{2}x) = 2 \sin(\frac{\pi}{2}x) \cos(\frac{\pi}{2}x)$. On obtient l'égalité : $\varphi(2x) = 4\varphi(x)(1 - \varphi(x))$, c'est-à-dire : $\varphi(2x) = T(\varphi(x))$. L'application T est donc semi-conjugée à l'application $S(x) = 2x$ définie de $\mathbf{R}$ dans $\mathbf{R}$.

Théorème

Soit $S : X \rightarrow X$ et $T : Y \rightarrow Y$ deux applications boréliennes semi-conjugées par le biais d'une application borélienne φ . Soit μ une mesure borélienne S -invariante. Alors son image $\varphi_*\mu$ est une mesure borélienne invariante par T . Si S est ergodique relativement à μ (resp. mélangeante) alors T est ergodique (resp. mélangeante) relativement à $\varphi_*\mu$.

Preuve

– Montrons l'invariance de $\varphi_*\mu$: $T_*(\varphi_*\mu) = (T \circ \varphi)_*\mu = (\varphi \circ S)_*\mu = \varphi_*(S_*\mu) = \varphi_*\mu$.

– Soit $A \subset Y$ un ensemble T -invariant. Montrons que $\varphi^{-1}(A)$ est S -invariant :

$$S^{-1}(\varphi^{-1}(A)) = (\varphi \circ S)^{-1}(A) = (T \circ \varphi)^{-1}(A) = \varphi^{-1}T^{-1}(A) = \varphi^{-1}(A).$$

Comme μ est ergodique, $\varphi^{-1}(A)$ est négligeable, ou de complémentaire négligeable, relativement à μ . Il en va donc de même pour A relativement à $\varphi_*\mu$.

– Pour le mélange, soit $A, B \subset Y$ deux sous-ensembles boréliens.

$$\varphi_*\mu(T^{-n}A \cap B) = \mu(\varphi^{-1}T^{-n}A \cap \varphi^{-1}B) = \mu(S^{-n}\varphi^{-1}A \cap \varphi^{-1}B).$$

Cette quantité converge vers $\mu(\varphi^{-1}A) \mu(\varphi^{-1}B)$, ce qui est la valeur recherchée.

Exemple

Posons $S(x) = 2x$ si $x \in [0, 1/2[$, $S(x) = 2 - 2x$ si $x \in [1/2, 1[$; c'est-à-dire $S(x) = 1 - |2x - 1|$ si $x \in [0, 1[$; cf fig.1. Un calcul direct montre que S est conjugué à $T(y) = 4y(1 - y)$ sur $[0, 1[$ par le biais de l'homéomorphisme $y = \varphi(x) = \sin^2(\frac{\pi}{2}x)$.

$$\begin{array}{ccc} [0, 1[& \xrightarrow{x \mapsto 1 - |2x - 1|} & [0, 1[\\ \varphi \downarrow & & \downarrow \varphi \\ [0, 1[& \xrightarrow{y \mapsto 4y(1 - y)} & [0, 1[\end{array}$$

Comme S est mélangeante relativement à la mesure dx , T est mélangeante relativement à la mesure $dy/(\pi\sqrt{y(1 - y)})$; cf fig.2. En effet, cette mesure est l'image de la mesure de Lebesgue par φ : $d(\sin^2(\frac{\pi}{2}x)) = 2\frac{\pi}{2} \sin(\frac{\pi}{2}x) \cos(\frac{\pi}{2}x)dx = \pi\sqrt{y(1 - y)} dy$.

A priori toute identité remarquable est susceptible de donner une conjugaison intéressante. Fixons un paramètre $k \in]0, 1[$ et définissons la fonction elliptique de Jacobi sn par le biais de sa fonction réciproque :

$$sn^{-1}(z) = \int_0^z \frac{dt}{\sqrt{(1-t^2)(1-k^2t^2)}}$$

Posons $K = sn^{-1}(1)$. La fonction sn dépend du paramètre k et définit une bijection continue de $[-1, 1]$ sur $[-K, K]$. Cette bijection est holomorphe au voisinage de 0. La dérivée de sn se calcule par la formule de dérivation des fonctions réciproques ; on obtient $sn'(x)^2 = (1 - sn^2(x))(1 - k^2 sn^2(x))$. La fonction sn satisfait l'identité :

$$sn(u+v) = \frac{sn(u)sn'(v) + sn'(u)sn(v)}{1 - k^2 sn^2(u)sn^2(v)}$$

Pour établir cette relation, il suffit de dériver le second membre, d'abord par rapport à u puis par rapport à v , et de constater que ces deux dérivées sont égales. Cela montre que le second membre peut s'écrire comme une fonction de la somme $u+v$, fonction qu'on calcule en prenant $v=0$. On en déduit :

$$sn(2u) = \frac{2sn(u)sn'(u)}{1 - k^2 sn^4(u)}$$

Cette égalité permet d'étendre sn en une fonction méromorphe du plan complexe : pour tout z , on peut trouver un entier n tel que $sn(z/2^n)$ soit bien défini ; on en déduit la valeur de $sn(z)$ en appliquant l'identité remarquable. Calculons à présent quelques valeurs de sn à l'aide des formules précédentes :

$$sn(K) = 1, \quad sn'(K) = 0, \quad sn(2K) = 0, \quad |sn'(2K)| = 1, \quad |sn(u+2K)| = |sn(u)|.$$

On en déduit : $sn(u+4K) = sn(u)$. La fonction sn est périodique, de période $4K$. Posons $K' = \int_1^{\frac{1}{k}} \frac{dt}{\sqrt{(t^2-1)(1-k^2t^2)}}$. D'après la définition de sn , $sn^{-1}(\frac{1}{k}) = K + iK'$.

$$sn(K + iK') = \frac{1}{k}, \quad sn'(K + iK') = 0, \quad sn(2K + 2iK') = 0.$$

On en déduit comme plus haut : $sn(u + 4iK') = sn(u + 4K + 4iK') = sn(u)$.

La fonction sn admet une seconde période $4iK'$! Voici deux exemples d'application :

Le pendule simple

Les fonctions elliptiques permettent d'intégrer explicitement certaines équations de la mécanique classique. Considérons le cas d'un pendule simple soumis au champ de gravitation terrestre. Notons par θ l'angle que fait le pendule avec la verticale ; cf fig.3. L'énergie est conservée au cours du mouvement, ce qui donne l'équation :

$$\theta'^2 - 2\omega^2 \cos \theta = E = C^{ste}$$

Posons $\psi = \frac{2\omega}{\sqrt{E+2\omega^2}} \sin(\frac{\theta}{2})$. Cette fonction vérifie : $\psi'^2 = \omega^2(1-\psi^2)(1 - (\frac{1}{2} + \frac{E}{4\omega^2})\psi^2)$. Lorsque $|E| < 2\omega^2$, cette équation admet comme solutions : $\psi(t) = sn(\omega t + C)$. Dans l'espace des phases (ψ, ψ') , la dynamique est conjuguée à une rotation.

Les exemples de Schröder (1871)

La fonction $z \mapsto sn^2(z)$ possède deux périodes. On peut donc passer au quotient et obtenir une application définie sur le tore $\mathbf{T}^2 = \mathbf{C}/(2K\mathbf{Z} + 2iK'\mathbf{Z})$, à valeurs dans $\mathbf{C} \cup \{\infty\}$. Cette application holomorphe est surjective car $sn^2(\mathbf{T}^2)$ est un ensemble à la fois ouvert et fermé. Elle donne une semi-conjugaison entre la transformation mélangeante du tore $x \mapsto 2x$ et la fraction rationnelle :

$$z \mapsto \frac{4z(1-z)(1-k^2z)}{(1-k^2z^2)^2}$$

Cette fraction rationnelle est donc transitive sur $\mathbf{C} \cup \{\infty\}$, ergodique et mélangeante relativement à une mesure absolument continue par rapport à Lebesgue.

figure 1 : Conjugaison de $T(x) = 4x(1-x)$ à l'application "tente"

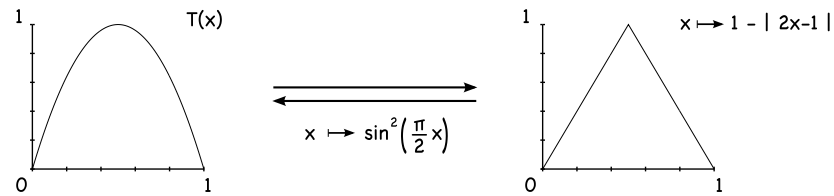


figure 2 : Comportement statistique des orbites de $T(x) = 4x(1-x)$

Cet histogramme comporte 100 bâtons de largeur 1/100. Il a été obtenu à partir du premier million d'itérés par T du point 0,9.

La distribution limite apparaît clairement sur l'histogramme.

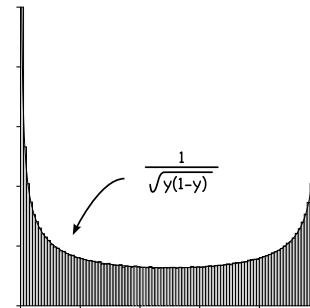
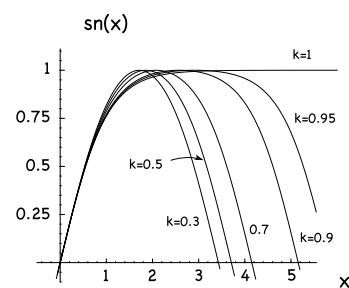
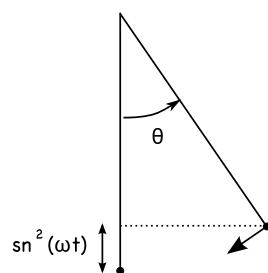


figure 3 : Pendule pesant



Exercices

exercice 1 :

Donnez une conjugaison C^0 explicite entre les applications $f_1, f_2 : \mathbf{R} \rightarrow \mathbf{R}$ définies par $f_1(x) = 2x$ et $f_2(x) = 3x$. Montrez qu'on peut trouver une conjugaison C^1 sur $\mathbf{R} - \{0\}$. Peut-on trouver une conjugaison C^1 sur $\mathbf{R}$ tout entier ?

exercice 2 :

Classez toutes les applications linéaires du plan à conjugaison C^0 près. Même question avec des conjugaisons C^1 .

exercice 3 :

Soit A une matrice $n \times n$ de norme inférieure stricte à 1 et de déterminant positif. On suppose que toutes ses valeurs propres sont distinctes. Montrez que l'application de $\mathbf{R}^n$ dans $\mathbf{R}^n$ donnée par $x \mapsto Ax$ est conjuguée C^0 à l'application $x \mapsto \frac{1}{2}x$. On pourra commencer par traiter le cas où toutes les valeurs propres sont réelles.

exercice 4 :

Soit A une matrice $n \times n$ à coefficients entiers, et v un vecteur de $\mathbf{R}^n$. On suppose que 1 n'est pas valeur propre de A . Montrez que les applications induites sur $\mathbf{T}^n$ par $x \mapsto Ax$ et $x \mapsto Ax + v$ sont conjuguées par le biais d'une translation.

En déduire que ces deux transformations ont les mêmes propriétés relativement à la mesure de Lebesgue.

exercice 5 :

Soit $\varphi : \mathbf{R} \rightarrow \mathbf{R}$ une application continue 1-périodique ; posons $f(x) = 2x + \varphi(x)$.

– Montrez que la série $\sum_{n=0}^{\infty} 2^{-n-1} \varphi(f^n(x))$ converge ; notons ψ sa limite.

– Montrez que ψ est continue et périodique de période 1.

– Montrez la relation : $2\psi(x) = \varphi(x) + \psi(f(x))$.

– Soit $h(x) = x + \psi(x)$. Montrez que h est surjective et que $h(f(x)) = 2h(x)$.

– Montrez que si f est C^1 et $f' > 1$, alors h est injective.

– Soit $\bar{f} : \mathbf{R}/\mathbf{Z} \rightarrow \mathbf{R}/\mathbf{Z}$ l'application obtenue à partir de f en passant au quotient. Montrez que $\bar{f}$ est transitive, et qu'elle admet une mesure finie invariante de support total qui est mélangeante.

exercice 6 :

Montrez que la transformation de l'intervalle $]0, 1/2[$ donnée par $x \mapsto 2x(1-x)$ est conjuguée à la multiplication par 2 sur $] -\infty, 0[$ par le biais de la conjugaison $x \mapsto \frac{1}{2}(1 - e^x)$.

exercice 7 :

Étudiez la dynamique de l'application $T : [-1, 1] \rightarrow [-1, 1]$ donnée par la formule : $T(x) = 4x^3 - 3x$. On pourra utiliser une formule trigonométrique concernant le cosinus.

exercice 8 :

Soit $a, b, c, d \in \mathbf{R}$. Montrez qu'une homographie $z \mapsto \frac{az+b}{cz+d}$ est conjuguée, par le biais d'une homographie, à une translation $z \mapsto z + \alpha$, $\alpha \in \mathbf{R}$, à une homothétie $z \mapsto \lambda z$, $\lambda \in \mathbf{R}$, ou à une rotation $z \mapsto e^{i\theta}z$, $\theta \in \mathbf{R}$.

exercice 9 :

Le mouvement du pendule simple est périodique ; que vaut sa période ? Supposons le pendule en position verticale au temps $t = 0$. Montrez que son altitude au temps t est proportionnelle à $\sin^2(\omega t)$.

exercice 10 :

Montrez que les exemples de Schröder admettent une infinité de points périodiques et que ces points périodiques sont denses dans $\mathbf{C} \cup \{\infty\}$. Calculez explicitement la mesure invariante absolument continue par rapport à Lebesgue.

Commentaires

Les premières formules de duplication pour les intégrales elliptiques apparaissent dès le *XVIII^e*. La formule suivante se trouve dans le traité de Giulio Fagnano, *Prodizioni Matematiche* (1750) :

$$\int_0^r \frac{dt}{\sqrt{1-t^4}} = 2 \int_0^\rho \frac{dt}{\sqrt{1-t^4}}, \quad r^2 = \frac{4\rho^2(1-\rho^4)}{(1+\rho^4)^2}.$$

Ces intégrales interviennent dans le problème du découpage d'une ellipse en arcs de même longueur.

Les fonctions sn peuvent être construites à partir des fonctions θ de Jacobi et sont reliées aux fonctions $\wp$ de Weierstrass par la formule : $\wp(u) = -\frac{1}{3}(1+k^2) + \frac{1}{sn^2(u)}$.

Voici quelques systèmes physiques dont les équations du mouvement peuvent être intégrées à l'aide de fonctions elliptiques ; cf D.F. Lawden, *Elliptic functions* [La89] : corps soumis à une force en $1/r^4$ ou $1/r^5$, effets relativistes dans le problème des deux corps, chaîne en rotation rapide, solide en rotation libre autour d'un point fixe, flot électrique dans une plaque rectangulaire conductrice.

Les fractions rationnelles qui agissent transitivement sur $\mathbf{C} \cup \{\infty\}$ sont appelées exemples de Lattès, même si le travail de Lattès (1918) est postérieur à celui de Schröder. J. Milnor (2004) donne un aperçu historique sur ce problème.

On peut obtenir d'autres familles de fractions rationnelles transitives sur $\mathbf{C} \cup \infty$ en utilisant des identités remarquables pour $sn(mt)$, $m \in \mathbf{N}$. L'exemple de Schröder est la seule famille à un paramètre de degré 4. Il existe des familles de degré n^2 pour tout $n > 1$. On peut également construire des exemples isolés qui s'obtiennent en utilisant les formules de transformation des fonctions elliptiques relativement au paramètre k . Par exemple, la fonction $sn(u; k)$, associée au paramètre k , s'exprime en fonction de $sn(u; l)$, $l = \frac{2\sqrt{k}}{1+k}$, par la formule :

$$sn((1+k)u; k) = (1+k) \frac{sn(u; l)}{1+ksn^2(u; l)}$$

L'équation $l = k$ admet pour solutions $k = \pm \frac{1}{2}(1 \pm i\sqrt{7})$. Pour ces valeurs, l'application $z \mapsto \frac{(1+k)z}{(1+kz^2)}$ est donc transitive.

Les résultats de classification à homéomorphisme près conjuguent en général à un modèle de nature algébrique. Il faut ensuite classer ces modèles. Par exemple, deux automorphismes d'un tore ou d'une nilvariété sont topologiquement conjugués si et seulement s'ils sont algébriquement conjugués (J. P. Conze, J. C. Marcuard 1970).

On sait classer à conjugaison topologique près les applications C^1 des variétés différentiables compactes connexes, qui dilatent une métrique riemannienne (Shub 1969, Gromov 1981). La variété doit être difféomorphe au quotient d'un groupe de Lie G nilpotent simplement connexe par un groupe de transformations affines opérant proprement discontinûment sans point fixe ; une telle variété est dite "infra-nilpotente". La transformation est conjuguée au quotient d'un automorphisme de G dont les valeurs propres sont strictement supérieures à 1.

On sait aussi montrer que les difféomorphismes Anosov sur les variétés infra-nilpotentes sont topologiquement conjugués à des automorphismes sans valeurs propres de module 1 (Manning 1974). Les variétés infra-nilpotentes portant un difféomorphisme Anosov sont classifiées en dimension inférieure à 9 (J. Lauret, C. E. Will 2004).

Il existe des exemples de variétés homéomorphes à des tores ou à des variétés infranilpotentes, mais pas difféomorphes à de telles variétés, qui admettent des difféomorphismes Anosov (Farrell Jones 1978, Farrell Gogolev 2010). On ignore cependant si une variété qui porte un difféomorphisme Anosov est nécessairement homéomorphe à une variété infra-nilpotente.

Linéarisation

I specifically remember discussions among ourselves and with visitors about what is now known as nonlinear mathematics – truly a strange expression, for it is like saying “I will discuss nonelephant animals”

S. Ulam (1909-1984)

Pour étudier la dynamique d’une transformation f définie sur un espace X , on peut chercher à conjuguer f à un modèle plus simple $A : Y \rightarrow Y$ par le biais d’un homéomorphisme $\varphi : X \rightarrow Y$ satisfaisant $A \circ \varphi = \varphi \circ f$. Lorsque le modèle A est linéaire, on parle de *linéarisation*.

La dynamique d’une application linéaire est suffisamment simple pour qu’on puisse décrire complètement le comportement asymptotique de ses orbites, et en déduire celui de ses conjugués. On a représenté dans la *figure 2* le portrait de phase des applications linéaires de $\mathbf{R}^2$ diagonalisables sur $\mathbf{C}$.

Quelles sont les applications qui peuvent être linéarisées ? Cette question peut être abordée par des méthodes perturbatives : une transformation proche d’une application linéaire A lui est-elle conjuguée ? C’est vrai dès que A n’a pas de valeurs propres de module 1 ; cf *fig. 1*. Une telle application est dite *hyperbolique*.

Ce résultat se généralise en dimension infinie, dès l’instant où l’espace peut se décomposer en deux sous-espaces respectivement contractés et dilatés par l’application linéaire. Pour construire la conjugaison, il faut résoudre l’équation $A \circ \varphi = \varphi \circ f$, c’est-à-dire construire un point fixe pour l’application $\varphi \mapsto A \circ \varphi \circ f^{-1}$. Pour cela, on applique un théorème du point fixe hyperbolique, qui se déduit du théorème du point fixe pour une application Lipschitzienne contractante.

On peut aussi aborder le problème de la conjugaison de manière locale. Supposons que la transformation f est différentiable et possède un point fixe p . Au voisinage de ce point fixe, la transformation f est proche de sa différentielle $D_p f$. P. Hartman et D. M. Grobman (1960) montrent qu’on peut conjuguer f à sa différentielle dans un petit voisinage du point fixe, si cette différentielle est hyperbolique.

Lorsque f est inversible et que toutes les valeurs propres de $D_p f$ sont de module plus petit que 1, la conjugaison peut même s’étendre à l’ensemble des points dont l’orbite tend vers p , ce qui montre, à posteriori, que cet ensemble est un ouvert homéomorphe à $\mathbf{R}^d$; cf *fig. 3*. Cet ensemble est appelé *bassin d’attraction* du point fixe p . De manière générale, pour un difféomorphisme f , le théorème de Hartman-Grobman montre qu’il existe une bijection continue d’un certain espace euclidien $\mathbf{R}^k$ sur la *feuille stable* du point p :

$$W^{ss}(p) = \{x \in X \mid d(f^n(x), f^n(p)) \xrightarrow{n \rightarrow \infty} 0\}$$

Cette bijection n’est en général pas un homéomorphisme, la feuille stable pouvant s’accumuler sur elle-même, et revenir arbitrairement proche de p .

Enfin, ces considérations s’étendent au cas d’un point périodique de période n en considérant la transformation f^n au lieu de f .

Linéarisation

Définition

Soit E un Banach. Une application linéaire continue $T : E \rightarrow E$ est dite hyperbolique si on peut trouver deux sous-espaces fermés E_s, E_u de E tels que $E = E_s \oplus E_u$ et :

$$\begin{aligned} T(E_s) &\subset E_s, & \|T|_{E_s}\| &< 1 \\ T(E_u) &= E_u, & T|_{E_u} &\text{est inversible,} & \|(T|_{E_u})^{-1}\| &< 1 \end{aligned}$$

Théorème du point fixe hyperbolique

Soit E un Banach, $T : E \rightarrow E$ une application linéaire hyperbolique et $F : E \rightarrow E$ une application continue. On suppose que $F - T$ est lipschitzienne, de constante de Lipschitz suffisamment petite. Alors F a un unique point fixe.

Preuve

Considérons les deux identités :

$$(id - T)(id + T + \dots + T^n) = id - T^{n+1} \quad \text{et} \quad (T - id)(T^{-1} + T^{-2} + \dots + T^{-n}) = id - T^{-n}$$

La première de ces deux identités montre que $id - T : E_s \rightarrow E_s$ est inversible, la deuxième que $id - T : E_u \rightarrow E_u$ est inversible. L'application linéaire $id - T : E \rightarrow E$ est donc inversible.

Un point x est point fixe de F si et seulement si $(F - T)(x) = x - T(x)$, c'est-à-dire si et seulement s'il est point fixe de l'application $(id - T)^{-1}(F - T)$. Cette application est contractante dès que la constante de Lipschitz de $F - T$ est strictement inférieure à $\|(id - T)^{-1}\|^{-1}$. Elle possède donc un unique point fixe.

Théorème de Hartman-Grobman, cas lipschitzien

Soit B un Banach, $A : B \rightarrow B$ une application linéaire continue hyperbolique inversible et $f : B \rightarrow B$ une application continue inversible. On suppose que $f - A$ est bornée, lipschitzienne, de constante de Lipschitz suffisamment petite. Alors f et A sont conjuguées.

Preuve

Soit $g : B \rightarrow B$ une application satisfaisant les mêmes hypothèses que f . Montrons qu'il existe φ continue telle que $g \circ \varphi = \varphi \circ f$. Une telle application φ doit être point fixe de la transformation donnée par $\varphi \mapsto g \circ \varphi \circ f^{-1}$. On cherche φ sous la forme $\varphi = id + \psi$, avec $\psi : B \rightarrow B$ continue bornée.

$$\begin{aligned} g \circ (id + \psi) \circ f^{-1} &= (g - A) \circ (id + \psi) \circ f^{-1} + A \circ (id + \psi) \circ f^{-1} \\ &= (g - A) \circ (id + \psi) \circ f^{-1} + A \circ f^{-1} + A \circ \psi \circ f^{-1} \end{aligned}$$

Posons $T(\psi) = A \circ \psi \circ f^{-1}$ et $F(\psi) = T(\psi) + (g - A) \circ (id + \psi) \circ f^{-1} + A \circ f^{-1} - id$.

Ces deux applications sont définies de E dans E , avec $E = C_b(B, B)$ l'espace de Banach des fonctions continues bornées de B dans B . On cherche à montrer que F a un point fixe. Pour cela, montrons que T est hyperbolique et que la constante de Lipschitz de $F - T$ est petite.

Soient B_s et B_u les sous-espaces stables et instables de A . Posons $E_s = C_b(B, B_s)$ et $E_u = C_b(B, B_u)$. Ces deux espaces sont les espaces stables et instables de T :

Si $\psi \in E_s$, $\|T(\psi)\| = \|A \circ \psi \circ f^{-1}\|_\infty = \|A \circ \psi\|_\infty \leq \|A|_{B_s}\| \|\psi\|_\infty$.
Si $\psi \in E_u$, $(T|_{E_u})^{-1} = (A|_{B_u})^{-1} \circ \psi \circ f$ et $\|(T|_{E_u})^{-1}(\psi)\|_\infty \leq \|(A|_{B_u})^{-1}\| \|\psi\|_\infty$

La constante de Lipschitz de $F - T$ est inférieure à celle de $g - A$, qui est notée K :

$$\|(F - T)(\psi) - (F - T)(\psi')\|_\infty = \|(g - A) \circ (id + \psi) - (g - A) \circ (id + \psi')\|_\infty \leq K \|\psi - \psi'\|_\infty$$

Il reste à montrer que φ est inversible. Pour cela, intervertissons le rôle de g et f . On obtient une fonction φ' qui satisfait $f \circ \varphi' = \varphi' \circ g$. Ceci donne $g \circ \varphi \circ \varphi' = \varphi \circ \varphi' \circ g$ et $f \circ \varphi' \circ \varphi = \varphi' \circ \varphi \circ f$. Par unicité de la conjugaison, $\varphi \circ \varphi' = \varphi' \circ \varphi = id$.

Remarque

Dans le théorème précédent, il n'est pas nécessaire de supposer f inversible. Soit K la constante de Lipschitz de $f - A$. Montrons que si $K < \|A^{-1}\|^{-1}$, f est inversible.

Pour cela, posons $A_y(x) = A(x) - y$; cette application est lipschitzienne, d'inverse $A_y^{-1}x = A^{-1}x + A^{-1}y$. L'application $A_y^{-1}(A - f)$ est contractante, donc possède un point fixe, qui vérifie $y = f(x)$. La relation $x = A^{-1}(A - f)(x) + A^{-1}y$ donne $\|x - x'\| \leq \|A^{-1}\| K \|x - x'\| + \|A^{-1}\| \|y - y'\|$; f^{-1} est donc lipschitzienne.

Définition

Soit M une variété différentiable, $f : M \rightarrow M$ une application C^1 . Un point périodique p de f de période n est dit hyperbolique si $D_p f^n$ est inversible et n'a pas de valeur propre de module 1. On parle de point attractif (resp. répulsif) lorsque toutes les valeurs propres ont un module strictement inférieur à 1 (resp. > 1).

Montrons que cette notion d'hyperbolicité coïncide avec celle introduite précédemment. Soit A une matrice $n \times n$ sans valeurs propres de module 1. Soit Λ l'ensemble des valeurs propres de A de module inférieur à 1, et $E_s = \bigoplus_{\lambda \in \Lambda} \ker(A - \lambda \text{id})^n$. Sur le sous-espace invariant $\ker(A - \lambda \text{id})^n$, on a :

$$A^m = (\text{id} + (A - \lambda \text{id}))^m = \sum_{k=0}^{n-1} C_m^k \lambda^{m-k} (A - \lambda \text{id})^k \ll m^k |\lambda|^m$$

Soit $\lambda_0, \lambda_1 \in \mathbf{R}$ tels que $\max_{\lambda \in \Lambda} \lambda < \lambda_0 < 1$ et $\lambda_0 < \lambda_1 < 1$. Comme $\|A^m\| \leq \lambda_0^m$ dès que m est assez grand, on a convergence de la série $\|x\|_s = \sum_{m=0}^{\infty} \lambda_1^{-m} \|A^m x\|$ pour tout $x \in E_s$. Cette norme satisfait $\|Ax\|_s = \lambda_1(\|x\|_s - \|x\|) \leq \lambda_1 \|x\|_s$. Pour les valeurs propres de module supérieur à 1, on procède de même en considérant A^{-1} . La matrice A est alors hyperbolique relativement aux max des normes construites. Remarquons que si A n'a pas de valeurs propres de module supérieur à 1, la boule $B(0, r)$, relativement à la norme $\|\cdot\|_s$, satisfait $A(B(0, r)) \subset B(0, \lambda_1 r) \subset B(0, r)$.

Théorème de linéarisation de Hartman-Grobman

Soit M une variété différentielle de dimension n , $f : M \rightarrow M$ une application différentiable C^1 avec un point fixe hyperbolique p . Alors on peut trouver des voisinages $U_1, U_2 \subset M$ de p , et $V_1, V_2 \subset \mathbf{R}^n$ de 0, ainsi qu'un homéomorphisme $\varphi : U_1 \cup U_2 \rightarrow V_1 \cup V_2$ qui conjugue f et $D_p f$:

$$\begin{array}{ccc} U_1 & \xrightarrow{f} & U_2 \\ \varphi \downarrow & & \downarrow \varphi \\ V_1 & \xrightarrow{D_p f} & V_2 \end{array}$$

$\forall x \in U_1, \quad \varphi(f(x)) = D_p f(\varphi(x))$

Preuve

On se place dans une carte centrée en p . Montrons que $f - D_p f$ s'étend en une application de $\mathbf{R}^n$ dont la constante de Lipschitz est petite.

Soit $g : \mathbf{R}^n \rightarrow [0, 1]$ une application C^1 , nulle pour $|x| > 2$, égale à 1 pour $|x| < 1$. Posons $g_r(x) = g(x/r)$. L'extension de f est donnée par $\tilde{f} = g_r f + (1 - g_r) D_p f$, pour r bien choisi. Afin d'obtenir la majoration sur la constante de Lipschitz de $\tilde{f} - D_p f$, majorons la norme de la différentielle de $g_r (f - D_p f)$ pour $|x| < 2r$:

$$\begin{aligned} \|D_x(g_r (f - D_p f))\| &\leq \frac{1}{r} \|Dg\|_{\infty} \|f(x) - D_p f(x)\| + \|g_r\|_{\infty} \|D_x f - D_p f\| \\ &\leq \frac{1}{r} \|Dg\|_{\infty} (\sup_{|x| < 2r} \|D_x f - D_p f\|) |x| + \|D_x f - D_p f\| \end{aligned}$$

Il suffit de prendre r tel que $\sup_{|x| < 2r} \|D_x f - D_p f\| < \varepsilon / (1 + 2\|Dg\|_{\infty})$, ε petit, pour pouvoir appliquer le théorème de conjugaison linéaire. On obtient une conjugaison globale $\varphi : \mathbf{R}^n \rightarrow \mathbf{R}^n$ entre $\tilde{f}$ et $D_p f$. Sur $B(0, r)$, on a l'égalité $\tilde{f} = f$. On prend pour U_2 l'ouvert de M correspondant à $B(0, r)$ et $U_1 = f^{-1}(U_2) \cap U_2$.

Remarque

Lorsque p est attractif, on peut prendre $\overline{U_2} \subset U_1$, $\overline{V_2} \subset V_1$. Si f est inversible, la conjugaison s'étend de $\bigcup f^{-k}(U_1)$ à $\mathbf{R}^n$ en posant $\varphi = (D_p f)^{-k} \circ \varphi \circ f^k$ sur $f^{-k}(U_1)$.

figure 1 : Perturbation d'une application linéaire hyperbolique

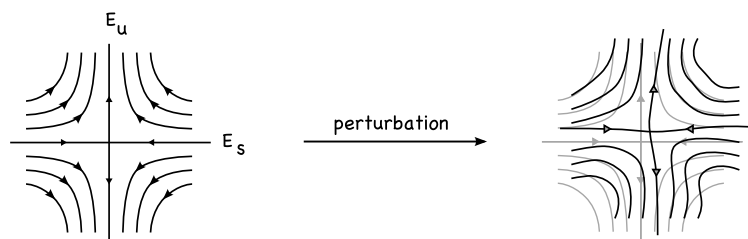


figure2 : Dynamique des applications linéaires de $\mathbb{R}^2$ dont les valeurs propres valent λ_1 et λ_2 .

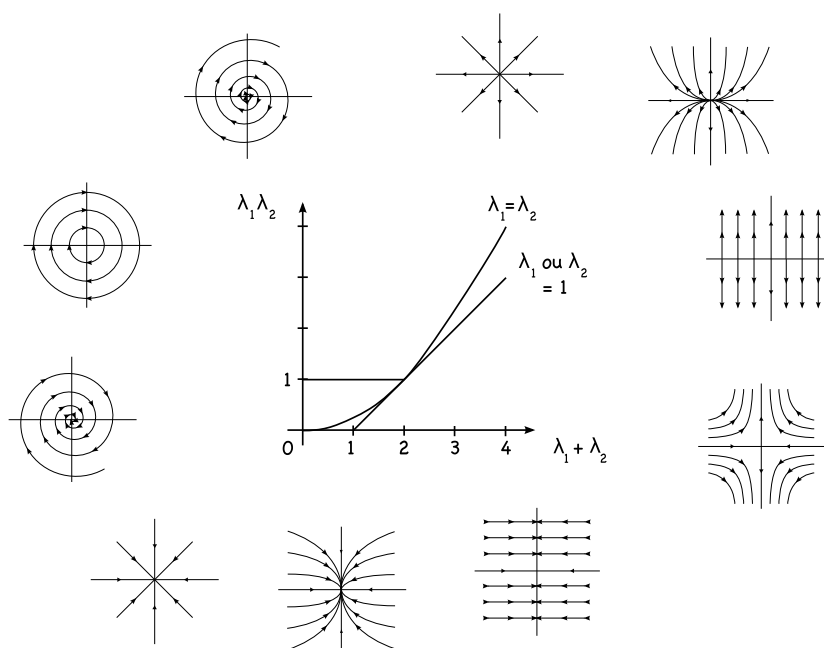
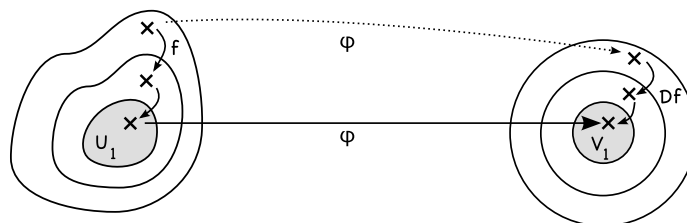


figure 3 : Extension de la conjugaison dans le cas contractant



Exercices

exercice 1 :

Soit X un espace métrique compact et $T : X \rightarrow X$ une application qui vérifie, pour tout couple de points distincts $x, y \in X$: $d(T(x), T(y)) < d(x, y)$.

Montrez que T a un unique point fixe x_0 et que $\forall x \in X, T^n(x) \xrightarrow{n \rightarrow \infty} x_0$.

exercice 2 :

Soit M une variété et $T : M \rightarrow M$ une application différentiable qui admet un point fixe hyperbolique p qui n'est pas attractif. Montrez que l'ensemble des points attirés par p est d'intérieur vide. (*Indic : Utiliser le théorème de Baire*)

exercice 3 :

On considère un difféomorphisme T défini sur une variété compacte M . On suppose que son ensemble non-errant est fini, composé de points périodiques hyperboliques. Montrez que l'union des bassins des points périodiques attractifs est un ouvert dense.

exercice 4 :

Soit f une application différentiable définie sur une variété compacte M . Soit p un point périodique hyperbolique. Montrez qu'il existe un voisinage de p tel que tous les points périodiques dans ce voisinage sont de période plus grande que celle de p . Supposons que tous les points périodiques de T sont hyperboliques. Soit $n \in \mathbf{N}^*$; montrez que l'ensemble des points périodiques de T de période n est fini.

exercice 5 :

Soit M une variété compacte connexe, $f : M \rightarrow M$ un difféomorphisme qui admet un point fixe attractif. Montrez que le complémentaire de son bassin d'attraction est compact connexe non vide. Contre-exemple lorsque M n'est pas compacte ?

exercice 6 :

Soit T un difféomorphisme défini sur une variété compacte M . Soit p un point fixe hyperbolique. On suppose qu'il existe un point $x \neq p$ tel que $x \in W^{ss}(p) \cap W^{su}(p)$. Montrez que x n'est pas récurrent. Montrez qu'il appartient à l'ensemble non-errant. En déduire que l'ensemble non-errant de T est infini.

exercice 7 :

On considère un difféomorphisme T défini sur une variété M et p un point fixe hyperbolique. Soit E_s le sous-espace stable de $D_p f$. Montrez qu'on a une semi-conjugaison bijective $\varphi : E_s \rightarrow W^{ss}(p)$ qui satisfait :

$$\forall x \in E_s, \quad f(\varphi(x)) = \varphi(D_p f(x))$$

$$\begin{array}{ccc} E_s & \xrightarrow{D_p f} & E_s \\ \varphi \downarrow & & \downarrow \varphi \\ W^{ss}(p) & \xrightarrow{f} & W^{ss}(p) \end{array}$$

exercice 8 * :

Soit T un difféomorphisme d'une variété compacte connexe M ($\neq S^1$). Supposons que l'ensemble non-errant Ω est fini, composé de points périodiques hyperboliques. Montrez que si $\text{Card}(\Omega) = 2$, M est homéomorphe à une sphère. Si $\text{Card}(\Omega) > 2$, alors Ω contient une orbite périodique qui n'est ni attractive ni répulsive.

exercice 9 * :

Montrez qu'il n'existe pas de difféomorphisme T , défini sur une variété compacte connexe, dont l'ensemble non-errant est composé de trois points fixes hyperboliques.

exercice 10 :

On se place sous les hypothèses du théorème de linéarisation lipschitzien. Soit $p_u, p_s : B \rightarrow B$ les projecteurs sur B_u et B_s . Montrez que la conjugaison φ entre A et f est donnée par les formules : $\varphi = id + (1 - T)^{-1}((A - f)f^{-1})$,

$$\varphi = id + \sum_{k \geq 0} p_u A^{-k-1} (A - f) f^k + p_s A^k (A - f) f^{-k-1} = \lim_{n \rightarrow \infty} p_u A^{-n} f^n + p_s A^n f^{-n}.$$

Commentaires

Il existe différentes méthodes pour obtenir une conjugaison entre deux systèmes dynamiques :

- appliquer un théorème de point fixe ;
- appliquer un théorème d'inversion locale ;
- chercher à obtenir cette conjugaison sous la forme d'une série ;
- construire la conjugaison de façon géométrique.

La preuve du théorème de linéarisation donnée plus haut suit Irwin [Irw01] et repose sur un théorème de point fixe. On aurait pu faire appel au théorème d'inversion locale suivant :

Soient U, V deux ouverts dans des espaces de Banach E et F , $f : U \rightarrow V$ un homéomorphisme dont l'inverse est lipschitzien. Soit $g : U \rightarrow F$ tel que la constante de Lipschitz de $f - g$ est strictement inférieure à l'inverse de celle de f^{-1} . Alors g est un homéomorphisme de U sur $g(U)$ et son inverse est lipschitzien.

La démonstration de ce résultat, qui fait appel au théorème du point fixe contractant, se trouve dans Shub [Sh78] App 5.1. Pour une preuve qui repose sur la géométrie des feuilles stables et instables, cf Palis-De Melo [PJ82]. Si on cherche à étudier la régularité de φ , il vaut mieux se servir d'un développement en série, par exemple celui proposé dans les exercices. Lorsque f est C^∞ et qu'il n'y a pas de résonances entre les valeurs propres $(\lambda_1, \dots, \lambda_n)$ de $D_p f$:

$$\forall i, \forall k_1, \dots, k_n \in \mathbf{N} \text{ tels que } \sum k_i > 1, \quad \lambda_i \neq \lambda_1^{k_1} \lambda_2^{k_2} \dots \lambda_n^{k_n},$$

S. Sternberg (1959) montre que la conjugaison est en fait C^∞ . cf par exemple Katok-Hasselblatt [HK95] Th 6.6.6. Un exemple de résonance est donné par un difféomorphisme de $\mathbf{R}^n$ ou $\mathbf{T}^n$ préservant la mesure de Lebesgue ($\prod \lambda_i = 1$). Dans ce cas, on peut montrer que deux telles applications C^∞ qui coïncident à tous les ordres en p sont conjuguées par le biais d'une conjugaison préservant le volume ; cf par exemple Banyaga, DeLaLlave, Wayne (1996).

A l'opposé, si on ne fait pas d'hypothèses sur la constante de Lipschitz de $f - A$, on a tout de même une semi-conjugaison continue φ telle que $A \circ \varphi = \varphi \circ f$ (prendre $g = A$ dans la démonstration). Si la constante de Lipschitz de $f - A$ est petite, φ satisfait la condition de régularité suivante : on peut trouver $C, D \in \mathbf{R}$, $0 < C < 1$, tels que $\forall n \in \mathbf{N}^*, ||x - y|| \leq D^n ||\varphi(x) - \varphi(y)|| + C^n$.

Une application lipschitzienne définie sur un sous-ensemble d'un espace métrique, à valeurs réelles, admet toujours une extension lipschitzienne avec même constante de Lipschitz ; cf Dudley [Du02] th 6.1.1. Pour toute application lipschitzienne f d'un sous-ensemble de $\mathbf{R}^n$ dans $\mathbf{R}^n$, $\mathbf{R}^n$ étant muni de la métrique euclidienne, on peut trouver une extension lipschitzienne avec même constante de Lipschitz (théorème de Kirszbraun, cf Federer [Fe69]).

Voici un théorème de conjugaison globale qui peut se déduire du théorème de Hartman-Grobman, cas Lipschitzien : toute application C^1 du tore $\mathbf{T}^n$, C^1 -proche d'un automorphisme hyperbolique, est en fait topologiquement conjuguée à cet automorphisme. Par conséquent, une telle application est transitive, topologiquement mélangeante, etc.

La situation est très différente lorsqu'on perturbe un automorphisme dont les valeurs propres sont de module 1. La famille standard $f_\lambda : \mathbf{T}^2 \rightarrow \mathbf{T}^2$ est définie par : $f_\lambda(x, y) = (x + y, y + \lambda \sin(2\pi(x + y)))$. Les applications f_λ préservent la mesure de Lebesgue, tous les points sont donc non-errants. Pour $\lambda = 0$, le système est intégrable ; les trajectoires de f_0 sont contenues dans les cercles $\{y = C^{ste}\}$ et il n'existe aucune orbite asymptotique aux points périodiques : $W^{ss}(p) = \{p\}$, $\forall p$ périodique. A l'opposé, pour $\lambda > 0$, on peut trouver un point périodique qui est hyperbolique ; sa feuille stable est donc non triviale. On ignore s'il existe une orbite dont l'adhérence est d'intérieur non vide. De même, on ne sait pas s'il peut exister un ouvert invariant en restriction duquel la mesure de Lebesgue est ergodique.

Un attracteur étrange

*Le grand carré n'a pas d'angles.
Le grand vase est lent à parfaire.
La grande musique n'a guère de sons.
La grande image n'a pas de forme.*
Lao Tseu

On peut montrer, à l'aide du théorème de Hartman-Grobman, qu'une petite perturbation f d'un automorphisme hyperbolique du tore est conjuguée à cet automorphisme. Pour une telle transformation, tous les points sont donc non-errants et il existe un ensemble dense de points récurrents.

Plaçons nous sur $\mathbf{T}^2$ et considérons l'automorphisme hyperbolique donné par la matrice $A = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$. Que se passe-t-il si on fait une perturbation localisée au voisinage de l'origine $(0,0)$, mais qui transforme ce point fixe en point attractif ? D'après le théorème d'Hartman-Grobman, il existe un ouvert U de points qui vont être attirés par $(0,0)$; cf fig.1. Que peut-on dire de cet ouvert ?

On va voir qu'il existe un point fixe hyperbolique p sur le bord de cet ouvert. Ses feuilles stables et instables :

$$W^{ss}(p) = \{x \in \mathbf{T}^2 \mid d(f^n(x), f^n(p)) \xrightarrow{n \rightarrow \infty} 0\} = \{x \mid d(f^n(x), p) \xrightarrow{n \rightarrow \infty} 0\}$$

$$W^{su}(p) = \{x \in \mathbf{T}^2 \mid d(f^{-n}(x), f^{-n}(p)) \xrightarrow{n \rightarrow \infty} 0\} = \{x \mid d(f^{-n}(x), p) \xrightarrow{n \rightarrow \infty} 0\}$$

forment deux sous-variétés immergées de dimension 1. La feuille stable de p ne peut bien sûr appartenir à l'ouvert U des points attirés par $(0,0)$. Nous allons montrer que son adhérence K est d'intérieur vide, et coïncide avec le complémentaire de U . Par conséquent, la plupart des orbites convergent vers l'origine.

Que peut-on dire de la dynamique de f^{-1} ? L'origine est maintenant point répulsif, et tous les autres points convergent vers le compact invariant K . De plus, la transformation est transitive en restriction à K . La structure de K est intéressante : il est d'intérieur vide dans $\mathbf{T}^2$ mais contient une sous-variété de dimension 1 immergée qui est à la fois dense dans K et d'intérieur vide dans K . Il s'agit donc d'un objet géométrique intermédiaire entre une droite et un plan. Il apparaît en noir sur la figure 1, tandis que son complémentaire, en blanc, correspond à l'ouvert U .

Les démonstrations sont basées sur le théorème de linéarisation de Hartman-Grobman, et sur l'existence d'une direction invariante définie sur K et dilatée par la différentielle de f . De fait, la transformation A possède une valeur propre strictement supérieure à 1 et la dilatation associée survit à la perturbation, lorsqu'on est loin de l'origine.

Sur le plan historique, le compact K est le premier exemple d'attracteur uniformément hyperbolique, qui n'est pas une sous-variété. Il est construit par S. Smale en 1972. Comme la transformation f^{-1} s'obtient à partir d'un automorphisme du tore, qui est l'exemple le plus simple de difféomorphisme Anosov, elle est qualifiée de *difféomorphisme dérivé d'Anosov*. On peut réaliser ce genre de construction sur toute transformation possédant un point fixe hyperbolique.

Perturbation d'un automorphisme du tore

On part de la matrice $A = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$. Le nombre d'or est noté $\lambda = \frac{1+\sqrt{5}}{2} \simeq 1.618$. Cette matrice admet deux valeurs propres λ^2 et λ^{-2} ; les vecteurs propres associés sont $\mathbf{e}_u = \frac{1}{\sqrt{1+\lambda^2}} \begin{pmatrix} \lambda \\ 1 \end{pmatrix}$ et $\mathbf{e}_s = \frac{1}{\sqrt{1+\lambda^2}} \begin{pmatrix} -1 \\ \lambda \end{pmatrix}$, et on a la relation :

$$\begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} = \frac{1}{\sqrt{1+\lambda^2}} \begin{pmatrix} \lambda & -1 \\ 1 & \lambda \end{pmatrix} \begin{pmatrix} \lambda^2 & 0 \\ 0 & \lambda^{-2} \end{pmatrix} \frac{1}{\sqrt{1+\lambda^2}} \begin{pmatrix} \lambda & 1 \\ -1 & \lambda \end{pmatrix}$$

Perturbons A de façon à rendre le point 0 attractif. Pour $(x, y) \in [-\frac{1}{2}, \frac{1}{2}]^2$, posons :

$$\begin{aligned} f \begin{pmatrix} x \\ y \end{pmatrix} &= \frac{1}{\sqrt{1+\lambda^2}} \begin{pmatrix} \lambda & -1 \\ 1 & \lambda \end{pmatrix} \begin{pmatrix} \lambda^2 + p_1 k(r/a) & 0 \\ 0 & \lambda^{-2} \end{pmatrix} \frac{1}{\sqrt{1+\lambda^2}} \begin{pmatrix} \lambda & 1 \\ -1 & \lambda \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \\ &= \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \frac{p_1}{1+\lambda^2} k(r/a) \begin{pmatrix} \lambda^2 & \lambda \\ \lambda & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \end{aligned}$$

avec $r = \sqrt{x^2 + y^2}$ et $k(r) = (1 - r^2)^2 \mathbf{1}_{[-1,1]}(r)$ utilisé comme "bosse" C^1 . Le paramètre a contrôle l'étendue de la perturbation, le paramètre p_1 son amplitude. Lorsque $a \in [0, 1/2]$, f passe au quotient et définit une application de $\mathbf{T}^2$ dans $\mathbf{T}^2$.

Propriétés

- Pour tout $(x, y) \in \mathbf{T}^2$, $f((x, y) + \mathbf{R}\mathbf{e}_u) \subset f(x, y) + \mathbf{R}\mathbf{e}_u$.
- Pour $p_1 \in]-\lambda^2, 0]$ et $a \in [0, 1/2]$, f est un difféomorphisme du tore $\mathbf{T}^2$.
- Pour $p_1 \in]-\lambda^2, 1-\lambda^2]$, 0 est un point fixe attractif. Soit U son bassin d'attraction.
- Pour $p_1 \in]-\lambda^2, 1-\lambda^2]$, f admet un point fixe $p \in]0, a[\mathbf{e}_u$ tel que $[0, p[\subset U$.
- $B(0, |p|)$ est inclus dans U .
- Pour tout $(x, y) \in U^c$, $|d_{(x,y)} f \cdot \mathbf{e}_u| > 1$.

Preuve

Pour tout $(x, y) \in \mathbf{T}^2$, $f(x, y) - A(x, y)$ appartient à $\mathbf{R}\mathbf{e}_u$. Par conséquent, $f((x, y) + t\mathbf{e}_u) - A((x, y) + t\mathbf{e}_u) \in \mathbf{R}\mathbf{e}_u$ et donc $f((x, y) + t\mathbf{e}_u) - f(x, y) \in \mathbf{R}\mathbf{e}_u$.

Calculons le jacobien de f dans la base orthonormale $(\mathbf{e}_u, \mathbf{e}_s)$:

$$\det(df) = \frac{\partial}{\partial x} (x + \lambda^{-2} p_1 x k(r/a)) = 1 + \lambda^{-2} p_1 k(r/a) + \lambda^{-2} p_1 \frac{x^2}{ra} k'(r/a) \geq 1 + \lambda^{-2} p_1$$

L'application f est donc un difféomorphisme local ; comme 0 a une seule pré-image dans $\mathbf{T}^2$, f est bijective. Le point fixe 0 est attractif pour les valeurs données plus haut car la différentielle $D_0 f$ admet comme valeurs propres λ^{-2} et $\lambda^2 + p_1$.

L'application $h(t) = \lambda^2 t + p_1 t k(t/a)$ admet un point fixe dans l'intervalle $]0, a[$ car $h(0) = 0$, $h'(0) \in]0, 1[$ et $h(a) > a$. Soit t_0 le premier point fixe de h dans $]0, a[$. On pose $p = t_0 \mathbf{e}_u$ et on remarque que $\lambda^2 + p_1 k(|p|/a) = 1$.

Montrons que $|f(x, y)| < |(x, y)|$ si $|(x, y)| < |p|$. On se place dans la base $(\mathbf{e}_u, \mathbf{e}_s)$. $|f(x, y)|^2 = \lambda^{-4} y^2 + (\lambda^2 x + p_1 x k(r/a))^2 < \lambda^{-4} y^2 + (\lambda^2 x + p_1 x k(|p|/a))^2 = \lambda^{-4} y^2 + x^2$. On a utilisé la décroissance stricte de k sur $]0, a[$ et l'égalité $p_1 k(|p|/a) = 1 - \lambda^2$. La fonction $(x, y) \mapsto \frac{f(x, y)}{|(x, y)|}$ atteint son maximum sur la couronne $\{\varepsilon \leq |m| \leq |p| - \varepsilon\}$. f est donc contractante sur cette couronne. Tout point de $B(0, |p|)$ a une orbite qui finit par entrer dans $B(0, \varepsilon)$. L'orbite du point converge donc vers 0.

Calculons la différentielle dans la direction $\mathbf{e}_u$ en se plaçant dans la base $(\mathbf{e}_u, \mathbf{e}_s)$.

$$d_{(x,y)} f \cdot \mathbf{e}_u = \lambda^2 + p_1 k(r/a) + p_1 \frac{x^2}{ra} k'(r/a) = 1 - p_1 (k(|p|/a) - k(r/a)) + p_1 \frac{x^2}{ra} k'(r/a)$$

Ceci est supérieur à 1 si $r \geq |p|$ et égal à 1 si $(x, y) = (0, |p|)$. Ce point est dans U .

Dans la suite, on prend $p_1 = -2.236$ et $a = 0.5$. Le bassin d'attraction de 0 est noté U , le complémentaire de U est noté K . Enfin, on fixe une linéarisation φ d'un voisinage V de p sur $]0, 1]^2$.

Théorème Soit $x \in K$. Alors pour tout $\varepsilon > 0$, $x - [0, \varepsilon] \mathbf{e}_u$ rencontre U . L'ouvert U est donc dense dans $\mathbf{T}^2$ et K est d'intérieur vide.

Preuve

Si $x - [0, \varepsilon] \mathbf{e}_u$ ne rencontre pas U , il en va de même pour tous ses itérés. Comme $D_y f \cdot \mathbf{e}_u > 1$ si $y \in K$, ces itérés sont de la forme $f^n(x) - [0, c_n \varepsilon] \mathbf{e}_u$, avec $c_n \geq c^n > 1$.

Comme l'ensemble $\mathbf{R}_+ \mathbf{e}_u$ est dense dans $\mathbf{T}^2$, on peut trouver $n \in \mathbf{N}$ tel que tout point de $\mathbf{T}^2$ soit à distance inférieure à $|p|$ de $[0, c^n \varepsilon] \mathbf{e}_u$. Le point 0 est à distance inférieure à $|p|$ de $f^n(x) - [0, c^n \varepsilon] \mathbf{e}_u$, d'où une contradiction.

Théorème L'ensemble $p + \mathbf{R}_+ \mathbf{e}_u$ est inclus dans $W^{su}(p)$, qui est dense dans $\mathbf{T}^2$.

Preuve

Soit $t_1 = \inf\{t \in \mathbf{R}_+ \mid p + t\mathbf{e}_u \notin W^{su}(p)\}$. D'après l'allure de $t \mapsto f(p + t\mathbf{e}_u)$, t_1 est strictement positif. L'image de $p + [0, t_1] \mathbf{e}_u$ par f est de la forme $p + [0, s] \mathbf{e}_u$. Comme $W^{su}(p)$ est invariant par f , $s = t_1$ et $p' = p + t_1 \mathbf{e}_u$ est un point fixe de f .

Vérifions que $p' \neq 0$: comme $p \in \mathbf{R}_+ \mathbf{e}_u$, on aurait $0 \in \mathbf{R}^* \mathbf{e}_u$, ce qui contredit l'irrationalité de λ . On en déduit que $p' \notin U$; la pente de la courbe $t \mapsto f(p + t\mathbf{e}_u)$ en t_1 est donc supérieure à 1. Par conséquent, on devrait avoir des points dont les itérés négatifs approchent à la fois p et p' , ce qui est absurde.

Théorème Soit $m \in U$ et t tel que $m + [0, t] \mathbf{e}_u \subset U$, $m + t\mathbf{e}_u \notin U$. Alors $m + t\mathbf{e}_u$ appartient à $W^{ss}(p)$. De plus, l'ensemble $W^{ss}(p) \cap W^{su}(p)$ est dense dans K .

Preuve

Soit $\varphi : V \rightarrow]-1, 1[$ un ouvert de linéarisation autour de p . Comme $W^{su}(p) \cap U$ contient $]0, p[$, on peut trouver $x' \in]0, p[\cap V$ tel que $[x', f(x')]$ soit dans $U \cap V$. Il existe donc dans $\varphi(V)$ un rectangle $[-\delta, \delta] \times [x', f(x')]$ contenu dans $\varphi(U)$. Ses itérés positifs sous l'action de $D_p f^{-1}$ sont aussi dans $\varphi(U)$ et recouvrent $[-\delta, \delta] \times [x', p[$; cf fig 2. Considérons une courbe dans l'ouvert $[-\delta, \delta] \times [x', -x'[$ issue du demi-plan inférieur, qui n'est pas tout entière contenue dans $\varphi(U)$. Le premier point de la courbe qui n'est pas dans $\varphi(U)$ doit se situer sur l'axe des x , c'est-à-dire sur $\varphi(W^{ss}(p))$.

Pour n grand, l'itéré $f^n(m + [0, t] \mathbf{e}_u)$ est une droite issue d'un petit voisinage de 0 et dirigée selon $\mathbf{e}_u$. Le premier point de la courbe qui appartient à K doit donc se trouver dans V et appartient à $W^{ss}(p)$.

Soit $m' \in K$ et $\varepsilon > 0$ tel que $m = m' - \varepsilon \mathbf{e}_u$ soit dans U . On peut trouver $a > 0$ tel que $p + a \mathbf{e}_u$ soit arbitrairement proche de m . Reprenant le raisonnement précédent, nous voyons que l'itéré $f^n(p + [a, a + 2t] \mathbf{e}_u)$ est proche de $f^n(m + [0, 2t] \mathbf{e}_u)$. Il coupe donc $W^{ss}(p)$ en un point x tel que $f^{-n}(x)$ soit aussi proche qu'on veut de $m + t \mathbf{e}_u$.

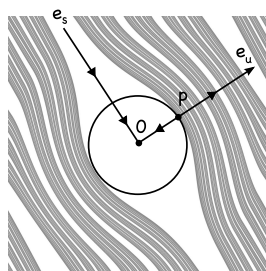
Corollaire L'application f est transitive en restriction à K .

Preuve

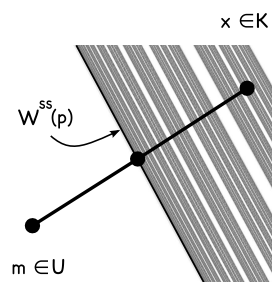
Soit U_1 un ouvert intersectant K , $x_1 \in W^{ss}(p) \cap W^{su}(p) \cap U_1$ et n_1 tel que $f^{-n_1}(x_1)$ est dans V pour tout $n \geq n_1$. Posons $x'_1 = f^{-n_1}(x_1)$. Montrons que tout segment orienté selon $\mathbf{e}_u$, passant près de x'_1 , coupe $W^{ss}(x'_1)$ au voisinage de x'_1 . Soit $R \subset V$ un petit rectangle centré en x'_1 et orienté selon $\mathbf{e}_s$ et $\mathbf{e}_u$. On peut trouver N tel que pour tout $n \geq N$, $f^n(x'_1)$ est dans V . L'ensemble $f^N(R)$ contient un petit rectangle R' centré sur $f^N(x'_1)$ et quitte à augmenter N , on peut supposer que R' traverse de bas en haut l'ouvert V . Comme $f^N(x'_1)$ est dans $\varphi(]-1, 1[\times \{0\})$, les verticales de R' coupent $\varphi(]-1, 1[\times \{0\}) \subset W^{ss}(f^N(x'_1))$ au voisinage de $f^N(x'_1)$. cf fig.3.

Soit U_2 un autre ouvert intersectant K . Il suffit de construire un point $x' \in K$ possédant un itéré négatif dans U_2 et un itéré positif dans U_1 pour obtenir la transitivité. Soit $x_2 \in U_2 \cap W^{ss}(p)$ et $\varepsilon > 0$ tel que $x_2 + [-\varepsilon, \varepsilon] \mathbf{e}_u$ est dans U_2 . Pour n grand, $f^n(x_2 + [-\varepsilon, \varepsilon] \mathbf{e}_u)$ est un segment dirigé par $\mathbf{e}_u$ proche de p qui traverse V de bas en haut. Il coupe donc $W^{ss}(x'_1) \cap f^{n_1}(U_1)$ en un point x' qui est dans K .

figure 1 : Bassin d'attraction de l'origine



O est point fixe attractif
p est point fixe hyperbolique



Un segment dirigé selon e_u
qui joint U à K, doit couper
la variété stable de p.

figure 2 : Etude de la variété stable de p

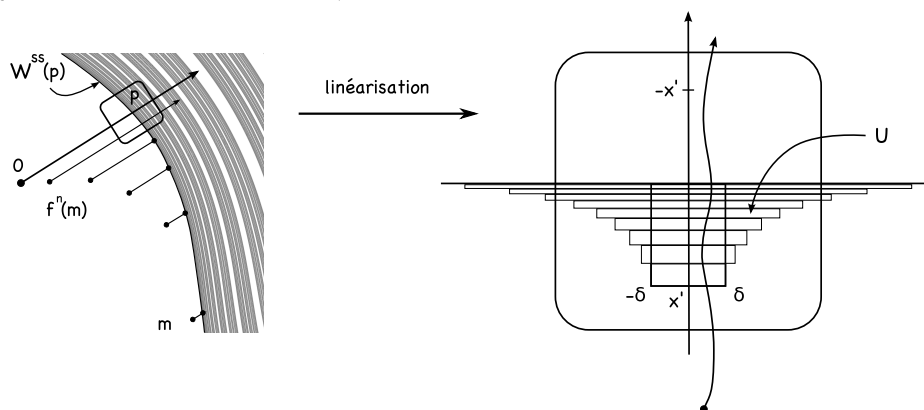
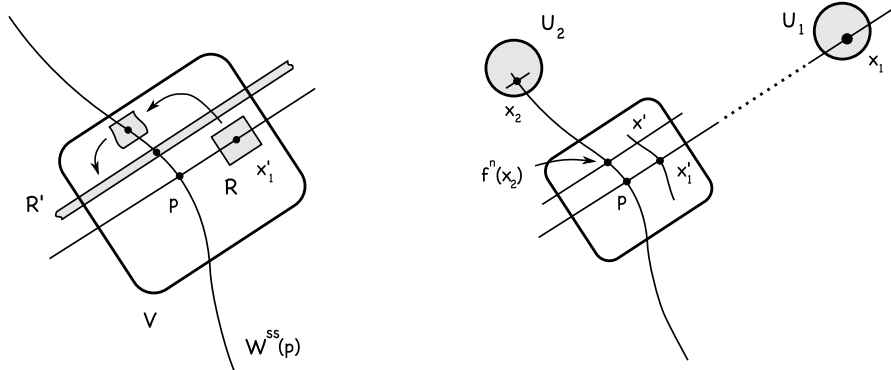


figure 3 : Preuve de la transitivité



Exercices

Les exercices 1 à 9 font référence au difféomorphisme f qui vient d'être étudié.

exercice 1 :

Montrez que K est compact connexe non dénombrable.

exercice 2 :

Montrez que pour tout $m \in \mathbf{T}^2$, $f(-m) = -f(m)$. En déduire qu'il existe un point fixe hyperbolique $p' \in [-a, 0] \mathbf{e}_u$ et que $W^{ss}(p')$ est dense dans K .

exercice 3 :

Montrez que $W^{ss}(p)$ est d'intérieur vide dans K . On pourra remarquer qu'aucun point de $W^{ss}(p)$ ne possède d'orbite dense.

exercice 4 :

Soit $\varepsilon > 0$. Montrez que $(p + [0, \varepsilon] \mathbf{e}_u) \cap K$ est compact, sans point isolé, et d'intérieur vide dans $p + [0, \varepsilon] \mathbf{e}_u$. En déduire qu'il est non dénombrable.

exercice 5 :

Montrez que les points de $W^{ss}(p)$ sont accessibles à partir de U au sens suivant : pour tout $x \in W^{ss}(p)$, on peut trouver une application continue $\gamma : [0, 1] \rightarrow \mathbf{T}^2$ telle que $\gamma([0, 1]) \subset U$ et $\gamma(1) = x$.

exercice 6 :

Montrez que K n'est pas localement connexe. On pourra remarquer que tout voisinage de p contient un point de U sur $W^{su}(p)$, et itérer un voisinage de ce point.

exercice 7 :

Soit $\gamma : [0, 1] \rightarrow K$ une application continue qui part de $p : \gamma(0) = p$. On se place dans un ouvert de linéarisation V autour de p et on considère une portion d'arc $\gamma([0, \delta])$ contenue dans V . Montrez que $\gamma([0, \delta]) \subset W^{ss}(p)$.

Pouvez-vous en déduire que $\gamma([0, 1]) \subset W^{ss}(p)$?

exercice 8 :

Soit $\gamma : [0, 1] \rightarrow \mathbf{T}^2$ une application continue qui satisfait $\gamma([0, 1]) \subset W^{ss}(p)$. Montrez que $\gamma(1) \in W^{ss}(p)$. On pourra raisonner par l'absurde et montrer que $\gamma(1)$ est un point fixe hyperbolique.

exercice 9 * :

Montrez que K n'est pas connexe par arc. On pourra montrer que p et $-p$ ne peuvent pas être reliés par un arc qui reste dans K .

exercice 10 :

Soit M une variété différentielle, $f : M \rightarrow M$ un difféomorphisme C^1 qui possède un point fixe hyperbolique p . On suppose que $D_p f$ possède une unique valeur propre de module inférieur à un, et que cette valeur propre est réelle positive. Montrez que $W^{ss}(p) \setminus \{p\}$ possède deux composantes connexes.

exercice 11 :

Montrez qu'on peut recoller deux systèmes dérivé d'Anosov de manière à obtenir un difféomorphisme f défini sur une surface de genre deux, et dont l'ensemble non-errant est union de deux compacts connexes non dénombrables K_1, K_2 en restriction auxquels f est transitive.

exercice 12 :

Soit M une variété différentielle, $f : M \rightarrow M$ un difféomorphisme C^1 qui possède un point fixe hyperbolique p . Montrez que si $W^{ss}(p)$ et $W^{su}(p)$ sont denses dans M , alors f est transitive.

Commentaires

La perturbation f étudiée dans ce chapitre est C^1 avec dérivée Lipschitzienne. On aurait pu construire une application C^∞ en prenant une fonction “bosse” du type $k(r) = \exp(-\frac{1}{1-r^2})\mathbf{1}_{[-1,1]}(r)$. Du point de vue numérique, une “bosse” polynomiale semble préférable.

Voici trois algorithmes qui permettent de visualiser le compact K :

- On choisit un point x au hasard et on l’itère un million de fois à l’aide de l’application f^{-1} . Si le point est différent de l’origine, sa trajectoire va venir s’accumuler sur l’attracteur K . La transformation f^{-1} est transitive en restriction à K . Pour la plupart des x , la trajectoire devrait donc venir s’approcher de tous les points de l’attracteur. C’est ce qui est observé en pratique. Cette méthode est la plus rapide du point de vue numérique.
- L’origine est point fixe attractif pour l’application f et K est le complémentaire de son bassin d’attraction. Pour visualiser ce bassin, on fixe un petit disque centré à l’origine, puis on colore les points du plan en fonction du nombre d’itérations nécessaire pour rejoindre ce petit disque. En pratique, la plupart des points atteignent le disque en moins de soixante dix étapes. On pourra par exemple colorer en noir tous les points nécessitant plus de vingt itérations, ce qui permet de représenter un petit voisinage de K .
- On peut montrer que les points périodiques de f sont denses dans K . L’ensemble des points périodiques de période inférieure à n , pour n suffisamment grand, constitue donc une bonne approximation de K . Le calcul des points périodiques s’avère particulièrement coûteux sur le plan numérique, cette méthode est donc peu recommandée.

On peut décrire la dynamique de f en restriction à K de manière plus précise. Sur K , la transformation f est semi-conjugée à un décalage de type fini topologiquement mélangeant. R. F. Williams (1974) montre qu’elle est conjuguée à un décalage sur un solénoïde généralisé. La dynamique est donc fortement “stochastique”.

Le compact K est localement homéomorphe au produit d’un segment avec un ensemble de Cantor. On le vérifie au voisinage de p en montrant que l’intersection de K avec la feuille instable locale de p est d’intérieur vide dans K . Pour le démontrer au voisinage de tout point x de K , il faut étudier plus en détail la structure des feuilles stables $W^{ss}(x)$, et montrer que ce sont des sous-variétés immergées de dimension 1.

Les points de $W^{ss}(p)$ et de $W^{ss}(-p)$ constituent le bord accessible de U . Ce sont les seuls points de ∂U qui sont extrémités d’une courbe $\gamma : [0, 1] \rightarrow \mathbf{T}^2$ contenue dans U pour $t \in [0, 1[$. Cette notion de bord accessible correspond sans doute mieux à l’idée intuitive qu’on peut se faire du bord d’un ensemble.

Le difféomorphisme dérivé d’Anosov est un exemple de difféomorphisme Axiom A : ses points périodiques forment un sous-ensemble dense de l’ensemble non-errant $\{0\} \cup K$; en restriction à cet ensemble non-errant, l’espace tangent peut-être décomposé en la somme de deux sous-fibrés invariants, respectivement contractés et dilatés par la différentielle de l’application.

L’ensemble non-errant d’un difféomorphisme Axiom A se décompose en un nombre fini de compacts invariants, en restriction desquels la transformation est transitive. Ceci se démontre en étudiant les feuilles stables et instables des points périodiques, comme pour le dérivé d’Anosov.

Entropie

The sun comes up just about as often as it goes down, in the long run, but this doesn't make its motion random.

D. Knuth

Intéressons nous au problème de la conjugaison sur le plan mesurable. Considérons deux systèmes mesurés donnés par une application $T_1 : X_1 \rightarrow X_1$ préservant une mesure μ_1 , et une application $T_2 : X_2 \rightarrow X_2$ préservant une mesure μ_2 . Ces deux systèmes sont *isomorphes* si on peut trouver deux sous-ensembles de X_1 et X_2 chacun de complémentaire négligeable, et une bijection mesurable φ entre ces deux ensembles, d'inverse mesurable, qui satisfait $\varphi \circ T_1 = T_2 \circ \varphi$ et qui envoie μ_1 sur $\mu_2 : \varphi_*\mu_1 = \mu_2$.

Quelles sont les propriétés préservées par isomorphisme ? Si T_1 est ergodique ou mélangeante, il en va de même pour T_2 . La masse de la mesure, le nombre moyen de pré-images, sont des quantités invariantes par isomorphisme. On peut construire d'autres invariants en faisant opérer les applications T_1 et T_2 sur l'espace L^2 par composition : on pose, pour tout $f \in L^2$, $U_1 f = f \circ T_1$, $U_2 f = f \circ T_2$. Si T_1 et T_2 sont isomorphes, les applications U_1, U_2 sont conjugués par le biais d'un opérateur unitaire. Elles ont donc même valeurs propres, ce qui donne de nouveaux invariants numériques.

On peut, à l'aide de ces invariants, classer les rotations du cercle ou les translations ergodiques des tores à isomorphisme près. Deux tels systèmes sont isomorphes dès que les applications induites sur L^2 sont conjuguées, et ceci se produit si et seulement si elles ont mêmes valeurs propres. La situation est plus complexe pour les automorphismes hyperboliques des tores ou les décalages de Bernoulli sur un alphabet fini. On peut montrer que les applications linéaires induites sur L^2 sont toutes conjuguées entre elles, mais cette conjugaison n'est pas obtenue à l'aide d'un isomorphisme. La question de la classification de ces systèmes à isomorphisme près, qui est entreprise dès les années 30, ne sera finalement résolue qu'en 1970 par D. Ornstein.

Une étape importante dans cette résolution, est la construction par A. N. Kolmogorov en 1958 d'un nouvel invariant d'isomorphisme, appelé *entropie*. Cet invariant numérique mesure la quantité de "hasard" contenue dans le système. Voici comment il est défini. On commence par partitionner l'espace en un nombre fini de morceaux, puis on itère chacun des morceaux, de manière à découper l'espace en morceaux de taille de plus en plus petite. Si la partition initiale a été bien choisie, la mesure des morceaux tend vers 0 et l'entropie s'obtient à partir du taux de décroissance moyen observé ; cf fig.2 et 3. Ce taux est polynomial pour les rotations mais exponentiel pour les décalages de Bernoulli, et c'est l'exposant qui permet de les différencier.

Le taux de décroissance dépend à priori de la partition choisie initialement. Ceci induit un certain nombre de complications techniques, et fait de l'entropie un concept délicat à définir et à calculer.

Entropie

Soit $(X, \mathcal{T}, \mu)$ un espace probabilisé. Une *partition* (finie, dénombrable) ξ de X est la donnée d'un ensemble (fini, dénombrable) de parties mesurables de X disjointes deux à deux, et dont l'union recouvre presque tout X . On notera $\xi(x)$ l'élément de la partition contenant le point $x \in X$.

Soit $T : X \rightarrow X$ une application mesurable et $\xi = \{A_i\}_{i=1\dots n}$ une partition. On pose $T^{-1}\xi = \{T^{-1}(A_i)\}_{i=1\dots n}$. Soit $\xi_1, \xi_2, \dots, \xi_n$ des partitions de X . La *partition engendrée* par les ξ_i est définie par la relation : $\bigvee_{i=1}^n \xi_i(x) = \bigcap_{i=1}^n \xi_i(x)$; cf fig.1.

Définition

Soit $(X, \mathcal{T}, \mu)$ un espace probabilisé, ξ une partition dénombrable de X . L'information et l'entropie de ξ sont définies par :

$$I(\xi)(x) = -\log(\mu(\xi(x))) = \sum_{A \in \xi} -\log(\mu(A)) \mathbf{1}_A(x)$$

$$H(\xi) = \int_X I(\xi) d\mu = \sum_{A \in \xi} -\mu(A) \log(\mu(A))$$

L'information $I(\xi)$ est une fonction mesurable positive ; l'entropie $H(\xi)$ est un réel positif, qui peut être infini si la partition n'est pas finie. Pour calculer ces quantités, on introduit une version conditionnelle de l'entropie.

Soit $\mathcal{A} \subset \mathcal{T}$ une tribu. L'entropie conditionnelle de ξ relativement à $\mathcal{A}$ est donnée par :

$$H(\xi | \mathcal{A}) = - \sum_{A \in \xi} \int E_\mu(\mathbf{1}_A | \mathcal{A}) \log(E_\mu(\mathbf{1}_A | \mathcal{A})) d\mu$$

Soit η une partition dénombrable ; on note $H(\xi | \eta)$ l'entropie conditionnelle de ξ relativement à la tribu engendrée par les éléments de η .

Propriétés

Soit ξ, ξ_1, ξ_2, η des partitions dénombrables, $T : X \rightarrow X$ une application mesurable qui préserve la mesure μ et $\mathcal{A}, \mathcal{B}$ des sous-tribus de $\mathcal{T}$.

- $H(\xi | \mathcal{T}) = 0$, $H(\xi | \{X\}) = H(\xi)$.
- Si $\mathcal{A} \subset \mathcal{B}$, alors $H(\xi) \geq H(\xi | \mathcal{A}) \geq H(\xi | \mathcal{B}) \geq 0$.
- $H(\xi_1 \vee \xi_2 | \eta) = H(\xi_1 | \xi_2 \vee \eta) + H(\xi_2 | \eta)$; en particulier, $H(\xi_1 \vee \xi_2) \leq H(\xi_1) + H(\xi_2)$.

Preuve

- C'est une conséquence des égalités : $E(\mathbf{1}_A | \{X\}) = \mu(A)$ et $E(\mathbf{1}_A | \mathcal{T}) = \mathbf{1}_A$.
- La fonction $f(x) = x \log x$ est strictement convexe. L'inégalité de Jensen donne :

$$f(E(\mathbf{1}_A | \mathcal{A})) \leq E(f(E(\mathbf{1}_A | \mathcal{B})) | \mathcal{A})$$

Par conséquent, $E(f(E(\mathbf{1}_A | \mathcal{A}))) \leq E(E(f(E(\mathbf{1}_A | \mathcal{B})) | \mathcal{A})) = E(f(E(\mathbf{1}_A | \mathcal{B})))$. Il suffit de sommer sur tous les éléments A de ξ pour obtenir l'inégalité recherchée.

- Posons $\xi = \xi_1 \vee \xi_2 = \{A_i\}$ et $\eta = \{B_j\}$. Montrons que $H(\xi | \eta) = H(\xi \vee \eta) - H(\eta)$.
 $H(\xi \vee \eta) - H(\xi | \eta) = \sum_{i,j} \mu(A_i \cap B_j) \log(\mu(A_i \cap B_j)) - \sum_{i,j} \mu(A_i \cap B_j) \log\left(\frac{\mu(A_i \cap B_j)}{\mu(A_i)}\right)$
 $= - \sum_{i,j} \mu(A_i \cap B_j) \log(\mu(A_i)) = - \sum_i \mu(A_i) \log(\mu(A_i)) = H(\xi)$.
 $H(\xi_1 | \xi_2 \vee \eta) + H(\xi_2 | \eta) = H(\xi_1 \vee \xi_2 \vee \eta) - H(\xi_2 \vee \eta) + H(\xi_2 \vee \eta) - H(\eta) = H(\xi_1 \vee \xi_2 | \eta)$.

Définition

Soit $(X, \mathcal{T}, \mu)$ un espace probabilisé, $T : X \rightarrow X$ une application mesurable qui préserve la mesure μ et ξ une partition dénombrable de X . L'entropie de T relativement à la partition ξ , et l'entropie de T , sont définies par :

$$h_\mu(T, \xi) := \lim_{N \rightarrow \infty} \frac{1}{N} H\left(\bigvee_{i=0}^{N-1} T^{-i}\xi\right)$$

$$h_\mu(T) = \sup \{ h_\mu(T, \xi) \mid \xi \text{ partition finie de } X \}$$

La suite $\frac{1}{n}H\left(\bigvee_{k=0}^{n-1} T^{-k}\xi\right)$ converge car elle est décroissante. On le démontre plus loin.

Propriétés

Soit ξ, η deux partitions dénombrables, $T : X \rightarrow X$ une application mesurable qui préserve la mesure μ , et $\mathcal{A}, \mathcal{B}$ des sous-tribus de $\mathcal{T}$.

- $H(\xi | \mathcal{A}) = H(T^{-1}\xi | T^{-1}\mathcal{A})$.
- $h_\mu(T, \eta) \leq h_\mu(T, \xi) + H(\eta | \xi)$.
- $h_\mu(T, \xi) = \lim_{n \rightarrow \infty} H(\xi | \bigvee_1^n T^{-i}\xi)$.
- Pour tout $n \in \mathbf{N}$, on a : $h_\mu(T, \xi) = h_\mu(T, \bigvee_{i=0}^n T^{-i}\xi)$.

Preuve

- Ceci découle de la relation $E_\mu(T^{-1}A | T^{-1}\mathcal{A}) = E_\mu(A | \mathcal{A}) \circ T$.
- On a : $H(\bigvee_0^n T^{-i}\eta) \leq H(\bigvee_0^n T^{-i}\eta | \bigvee_0^n T^{-i}\xi) = H(\bigvee_0^n T^{-i}\xi) + H(\bigvee_0^n T^{-i}\eta | \bigvee_0^n T^{-i}\xi)$
 $H(\bigvee_0^n T^{-i}\eta | \bigvee_0^n T^{-i}\xi) \leq \sum_i H(T^{-i}\eta | T^{-i}\xi) \leq \sum_i H(T^{-i}\eta | T^{-i}\xi) = (n+1)H(\eta | \xi)$
L'inégalité recherchée s'en déduit en divisant par $n+1$ et en passant à la limite.
- On a : $H(\bigvee_0^n T^{-i}\xi) = H(\bigvee_0^n T^{-i}\xi) + H(\xi | \bigvee_1^n T^{-i}\xi) = H(\bigvee_0^{n-1} T^{-i}\xi) + H(\xi | \bigvee_1^n T^{-i}\xi)$
Ceci implique : $H(\bigvee_{k=0}^n T^{-k}\xi) = H(\xi) + \sum_{j=1}^n H(\xi | \bigvee_{k=1}^j T^{-k}\xi)$. Comme la suite $H(\xi | \bigvee_{k=1}^n T^{-k}\xi)$ est décroissante, il en va de même de sa moyenne $\frac{1}{n+1}H(\bigvee_{k=0}^n T^{-k}\xi)$ et ces deux suites ont même limite.
- $h(T, \bigvee_{i=0}^n T^{-i}\xi) = \lim \frac{1}{N}H\left(\bigvee_{j=0}^{N-1} \bigvee_{i=0}^n T^{-i-j}\xi\right) = \lim \frac{1}{N+n}H\left(\bigvee_{k=0}^{N+n-1} T^{-k}\xi\right) = h(T, \xi)$.

Théorème (Kolmogorov-Sinaï 1958)

Soit $(X, \mathcal{T}, \mu)$ un espace probabilisé, $T : X \rightarrow X$ une application mesurable qui préserve la mesure μ et ξ une partition finie de X . La partition ξ est dite *génératrice unilatère* si la tribu engendrée par les éléments des partitions $T^{-i}\xi$, $i \in \mathbf{N}$, et les ensembles négligeables est égale à $\mathcal{T}$. Dans ce cas :

$$h_\mu(T) = h_\mu(T, \xi) = H(\xi | T^{-1}\mathcal{T})$$

Preuve

Soit ξ_n une suite de partitions dénombrables satisfaisant $\xi_{n+1}(x) \subset \xi_n(x)$ pour tout $n \in \mathbf{N}$ et tout $x \in X$. Soit $\mathcal{A}$ la tribu engendrée par les éléments de tous les ξ_n . Soit enfin η une partition finie de X . Montrons que $H(\eta | \xi_n) \rightarrow H(\eta | \mathcal{A})$.

Soit A un élément de η . Par le théorème de convergence L^2 des Martingales, on a : $E_\mu(\mathbf{1}_A | \xi_n) \rightarrow E_\mu(\mathbf{1}_A | \mathcal{A})$. Quitte à extraire, on peut trouver une sous-suite n_i qui converge presque partout. Ceci implique :

$$E_\mu(\mathbf{1}_A | \xi_{n_i}) \log(E_\mu(\mathbf{1}_A | \xi_{n_i})) \rightarrow E_\mu(\mathbf{1}_A | \mathcal{A}) \log(E_\mu(\mathbf{1}_A | \mathcal{A})) \quad \text{pp.}$$

Le terme de gauche est borné par e^{-1} . Le théorème de convergence dominée entraîne :

$$\int E_\mu(\mathbf{1}_A | \xi_{n_i}) \log(E_\mu(\mathbf{1}_A | \xi_{n_i})) \, d\mu \longrightarrow \int E_\mu(\mathbf{1}_A | \mathcal{A}) \log(E_\mu(\mathbf{1}_A | \mathcal{A})) \, d\mu$$

Sommons sur tous les éléments A de η ; on obtient : $H(\eta | \xi_{n_i}) \rightarrow H(\eta | \mathcal{A})$. Comme la suite $H(\eta | \xi_n)$ est décroissante, on a la relation recherchée.

Si on prend : $\xi_n = \bigvee_{i=1}^n T^{-i}\xi$, on obtient : $h_\mu(T, \xi) = \lim H(\xi | \xi_n) = H(\xi | T^{-1}\mathcal{T})$.

Si on prend : $\xi_n = \bigvee_{i=0}^n T^{-i}\xi$, on obtient : $\lim H(\eta | \xi_n) = H(\eta | \mathcal{T}) = 0$, et donc :

$$h_\mu(T, \eta) \leq h_\mu(T, \xi_n) + H(\eta | \xi_n) = h_\mu(T, \xi) + H(\eta | \xi_n) \longrightarrow h_\mu(T, \xi).$$

Remarque : Lorsque T est inversible, on montre de même que $h_\mu(T) = h_\mu(T, \xi)$ dès que l'ensemble de tous les éléments des partitions $T^{-i}\xi$, $i \in \mathbf{Z}$, engendrent $\mathcal{T}$. On dit dans ce cas que ξ est une *partition génératrice bilatère*.

figure 1 : Partition engendrée

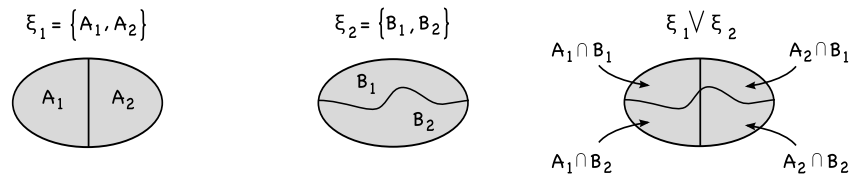


figure 2 : Itération d'une partition du tore

On considère la partition

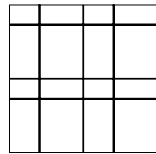
$$\xi = \{ [0, 1/2[^2, [0, 1/2[\times [1/2, 1[, [1/2, 1[\times [0, 1/2[, [1/2, 1[^2 \}$$

et ses itérés $\bigvee_{k=0}^n T^{-k} \xi$

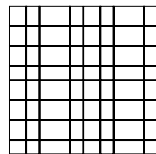
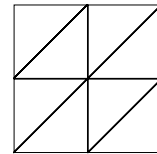
par les applications suivantes :

$$\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} x + \sqrt{2} \\ y + \sqrt{3} \end{pmatrix}$$

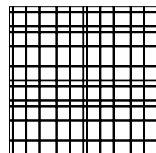
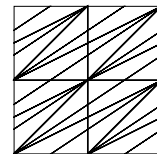
$$\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$



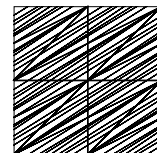
n=1



n=3



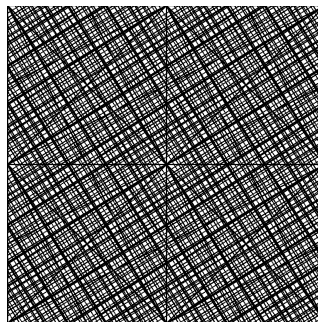
n=5



A l'étape n , on obtient $(2n+2)^2$ morceaux sensiblement de même taille pour la translation, tandis que pour l'automorphisme du tore, on obtient 2^{n+2} morceaux.

figure 3 : Itération de la partition ξ par l'automorphisme $\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$

$$\bigvee_{k=-7}^7 T^{-k} \xi$$



Exercices

Dans la suite, $(X, \mathcal{T}, \mu)$ est un espace probabilisé.

exercice 1 :

On considère l'application $T : [0, 1[\rightarrow [0, 1[$ donnée par $T(x) = 2x \bmod 1$. Cette application préserve la mesure de Lebesgue. Soit $\xi = \{[0, 1/2[, [1/2, 1[\}$.

- Calculez $T^{-1}\xi$, puis $\xi \vee T^{-1}\xi$.
- Calculez $H(\xi)$, puis $H(\xi \vee T^{-1}\xi)$.
- Montrez par récurrence la formule : $\bigvee_0^{n-1} T^{-i}\xi = \{[\frac{l-1}{2^n}, \frac{l}{2^n}[\mid l \in \{1 \dots 2^n\}\}$
- Calculez $H(\bigvee_0^{n-1} T^{-i}\xi)$ et $h(T, \xi)$.
- Montrez que la partition ξ est génératrice. En déduire $h(T)$.

exercice 2 :

Soit ξ une partition finie de X . Montrez l'inégalité $H(\xi) \leq \ln(\text{Card}(\xi))$. Montrez qu'on a égalité si et seulement si tous les éléments de ξ ont la même mesure.

exercice 3 :

Soit ξ et η deux partitions dénombrables de X . Montrez que $H(\xi \mid \eta) = 0$ si et seulement si pour presque tout $x \in X$, $\eta(x) \subset \xi(x)$.

Supposons $H(\xi) < \infty$ et $H(\eta) < \infty$. Montrez que l'égalité $H(\xi \vee \eta) = H(\xi) + H(\eta)$ est satisfaite si et seulement si les partitions ξ et η sont indépendantes :

$$\forall A \in \xi, \forall B \in \eta, \quad \mu(A \cap B) = \mu(A)\mu(B).$$

exercice 4 :

Soit $(X, \mathcal{T}, \mu)$ un espace probabilisé, $T : X \rightarrow X$ une application mesurable qui préserve la mesure μ et ξ une partition dénombrable de X . Soit k un entier positif. Montrez les égalités : $h_\mu(T^k, \bigvee_0^{k-1} T^{-i}\xi) = kh_\mu(T, \xi)$ et $h_\mu(T^k) = kh_\mu(T)$.

exercice 5 :

Soient ξ et η deux partitions finies de X . On pose $d(\xi, \eta) = H(\xi \mid \eta) + H(\eta \mid \xi)$. Soient ξ_1, ξ_2, ξ_3 trois partitions finies de X . Montrez l'inégalité triangulaire :

$$d(\xi_1, \xi_3) \leq d(\xi_1, \xi_2) + d(\xi_2, \xi_3)$$

exercice 6 :

Soit $(X, \mathcal{T}, \mu)$ et $(Y, \mathcal{S}, \nu)$ deux espaces probabilisés et $T : X \rightarrow X, S : Y \rightarrow Y$ deux applications mesurables préservant la mesure. On suppose que ces deux systèmes dynamiques mesurés sont isomorphes. Montrez qu'ils ont même entropie.

exercice 7 :

Soit $(X, \mathcal{T}, \mu)$ un espace probabilisé non atomique ($\mu(\{x\}) = 0, \forall x \in X$). Montrez que l'application identité n'admet pas de partition génératrice. Généralisez au cas d'une transformation pour laquelle tous les points sont périodiques.

exercice 8 :

Considérons $T : X \rightarrow X$ une application mesurable qui préserve la mesure μ . Soit ξ une partition finie de X telle que les éléments de $T^{-i}\xi, i \in \mathbf{N}$, engendrent $\mathcal{T}$. On suppose T inversible. Montrez que $h_\mu(T) = 0$.

exercice 9 :

Dans la preuve du théorème de Kolmogorov-Sinai, on a vu que si une suite de partitions croissantes ξ_n engendrent $\mathcal{T}$, alors $H(\eta \mid \xi_n)$ converge vers 0 pour toute partition finie η . Montrez la réciproque.

Indic : montrez que $H(\{A, A^c\} \mid \xi_n) = 0$ pour tout $A \in \mathcal{T}$.

Commentaires

La suite $a_n = H(\bigvee_{k=0}^n T^{-k}\xi)$ vérifie l'inégalité $a_{n+m} \leq a_m + a_n$. On dit qu'elle est sous-additive. La convergence de la suite $\frac{a_n}{n}$ aurait donc pu être déduite du lemme élémentaire suivant : soit $a_n \in \mathbf{R}$ une suite satisfaisant $a_{n+m} \leq a_m + a_n$. Alors la suite $\frac{a_n}{n}$ admet une limite dans $[-\infty, \infty[$ et cette limite coïncide avec $\inf \frac{a_n}{n}$.

On aurait pu définir l'entropie en prenant la borne supérieure sur toutes les partitions dénombrables d'entropie finie. On aurait obtenu la même valeur pour $h_\mu(T)$. Le théorème de Kolmogorov-Sinai est encore vrai pour les partitions génératrices dénombrables, mais il faut justifier l'emploi du théorème de convergence dominée. Pour cela, on peut employer le lemme suivant, dû à K. L. Chung (1961) : Soit ξ une partition dénombrable et $\{\xi_n\}_{n \in \mathbf{N}}$ une suite croissante de partitions dénombrables. Alors $\int \sup_n I(\xi|\xi_n) d\mu \leq H(\xi) + 1$.

Supposons que l'espace sur lequel est défini la transformation est un espace de Lebesgue. Une partition finie ou dénombrable ξ est génératrice unilatère si et seulement si on peut trouver un ensemble de mesure totale en restriction duquel l'application suivante est injective :

$$\begin{aligned} X &\rightarrow \xi^{\mathbf{N}} \\ x &\mapsto \{\xi(T^n(x))\}_{n \in \mathbf{N}} \end{aligned}$$

Une partition génératrice donne lieu à un codage qui permet de distinguer presque tous les points entre eux. Ce résultat peut se démontrer en employant une version mesurable du théorème de Stone-Weierstrass : une sous-algèbre dénombrable de fonctions bornées de L^2 engendre tout L^2 si et seulement si, après restriction à un ensemble de mesure totale, cette algèbre sépare les points. Il suffit d'appliquer ce théorème à l'algèbre engendrée par les fonctions indicatrices des éléments des $T^{-k}\xi$, $k \in \mathbf{N}$. On peut énoncer un théorème analogue pour les partitions génératrices bilatères des transformations inversibles, en codant à l'aide de $\xi^{\mathbf{Z}}$.

Dans quelle mesure un système d'entropie positive peut-il être qualifié d'aléatoire ? Les décalages de Bernoulli, qui modélisent le lancer d'un dé ou d'une pièce de monnaie, tombent certainement dans cette catégorie. Y. G. Sinai montre en 1964 que tout système dynamique mesuré ergodique d'entropie positive admet un quotient isomorphe à un décalage de Bernoulli de même entropie. Ce résultat permet d'interpréter l'entropie en termes probabilistes : un système est d'entropie supérieure ou égale à $-\sum p_i \log p_i$ si et seulement si on peut trouver une partition de l'espace $\{A_1, \dots, A_n\}$ satisfaisant $\mu(A_i) = p_i$, et telle que les $X_n(x) = \sum i \mathbf{1}_{A_i}(T^n(x))$ forment une suite de variables aléatoires indépendantes identiquement distribuées. Par exemple, un système est d'entropie plus grande que $\log(6)$ si et seulement s'il permet de simuler le lancer d'un dé à six faces.

Les premiers systèmes réguliers à avoir été classifiés par le biais de l'entropie sont les automorphismes ergodiques du tore T^2 (R. L. Adler et B. Weiss, 1967). Deux tels systèmes sont isomorphes si et seulement si ils ont même entropie. En particulier, il existe des automorphismes de T^2 qui sont isomorphes sans être conjugués algébriquement, par exemple $\begin{pmatrix} 5 & 2 \\ 2 & 1 \end{pmatrix}$ et $\begin{pmatrix} 5 & 4 \\ 1 & 1 \end{pmatrix}$. L'isomorphisme est construit à l'aide d'une partition de Markov.

D. Ornstein montre en 1970 que deux systèmes de Bernoulli bilatères sont isomorphes si et seulement s'ils ont même entropie. Ce résultat s'étend aux décalages de type fini mélangeants. Y. G. Sinai, R. Bowen, M. Ratner montrent ensuite, en utilisant des partitions markoviennes, que la plupart des systèmes uniformément hyperboliques sont isomorphes à des systèmes de Bernoulli : difféomorphismes Anosov C^2 préservant la mesure de Lebesgue, difféomorphismes Axiom A C^1 préservant une mesure de Gibbs mélangeante, etc. La théorie de Pesin permet d'étendre ces résultats à certaines classes de systèmes non uniformément hyperboliques. La dynamique de ces systèmes est donc, sur le plan mesurable, fortement stochastique.

Classer les transformations non inversibles à isomorphisme près est plus délicat. Les seuls isomorphismes entre deux décalages de Bernoulli unilatères sont obtenus par permutation des symboles. B. Marcus et S. Tuncel donnent en 2001 une classification des décalages de type fini unilatères. Une transformation rationnelle de la sphère de Riemann possède une unique mesure de probabilité d'entropie maximale ; D. Heicklen et C. Hoffman démontre en 2002 que ce système est isomorphe à un décalage de Bernoulli.

Un système dynamique mesuré peut être d'entropie infinie ; considérons une suite p_i de réels dans $[0, 1]$ satisfaisant $\sum p_i = 1$ et $-\sum p_i \log p_i = \infty$. La suite des p_i définit une mesure de probabilité μ sur l'ensemble des entiers $\mathbf{N}$, et le décalage sur l'espace produit $\mathbf{N}^{\mathbf{Z}}$, relativement à la mesure produit $\mu^{\otimes \mathbf{Z}}$, est d'entropie infinie. On peut cependant montrer qu'un homéomorphisme Lipschitzien de constante de Lipschitz K , défini sur un espace métrique compact de dimension finie et préservant une mesure de probabilité borélienne, est d'entropie finie, majorée par $\dim(X) \log^+ K$ (Kouchnirenko, 1965). En particulier, un difféomorphisme C^1 défini sur une variété compacte, et préservant une mesure de probabilité borélienne, est d'entropie finie.

W. Krieger démontre en 1970 que toute transformation inversible ergodique définie sur un espace de Lebesgue, qui préserve la mesure et dont l'entropie est finie, admet une partition génératrice bilatère avec au plus $e^{h_\mu(T)} + 1$ éléments. Les transformations inversibles ergodiques d'entropie strictement positive n'ont pas de partitions génératrices unilatères.

D. A. Lind et J. P. Thouvenot montrent en 1977 que toute transformation ergodique d'entropie finie d'un espace de Lebesgue est isomorphe à un homéomorphisme du tore préservant la mesure de Lebesgue. L'existence de réalisations C^1 pour les transformations d'entropie finie reste ouverte.

Entropie et théorie de l'information

Von Neumann told me, 'You should call it entropy, for two reasons. In the first place your uncertainty function has been used in statistical mechanics under that name, so it already has a name. In the second place, and more important, no one really knows what entropy really is, so in a debate you will always have the advantage.'

C. E. Shannon (1916-2001)

Le terme d'entropie est utilisé pour la première fois par R. Clausius en 1865, dans le cadre de ses recherches sur la chaleur. Le concept sous-jacent va jouer un rôle crucial dans le développement de la thermodynamique et de la mécanique statistique, avec les travaux de J. W. Gibbs et L. Boltzmann, à la fin du XIX^e siècle. Ce ne sont cependant pas ces deux théories qui inspirent A. N. Kolmogorov lorsqu'il introduit un nouvel invariant appelé "entropie" pour l'étude des systèmes dynamiques, mais bien les travaux de C. E. Shannon en théorie de l'information.

Dans un article fameux publié en 1948, et marquant la naissance de la théorie de l'information, C. E. Shannon introduit une quantité destinée à quantifier l'information perdue dans les transmissions téléphoniques, lorsque le signal est bruité.

L'expérience suivante montre comment appréhender l'entropie, du point de vue de cette théorie de l'information. Considérons une source dont le résultat appartient à un ensemble de n symboles $x_1, \dots, x_n$ de probabilités respectives $p_1, \dots, p_n$. Nous cherchons à déterminer quel est le résultat produit par la source en posant des questions de type oui/non : "le résultat est-il égal à x_1 ?", "appartient-il à tel sous-ensemble" etc. Posons $H = -\sum p_i \log p_i$, en utilisant le logarithme en base deux. C. E. Shannon montre que le nombre moyen de questions qu'il faut poser est compris entre H et $H + 1$, dès l'instant où on a fait un choix optimal pour ces questions.

On a introduit au chapitre précédent le concept de partition et d'information associée. Quel lien y a-t-il avec le problème qui vient d'être exposé ? Remarquons tout d'abord qu'une question partitionne l'espace des résultats en deux sous-ensembles. Une suite de questions donne donc une suite de partitions de notre espace. Si cette suite de questions est en mesure de distinguer entre tous les résultats possibles, cela signifie que la partition engendrée est la partition en points, $\xi(x) = \{x\}$ pour tout x . L'entropie de cette partition est précisément égale à H .

La valeur de H apparaît comme la quantité moyenne d'information nécessaire pour déterminer le résultat produit par la source. On va expliquer dans la suite comment quantifier ce concept d'information et traiter quelques exemples concrets. On calculera explicitement l'entropie H lorsque la source produit une suite de symboles indépendants entre eux, ou lorsque la probabilité d'un symbole ne dépend que du symbole qui le précède (cas markovien).

Entropie et théorie de l'information

La notion d'information

Axelle prend connaissance du résultat obtenu à l'issu d'une certaine expérience aléatoire. François cherche à déterminer ce résultat et demande à Axelle si elle peut lui fournir des informations. Axelle autorise François à lui poser une question, à laquelle elle répondra par oui ou par non.

François pose sa question et Axelle répond par l'affirmative. Quelle est la quantité d'information que François a obtenue ?

Cherchons à comprendre ce que peut valoir cette quantité d'information. Remarquons tout d'abord que la question de François partitionne l'ensemble des résultats Ω en deux sous-ensembles : d'un côté les résultats qui vont entraîner une réponse positive de la part d'Axelle, et de l'autre ceux qui donneront une réponse négative.

La quantité d'information reçue par François dépend de la probabilité p d'obtenir une réponse positive à sa question ; notons cette quantité $I(p)$. Supposons maintenant qu'Axelle répète deux fois l'expérience et que la réponse à la question initiale est positive dans les deux cas, et faisons le postulat suivant :

La valeur de l'information fournie par deux résultats obtenus de manière indépendante est égale à la somme des quantités d'information associées à chacun des résultats.

La quantité d'information obtenue par François est donc égale au double de l'information qui aurait résulté d'une réponse positive à une seule réalisation de l'expérience. D'un autre côté, la probabilité d'obtenir deux fois une réponse positive est égale à p^2 . On a donc $I(p^2) = 2I(p)$. On en déduit que $I(p^m) = mI(p)$, en répétant m fois l'expérience. Un petit calcul montre alors que $I(p^m) = mI(p)$ pour tout nombre rationnel positif m . Si la fonction $p \mapsto I(p)$ est continue sur $]0, 1[$, cela entraîne l'égalité $I(p^x) = xI(p)$ pour tout réel x . Par convention, on pose $I(\frac{1}{2}) = 1$, ce qui donne $I(y) = -\log_2(y)$.

Calculons l'information moyenne que donne la réponse d'Axelle : avec probabilité p la réponse est positive et la quantité d'information reçue est égale à $\log_2 p$; avec probabilité $1 - p$ la réponse est négative et l'information reçue vaut $\log_2(1 - p)$.

$$H(p) = -p \log_2 p - (1 - p) \log_2(1 - p)$$

Notons $\xi = \{\text{oui}, \text{non}\}$ la partition de Ω associée à la question d'Axelle, x le résultat de l'épreuve aléatoire et μ la mesure de probabilité définie sur Ω . Par définition de p , $p = P(\text{oui}) = \mu(\xi(x))$, et on retrouve la formule usuelle pour l'entropie d'une partition à deux éléments. Traitons maintenant un exemple concret.

Le jeu des questions et des réponses

Axelle lance deux dés à six faces et fait la somme des chiffres obtenus. François essaie de deviner le résultat en posant des questions auxquelles Axelle répond par oui ou par non.

François demande par exemple si le résultat est supérieur ou égal à 7, puis s'il est pair. Il reçoit les réponses "non" puis "oui" à ces questions. Il demande ensuite si le résultat obtenu est égal à 6 et, après avoir reçu une réponse négative, s'il est égal à 2. Ceci est récapitulé dans le *tableau 2*. On peut calculer explicitement l'information que lui donnent les réponses d'Axelle. Notons x le résultat obtenu par Axelle et $\xi_1, \dots, \xi_n$ les partitions associées aux questions successives de François. L'information qu'apportent à François les réponses aux questions 1 à n est égale à $I(\xi_1 \vee \xi_2 \vee \dots \vee \xi_n)(x)$; elle figure dans l'avant-dernière colonne du tableau.

Le gain d'information apporté par la réponse n vaut $I(\xi_n | \xi_1 \vee \dots \vee \xi_{n-1})(x)$, c'est-à-dire $-\log_2 \mu(\xi_n(x) | \xi_1 \vee \dots \vee \xi_{n-1}(x))$. C'est la différence entre $I(\xi_1 \vee \xi_2 \vee \dots \vee \xi_n)(x)$ et $I(\xi_1 \vee \xi_2 \vee \dots \vee \xi_{n-1})(x)$, elle figure dans la dernière colonne du tableau.

Pour deviner le résultat x , François doit totaliser un gain d'information égal à $-\log_2(P(\{x\}))$. On peut suivre sa progression dans le tableau. Par exemple, la réponse à la question 3 : “le résultat est-il égal à 6” lui est plutôt favorable (information supérieure à 1) même si en moyenne, une telle question apporte peu d'information dans ce contexte (information relative moyenne égale à 0.25). Après avoir posé la question 4 : “le résultat est-il égal à 2”, François dispose de suffisamment d'informations pour deviner le nombre recherché. En effet, 4 est le seul résultat qui induit la série de réponses non-oui-non-non à ses questions. Et de fait, François a bien atteint la quantité d'information nécessaire : $\log_2(12) \simeq 3.58$.

Information et chaînes de Markov

Voici un autre exemple issu de la théorie des probabilités. Soit $X_0, X_1, \dots, X_n \dots$ une suite de variables aléatoires stationnaires définies sur un espace probabilisé $(\Omega, \mathcal{T}, P)$, à valeurs dans l'ensemble fini $\mathcal{A} = \{1, \dots, N\}$. Supposons connues les valeurs des $X_i, i \geq 1$. De quelle quantité d'information a-t-on besoin en moyenne, pour connaître la valeur de X_0 ?

On peut supposer que l'espace Ω sur lequel sont définies les X_i est égal à $\mathcal{A}^{\mathbb{N}}$, la variable aléatoire X_i correspondant à la projection sur la coordonnée i . On note $\sigma : \Omega \rightarrow \Omega$ le décalage vers la gauche. Considérons les partitions ξ_n de Ω définies par $\xi_n = \{(X_n = a) | a \in \mathcal{A}\}$ et remarquons que $\xi_n = \sigma^{-n}\xi_0$. Soit $H(X_0 | X_1, \dots, X_n)$ la quantité d'information moyenne nécessaire pour connaître la valeur de X_0 connaissant celle des $X_i, i = 1$ à n . On a l'égalité :

$$H(X_0 | X_1, \dots, X_n) = H(\xi_0 | \xi_1 \vee \xi_2 \vee \dots \vee \xi_n) = H(\xi_0 | \bigvee_{i=1}^n \sigma^{-i}\xi_0)$$

Cette quantité converge vers l'entropie du décalage $h(\sigma, \xi_0)$. Lorsque $\Omega = \mathcal{A}^{\mathbb{N}}$, la partition ξ_0 est génératrice car les éléments de $\bigvee_{i=0}^n \sigma^{-i}\xi_0$ sont les cylindres de longueur $n+1$. L'entropie du décalage s'interprète donc comme l'information moyenne nécessaire pour connaître la valeur “présente” X_0 connaissant les valeurs “passées” $X_i, i \geq 1$, du processus. Faisons le calcul quand les X_i forment une chaîne de Markov.

Théorème

Soit $\mathcal{A}$ un ensemble fini ; le décalage $\sigma : \mathcal{A}^{\mathbb{N}} \rightarrow \mathcal{A}^{\mathbb{N}}$ est défini par la formule $\sigma(\{x_i\}_{i \in \mathbb{N}}) = \{x_{i+1}\}_{i \in \mathbb{N}}$. Pour $i, j \in \mathcal{A}$, on se donne des éléments $p_i, p_{i,j}$ de $[0, 1]$ vérifiant : $\sum_i p_i = 1, \sum_j p_{i,j} = 1$ et $\sum_i p_i p_{i,j} = p_j$. Notons P la probabilité satisfaisant :

$$P(\{x_i\}_{i \in \mathbb{N}} | x_0 = i_0, \dots, x_n = i_n) = p_{i_0} p_{i_0, i_1} p_{i_1, i_2} \dots p_{i_{n-1}, i_n}$$

L'entropie de σ par rapport à P est donnée par :

$$h_P(\sigma) = - \sum_{i,j} p_i p_{i,j} \log p_{i,j}.$$

Preuve

En vertu des égalités $P(X_k = j | X_{k+1} = i) = p_{i,j}, p_i = P(X_1 = i)$ et de la propriété de Markov, on a :

$$\begin{aligned} H(X_0 | X_1, \dots, X_n) &= H(X_0 | X_1) \\ &= - \sum_{i,j} P(X_0 = j, X_1 = i) \log P(X_0 = j | X_1 = i) \\ &= - \sum_{i,j} p_i p_{i,j} \log p_{i,j}. \end{aligned}$$

Comme corollaire, on en déduit qu'un système de Bernoulli, sur un alphabet à n symboles de probabilités $p_1, \dots, p_n$, a pour entropie $-\sum p_i \log p_i$.

Tableau 1 : Loi suivie par la somme de deux lancers d'un dé à six faces

résultat x	2	3	4	5	6	7	8	9	10	11	12
probabilité p	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$
information	5.17	4.17	3.58	3.17	2.85	2.58	2.85	3.17	3.58	4.17	5.17

Tableau 2 : liste de questions/réponses ($x = 4$)

i	Question	Un élément de ξ_i	Entropie cumulée	Gain	Réponse	Information cumulée	Gain
1	≥ 7 ?	{7, 8, 9, 10, 11, 12}	0.98	0.98	non	1.26	1.26
2	pair ?	{2, 4, 6, 8, 10, 12}	1.96	0.98	oui	2	0.74
3	$= 6$?	{6}	2.21	0.25	non	3.17	1.17
4	$= 2$?	{2}	2.3	0.09	non	3.58	0.41

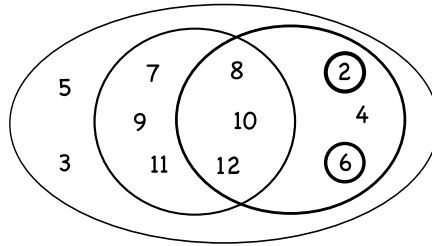
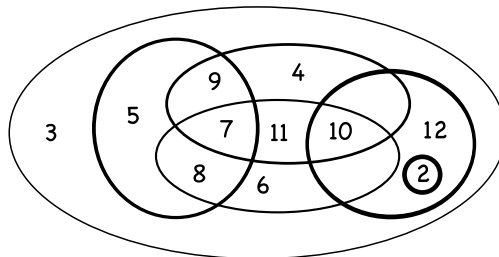


Tableau 3 : liste de questions/réponses ($x = 12$)

i	Question	Un élément de ξ_i	Entropie cumulée	Gain	Réponse	Information cumulée	Gain
1	6,7,8,10 ou 11	{6, 7, 8, 10, 11}	0.98	0.98	non	1.26	1.26
2	4,7,9,10 ou 11	{4, 7, 9, 10, 11}	1.98	1	non	2.16	0.9
3	5,7,8 ou 9	{5, 7, 8, 9}	2.97	0.99	non	3.18	1.02
4	2,10 ou 12	{2, 10, 12}	3.22	0.25	oui	4.17	0.99
5	$= 2$?	{2}	3.28	0.06	non	5.17	1



Exercices

exercice 1 :

Axelle lance deux dés à six faces et fait la somme des deux chiffres obtenus. Elle accepte de répondre par oui ou par non aux questions de François portant sur la valeur de la somme.

François peut-il deviner le résultat à coup sûr, en trois questions seulement ? Quel est le nombre minimal de questions qu'il doit poser, de façon à pouvoir conclure quel que soit le résultat du lancer ? Même problème pour le lancer de trois et quatre dés.

exercice 2 :

On considère une expérience aléatoire admettant n résultats possibles, de probabilités respectives $p_1, \dots, p_n$. François cherche à deviner le résultat de l'expérience en posant des questions de type oui/non. Montrez que le nombre minimal de questions qu'il doit poser, de façon à pouvoir conclure quel que soit le résultat de l'épreuve, est toujours supérieur à l'entropie $-\sum p_i \log_2 p_i$. Montrez qu'on ne peut avoir égalité que si tous les résultats de l'épreuve sont équiprobables. Le nombre de résultats possible doit alors être une puissance de 2.

exercice 3 :

Axelle lance deux dés à six faces et François cherche à deviner la somme des deux chiffres obtenus par Axelle. Pour cela, il peut poser quatre questions.

Les trois premières réponses d'Axelle lui ont apporté une quantité d'information égale à 3.17, et il ne lui reste plus qu'une question à poser.

- Le résultat peut-il être égal à 6 ?
- Le résultat peut-il être égal à 4 ?
- Si c'est le cas, quelle question suggèreriez vous à François ?

On rappelle que $\log_2(3) = 1.58$ et $\log_2(5) = 2.32$

exercice 4 :

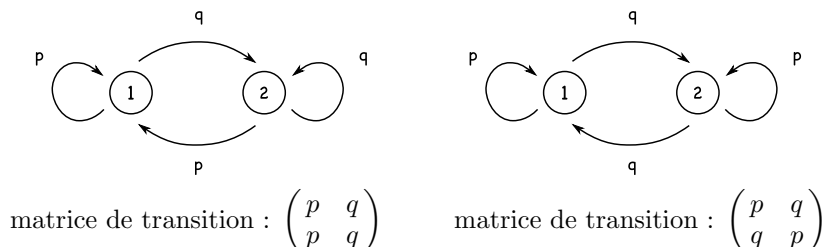
Soit $X_0, X_1, \dots, X_n, \dots$ une suite de variables aléatoires stationnaire markovienne. Montrez l'inégalité : $H(X_n | X_0) \leq H(X_{n+1} | X_0)$.

exercice 5 :

On lance deux dés à n faces et on fait la somme des deux résultats obtenus. Calculez l'entropie du système obtenu en répétant de manière indépendante une telle épreuve. Comparez avec l'entropie associée à la répétition, de manière indépendante, d'une épreuve obéissant à la loi uniforme sur un ensemble à $2n - 1$ éléments.

exercice 6 :

Soit $p, q \in [0, 1]$ tels que $p + q = 1$. Calculez l'entropie des chaînes de Markov suivantes :



exercice 7 :

Montrez que le système de Bernoulli sur un alphabet à trois symboles de probabilités respectives $1/3, 1/3$ et $1/3$, n'est pas isomorphe au système de Bernoulli sur un alphabet à deux symboles de probabilités respectives $1/2, 1/2$.

Commentaires

Revenons au cas du lancer des dés à six faces. Considérons un ensemble de questions $\xi_1 \dots \xi_n$ qui permet de conclure quelque soit le résultat : $\xi_1 \vee \dots \vee \xi_n(x) = \{x\}$, c'est-à-dire $H(\xi_1 \vee \dots \vee \xi_n) = -\sum p_i \log p_i = 3.27$. Pour certains résultats, il n'est pas nécessaire de poser les n questions pour être en mesure de conclure. Par exemple, pour les questions du tableau 3, il suffit de poser les trois premières questions pour trouver le résultat si celui-ci vaut 4,5,6,7,8 ou 9.

En moyenne, quel est le nombre de questions effectivement posées pour trouver le résultat ? Dans son papier fondateur de 1948, C. E. Shannon montre que ce nombre moyen est toujours supérieur à l'entropie. D. Huffman propose en 1952 un algorithme pour construire une suite de questions minimisant le nombre moyen de questions à poser. Pour le cas du lancer de deux dés, le tableau 3 a été obtenu à l'aide de cet algorithme. Le nombre moyen de questions effectivement posées vaut 3.306 ; il est optimal. Les méthodes de compression jpeg, mp3 et pkzip utilisent cet algorithme de Huffman.

Remarquons qu'il faut poser au moins quatre questions pour être en mesure de distinguer tous les résultats entre eux. En effet, trois questions partitionnent l'ensemble des résultats en au plus $2^3 = 8$ parties, alors qu'il y a 11 résultats différents.

On peut chercher à déterminer un ensemble de quatre questions, qui minimise le nombre moyen de questions à poser, parmi les ensembles de quatre questions. Ceci s'obtient par un algorithme "numismatique", et donne : 6, 7, 8 ou 9 ? 4, 5, 7, 8 ou 10 ? 3, 5, 6, 7 ou 11 ? 2, 3 ou 4 ? pour un nombre moyen de questions égal à 3.333.

L'entropie associée au lancer de deux dés à n faces est donnée par :

$$H = -\sum p_i \log p_i = 2 \log n - \frac{1}{n^2} (\sum_{i=1}^{n-1} 2i \log i + n \log n) \sim \log(n) + \frac{1}{2} + o(1/n)$$
On peut vouloir comparer cela avec l'entropie de la loi uniforme sur un ensemble à $2n - 1$ éléments :

$$H = \log(2n - 1) \sim \log(n) + \log(2) + o(1).$$

Calculs d'entropie

L'entropie est une loi générale de l'univers : la tendance naturelle des choses à passer de l'ordre au désordre sous l'effet d'un hasard calculable.

F. Jacob

Calculer l'entropie d'une transformation préservant une mesure est en général un problème délicat. On va s'intéresser à une classe d'applications non inversibles pour lesquelles on peut mener le calcul sans trop de difficulté.

Une application T définie sur un espace métrique X est dite *uniformément dilatante* si on peut trouver une partition $\{X_i\}$ de X telle que, en restriction à chacun des X_i , T dilate la métrique au sens suivant : on peut trouver une constante $K > 1$ telle que, pour tout i ,

$$\forall x, y \in X_i, \quad d(T(x), T(y)) \geq K d(x, y).$$

Nous avons déjà rencontré plusieurs exemples d'applications dilatantes : les automorphismes des tores dont toutes les valeurs propres sont de module strictement supérieur à un, les applications C^1 par morceau de l'intervalle dont la dérivée est plus grande qu'une certaine constante $K > 1$ (*cf fig.3*), satisfont la propriété ci-dessus.

Pour ces applications, il est possible de donner une expression explicite pour l'entropie des mesures de probabilité invariantes. Cette expression fait intervenir le taux de dilatation de la mesure sous l'action de la transformation. D'un point de vue informel, on peut dire que le désordre inhérent au système est proportionnel à ce facteur de dilatation.

Ce facteur se calcule aisément quand la transformation est régulière et préserve une mesure invariante absolument continue par rapport à la mesure de Lebesgue. Il s'obtient par un simple changement de variables. L'entropie est alors égale à l'intégrale du logarithme du jacobien de la transformation. On relie donc une quantité mesurable, définie de manière globale, et mesurant l'incertitude en ce qui concerne l'évolution du système au cours du temps, à une quantité obtenue en moyennant la dilatation infinitésimale observée au voisinage de chacun des points de l'espace.

Les systèmes de Bernoulli définis sur un alphabet fini I peuvent s'interpréter en terme d'application dilatante. Il suffit pour cela de munir l'espace compact $I^{\mathbf{N}}$ de la distance suivante :

$$d(\{x_i\}, \{y_i\}) = 2^{-\min\{j \in \mathbf{N} \mid x_j \neq y_j\}}$$

Le décalage σ défini sur cet espace est alors dilatant en restriction aux cylindres $[a] = \{ \{x_i\} \in I^{\mathbf{N}} \mid x_0 = a \}$, pour tout $a \in I$:

$$\forall x, y \in [a], \quad d(\sigma(x), \sigma(y)) \geq 2 d(x, y).$$

On peut raisonner de même pour une chaîne de Markov à espace d'état fini et obtenir une expression pour l'entropie du décalage en fonction des probabilités de transition. Cette méthode de calcul n'est pas la plus élémentaire, mais elle permet d'illustrer le concept d'application dilatante.

Calculs d'entropie

Soit $(X, \mathcal{T}, \mu)$ un espace probabilisé, $T : X \rightarrow X$ une application mesurable qui préserve la mesure μ . Supposons que T soit injective en restriction à un ensemble mesurable $A \subset X$. On définit l'inverse du jacobien de T sur A par la formule :

$$\frac{1}{|T'_\mu|} = \frac{d}{d\mu} (T_*(\mu|_A)) \circ T$$

de façon à avoir le changement de variable habituel : $\int_A g \circ T \, d\mu = \int_{TA} \frac{g}{|T'_\mu| \circ T|_A} \, d\mu$.

Formule de Rokhlin

Soit $(X, \mathcal{T}, \mu)$ un espace probabilisé, $T : X \rightarrow X$ une application mesurable qui préserve la mesure μ . Supposons qu'il existe une partition finie X_i , i allant de 1 à k , telle que les TX_i soient mesurables et $T : X_i \rightarrow TX_i$ soit bijective d'inverse mesurable. Alors :

$$pp \, x_0 \in X, \quad E(f | T^{-1}\mathcal{T})(x_0) = \sum_{Tx=Tx_0} \frac{f(x)}{|T'_\mu(x)|}$$

$$H(\{X_i\}_{i=1..k} | T^{-1}\mathcal{T}) = \int_X \log |T'_\mu| \, d\mu$$

Preuve

Notons $T_i^{-1} : TX_i \rightarrow X_i$ l'inverse de T sur les domaines considérés ; on prolonge cette fonction à X de manière arbitraire. Remarquons que pour tous les $x_0 \in X_i$,

$$\{x | Tx = Tx_0\} = \{T_i^{-1}(Tx_0) | i \text{ tel que } T(x_0) \in T(X_i)\}.$$

Pour calculer l'espérance conditionnelle, on écrit : $f = \sum_i \mathbf{1}_{X_i} f \circ T_i^{-1} \circ T$, ce qui donne :

$$E(f | T^{-1}\mathcal{T}) = \sum_i E(\mathbf{1}_{X_i} | T^{-1}\mathcal{T}) f \circ T_i^{-1} \circ T$$

Il suffit donc de démontrer l'égalité : $E(\mathbf{1}_{X_i} | T^{-1}\mathcal{T}) = \frac{\mathbf{1}_{TX_i} \circ T}{|T'_\mu| \circ T_i^{-1} \circ T}$.

$$\int g \circ T \, E(\mathbf{1}_{X_i} | T^{-1}\mathcal{T}) \, d\mu = \int_{X_i} g \circ T \, d\mu = \int \frac{\mathbf{1}_{TX_i} g}{|T'_\mu| \circ T_i^{-1}} \, d\mu = \int \frac{\mathbf{1}_{TX_i} \circ T \, g \circ T}{|T'_\mu| \circ T_i^{-1} \circ T} \, d\mu.$$

Il reste à calculer l'entropie relative de la partition $\{X_i\}$:

$$\begin{aligned} H(\{X_i\} | T^{-1}\mathcal{T}) &= \sum_i \int -\mathbf{1}_{X_i} \log E(\mathbf{1}_{X_i} | T^{-1}\mathcal{T}) \, d\mu \\ &= \sum_i \int_{X_i} \log |T'_\mu| \circ T_i^{-1} \circ T \, d\mu \\ &= \sum_i \int_{X_i} \log |T'_\mu| \, d\mu \\ &= \int_X \log |T'_\mu| \, d\mu \end{aligned}$$

Corollaire

$h_\mu(T) \geq \int_X \log |T'_\mu| \, d\mu$ avec égalité si la partition $\{X_i\}$ est génératrice unilatère.

Remarques

– Supposons $X \subset \mathbf{R}^n$, μ est la mesure de Lebesgue, T est C^1 sur l'intérieur des X_i (Lipschitz suffit) et $\mu(\partial X_i) = 0$. Alors $|T'_\mu| = |\det(DT)|$ presque partout.

– La quantité $E(\mathbf{1}_{X_i} | T^{-1}\mathcal{T})(x)$ s'interprète comme la probabilité que x soit dans X_i , sachant la valeur de $T(x)$.

– Si T possède un générateur unilatère, $h(T)$ s'interprète comme la quantité moyenne d'information nécessaire pour connaître x sachant que Tx est connu.

– L'opérateur $L_\mu f(y) = \sum_{Tx=y} \frac{f(x)}{|T'_\mu(x)|}$ est l'opérateur de transfert associé à T ; c'est l'adjoint de l'isométrie $f \mapsto f \circ T$ défini sur $L^2(X, \mu)$.

Entropie des décalages

Calculons l'entropie du décalage défini sur une chaîne de Markov. Soit I un alphabet fini. On se place sur $X = I^{\mathbb{N}}$ et on considère le décalage $T(\{x_i\}_{i \in \mathbb{N}}) = \{x_{i+1}\}_{i \in \mathbb{N}}$. Pour chaque indice $i, j \in I$, on se donne des nombres réels $p_{ij} \in [0, 1]$, $p_i \in [0, 1]$ qui satisfont $\sum_i p_i = 1$, $\sum_j p_{ij} = 1$ et $\sum_i p_i p_{ij} = p_j$.

Soit $a_1, \dots, a_n$ des éléments de I . Rappelons que le cylindre $[a_1, \dots, a_n] \subset X$ est composé des éléments de X qui commencent par la suite $a_1, \dots, a_n$. D'après le théorème d'extension de Kolmogorov, il existe une mesure de probabilité μ sur X qui satisfait $\mu([a_1, a_2, \dots, a_n]) = p_{a_1} p_{a_1 a_2} \dots p_{a_{n-1} a_n}$. Cette mesure est T -invariante.

Théorème

$$h_\mu(T) = - \sum_{i,j} p_i p_{ij} \log p_{ij}$$

Preuve

Soit N le cardinal de I . Chaque point de X a exactement N pré-images et T est bijective de $[i]$ dans X . On prend donc $X_i = [i]$. La partition $\{[i] \mid i \in I\}$ est génératrice car la partition engendrée par ses k premières images réciproques est composée de tous les cylindres de longueur k . Calculons le jacobien de $T : [i] \rightarrow X$.

$$\mu(T^{-1}[a_1, \dots, a_n] \cap [i]) = \mu([i, a_1, \dots, a_n]) = \frac{p_i p_{ia_1}}{p_{a_1}} \mu([a_1, \dots, a_n])$$

Ceci montre que $|T'_\mu|$ est constant sur le cylindre $[i, j]$ et vaut $\frac{p_j}{p_i p_{ij}}$. Au final,

$$h_\mu(T) = - \sum_{i,j} p_i p_{ij} \log(p_{ij}) - \sum_i p_i \log(p_i) (\sum_j p_{ij}) + \sum_j \log(p_j) (\sum_i p_i p_{ij})$$

Et les deux derniers termes se compensent.

Entropie des applications dilatantes par morceaux

Montrons que la partition $\{X_i\}$ est génératrice dès que T est dilatante en restriction à chacun des morceaux de la partition (cf fig.1).

Proposition

Soit X un espace métrique, μ une mesure de probabilité borélienne et $T : X \rightarrow X$ une application borélienne qui préserve la mesure μ . Soit ξ une partition finie, dont le diamètre des éléments est majoré ; on suppose que pour un certain $K > 1$ et pour tout $A \in \xi$,

$$\forall x, y \in A, \quad d(Tx, Ty) \geq K d(x, y)$$

Alors ξ est une partition génératrice.

Preuve

Montrons que le diamètre des ensembles $(\bigvee_0^n T^{-i} \xi)(x)$ tend vers 0 quand $n \rightarrow \infty$.

Si ce n'est pas le cas, on peut trouver $\delta > 0$ et, pour tout entier $n \in \mathbb{N}$, des points x_n dans $\bigvee_0^n T^{-i} \xi(x)$ tels que $d(x_n, x) > \delta$. Pour tout $i \in \{0..n\}$, Le point $T^i x_n$ appartient à $\xi(T^i x)$, ce qui donne :

$$d(T^{i+1} x_n, T^{i+1} x) \geq K d(T^i x_n, T^i x)$$

par conséquent : $\text{diam } \xi(T^n x) \geq d(T^n x_n, T^n x) \geq K^n d(x_n, x) \geq K^n \delta$.

Le diamètre des $\xi(T^n x)$ ne serait donc pas majoré. On conclut à l'aide du lemme :

Lemme

Soit X un espace métrique et ξ_n une suite de partitions dénombrables qui satisfait pour tout $x \in X$, $\text{diam } \xi_n(x) \rightarrow 0$. Alors les éléments des ξ_n , $n \in \mathbb{N}$, engendrent la tribu des boréliens de X ; cf fig.2.

Preuve

Soit U un ouvert de X . Pour tout $x \in U$, on peut trouver $n \in \mathbb{N}$ tel que $\xi_n(x) \subset U$. On a donc : $U = \bigcup_{n \in \mathbb{N}} \bigcup_{\substack{A \in \xi_n \\ \text{et } A \subset U}} A$. L'ouvert U est dans la tribu engendrée par les ξ_n .

figure 1 : Dilatation et partition génératrice

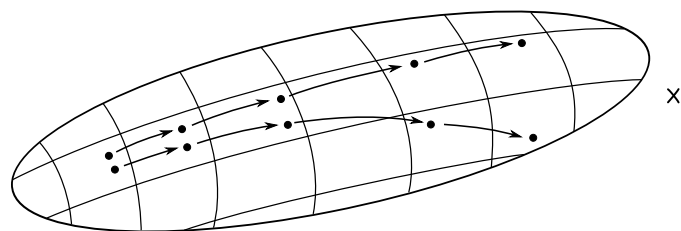


figure 2 : Suite de partitions de diamètre arbitrairement petit

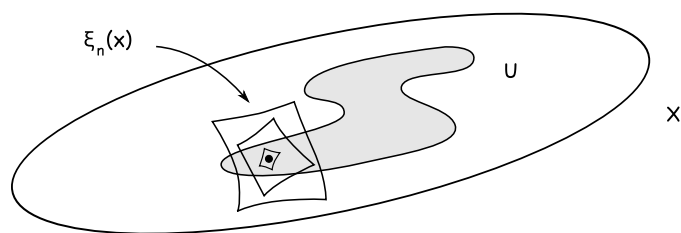
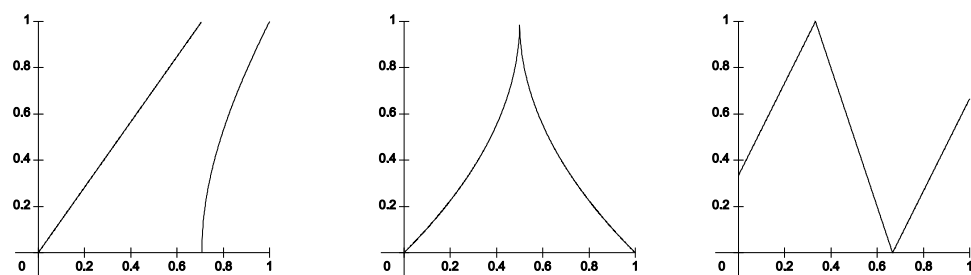


figure 3 : Quelques applications dilatantes de l'intervalle



Exercices

exercice 1 :

On considère l'application $T : [0, 1] \rightarrow [0, 1]$ donnée par :

$$T(x) = \begin{cases} \sqrt{2}x & \text{sur } [0, 1/\sqrt{2}] \\ \sqrt{2x^2 - 1} & \text{sur } [1/\sqrt{2}, 1] \end{cases}$$

Montrez que T préserve la mesure $2x \, dx$ et calculez son entropie.

exercice 2 :

Soit A une matrice $n \times n$ inversible dont les valeurs propres sont toutes de module strictement supérieur à un. Cette matrice induit une application du tore $\mathbf{T}^n$ par passage au quotient. Montrez que cette application préserve la mesure de Lebesgue, puis qu'elle est injective et dilatante sur tout ensemble de diamètre suffisamment petit. Calculez son entropie.

exercice 3 :

Soit U un ouvert de $\mathbf{R}^n$ et $T : U \rightarrow \mathbf{R}^n$ une application C^1 injective préservant une mesure finie de la forme $d\mu = h \, dx$, avec h mesurable. Calculez $|T'_\mu|$.

On suppose que la fonction $\log(h)$ est μ -intégrable. Montrez la formule suivante :

$$\int_U \log |T'_\mu| \, d\mu = \int_U \log |\det D_x T| \, d\mu.$$

exercice 4 :

Soit $T : [0, 1] \rightarrow [0, 1]$ l'application définie par :

$$T(x) = \begin{cases} 2x + \frac{1}{3} & \text{sur } [0, \frac{1}{3}] \\ -3x + 2 & \text{sur } [\frac{1}{3}, \frac{2}{3}] \\ 2x - \frac{4}{3} & \text{sur } [\frac{2}{3}, 1] \end{cases}$$

Montrez que T préserve une mesure de probabilité absolument continue par rapport à la mesure de Lebesgue. Calculez l'entropie de T par rapport à cette mesure. Peut-on trouver une partition génératrice ayant deux éléments ?

exercice 5 :

Soit X un espace métrique compact, $T : X \rightarrow X$ une application continue et μ une mesure de probabilité borélienne invariante par T . Soit ξ une partition finie satisfaisant la propriété suivante :

$$\forall A \in \xi, \forall x, y \in \bar{A} \text{ distincts}, \quad d(Tx, Ty) > d(x, y).$$

Montrez que la partition ξ est génératrice.

exercice 6 :

Montrez que l'application $T : [0, 1] \rightarrow [0, 1]$ suivante préserve la mesure de Lebesgue :

$$\begin{aligned} T(x) &= \sqrt{|2x - 1|} & \text{si } x \in [0, 1/2] \\ &= 1 - \sqrt{|2x - 1|} & \text{si } x \in]1/2, 1] \end{aligned}$$

Calculez son entropie.

exercice 7 :

Soit $(X, \mathcal{T}, \mu)$ un espace probabilisé, $T : X \rightarrow X$ une application mesurable qui préserve μ . On suppose qu'il existe une partition finie $\{X_i\}$ telle que les TX_i soient mesurables et telle que la restriction de T à chacun des X_i soit bijective, bi-mesurable, de X_i sur TX_i . Montrez l'égalité : $\int_X |T'_\mu| \, d\mu = \int_X \text{Card}(T^{-1}(y)) \, d\mu(y)$.

exercice 8 * :

Soit $p, q \in [0, 1]$ tels que $p+q = 1$. Calculez le jacobien du décalage unilatère associé à la chaîne de Markov de matrice de transition $\begin{pmatrix} p & q \\ p & q \end{pmatrix}$. Même question lorsque la matrice de transition vaut $\begin{pmatrix} p & q \\ q & p \end{pmatrix}$. En déduire que les systèmes dynamiques mesurés associés ne sont pas isomorphes.

Commentaires

Pour une application mesurable $T : X \rightarrow X$ injective en restriction aux éléments d'une partition finie X_i , le jacobien $|T'_\mu|$ tel qu'il a été défini plus haut dépend à priori de la partition X_i choisie. Cependant, si $\{Y_j\}$ est une autre partition telle que T soit injective en restriction à chacun de ses éléments, les jacobiens $T'_{\mu|X_i}$ et $T'_{\mu|Y_j}$ coïncident presque partout en restriction à $X_i \cap Y_j$. La fonction $|T'_\mu|$ peut donc être considéré comme bien défini, à un ensemble de mesure nulle près, sans faire référence à une partition spécifique.

L'inverse du jacobien admet une interprétation probabiliste. La quantité $1/|T'_\mu(x)|$ correspond à la probabilité d'obtenir la valeur x parmi toutes les valeurs de $T^{-1}(T(x))$, connaissant la valeur de $T(x)$.

On s'est restreint au cas d'une partition $\{X_i\}$ de cardinal fini car c'est dans ce contexte que le concept d'entropie a été défini. Les résultats démontrés plus haut sont encore vrais si la partition $\{X_i\}$ est dénombrable, les preuves restent inchangées.

Un espace de Lebesgue X sur lequel est défini une application mesurable $T : X \rightarrow X$ qui préserve la mesure, et dont le cardinal des fibres $T^{-1}(x)$ est dénombrable pour tout $x \in X$, possède automatiquement une partition dénombrable $\{X_i\}_{i \in \mathbb{N}}$, telle que T soit injective en restriction à chacun des X_i . Ce résultat est démontré dans le livre de W. Parry, *entropy and generators in ergodic theory* [Pa69].

On peut relaxer les hypothèses de dilatation uniforme, dans la proposition montrant que la partition $\{X_i\}$ est génératrice. D'abord, il suffit qu'une puissance de la transformation soit dilatante. Lorsque l'espace X est compact, il suffit d'avoir une estimée de la forme $d(Tx, Ty) > d(x, y)$ sur les éléments de la partition. Lorsque la transformation est ergodique, on peut se contenter d'une dilatation uniforme sur un seul élément, et d'une estimée de la forme $d(Tx, Ty) \geq d(x, y)$ sur les autres éléments. Presque tout élément a une orbite qui passe une infinité de fois dans l'élément sur lequel on a la dilatation, et c'est suffisant pour conclure.

Soit X une variété compacte et T une application différentiable C^1 définie sur X . La condition $\det D_x T > 1$ en tout point assure le caractère uniformément dilatant de la transformation en restriction à tout ensemble de diamètre suffisamment petit. Toutes les variétés différentielles n'admettent pas une telle transformation. Seules les variétés qui peuvent s'écrire comme quotient d'un groupe de Lie nilpotent par un sous-groupe discret sont susceptibles de porter de telles applications.

Le jacobien est un invariant de conjugaison mesurable. Il peut être utilisé pour distinguer entre les décalages unilatères. Par exemple le décalage unilatère associé à la chaîne de Markov de matrice de transition $\begin{pmatrix} p & q \\ p & q \end{pmatrix}$ et celui associé à la chaîne de matrice de transition $\begin{pmatrix} p & q \\ q & p \end{pmatrix}$ ont même entropie. Mais les partitions données par les lignes de niveau du jacobien ne peuvent pas être isomorphes : dans le premier cas, cette partition engendre les boréliens sous l'action du décalage, tandis que dans le second cas, la partition engendrée est invariante par permutation des deux symboles.

Espaces de Lebesgue et isomorphisme

Nous avons la chance unique d'avoir à notre disposition une langue universelle, la numérotation décimale écrite, utilisons la.

H. Lebesgue (1875-1941)

Intéressons nous aux espaces sur lesquels sont définis les systèmes dynamiques mesurés. Nous dirons que deux espaces probabilisés sont isomorphes si après avoir écarté un ensemble négligeable de points dans ces deux espaces, on peut trouver une bijection mesurable qui préserve la mesure, et dont l'application réciproque est elle aussi mesurable. Si deux systèmes dynamiques sont isomorphes, les espaces sous-jacents sont eux aussi isomorphes.

Peut-on classer les espaces probabilisés à isomorphisme près ? Remarquons tout d'abord qu'un isomorphisme envoie les points de mesure strictement positive (les atomes de la mesure) sur des points qui sont eux aussi de mesure strictement positive. Le nombre et la masse de ces atomes est donc un invariant d'isomorphisme.

De manière surprenante, la plupart des espaces probabilisés sans atomes sont isomorphes à $[0, 1]$, muni de la mesure de Lebesgue. Seul un manque de mesurabilité ou de séparabilité peut mettre en défaut un tel isomorphisme.

Expliquons pourquoi il en est ainsi, en considérant le cas d'un borélien X dans un espace métrique séparable complet. Du fait de la séparabilité de X , il est possible de coder les points de X par des suites de 0 et de 1, de la même façon que l'on peut représenter un nombre réel par la suite de ses chiffres en base deux. On peut donc trouver une injection mesurable de X dans $\{0, 1\}^{\mathbf{N}}$. Le seul problème est de montrer que cette application envoie les ensembles mesurables sur des ensembles mesurables.

Cette question apparaît très tôt dans l'histoire de la théorie de la mesure. Déjà H. Lebesgue se demandait à quelle condition une application borélienne transforme un borélien en borélien. L'injectivité de la transformation s'avère être une condition suffisante. Pour démontrer cela, il a fallu introduire une nouvelle classe d'ensembles, les ensembles sousliniens (ou analytiques), et mener une étude poussée de la structure de ces ensembles.

Il est possible de contourner les difficultés que posent ces notions, en se contentant de montrer que l'image d'un borélien par une application borélienne injective est juste mesurable. C'est un résultat relativement aisé à établir, car la mesurabilité fait référence à une mesure. On peut donc l'obtenir par un calcul direct.

Rappelons la différence entre ensemble borélien et ensemble mesurable. Soit X un espace topologique, $\mathcal{B}(X)$ la tribu des boréliens de X et μ une mesure sur cette tribu. Nous dirons qu'un sous-ensemble de X est μ -mesurable s'il appartient à la complétion $\overline{\mathcal{B}(X)}$ de la tribu des boréliens relativement à μ .

$$\overline{\mathcal{B}(X)} = \{A \subset X \mid \exists B', B'' \in \mathcal{B}(X) \text{ tels que } B' \subset A \subset B'' \text{ et } \mu(B'' \setminus B') = 0\}$$

La tribu $\overline{\mathcal{B}(X)}$ est *complète* au sens suivant : les sous-ensembles des ensembles négligeables appartiennent tous à la tribu.

Rappelons également qu'une fonction $f : X \rightarrow \mathbf{R}$ définie sur un espace mesuré $(X, \mathcal{T}, \mu)$ est dite *mesurable* si l'image réciproque de tout borélien de $\mathbf{R}$ est dans $\mathcal{T}$. Si la tribu $\mathcal{T}$ est *complète*, cela implique que f est une application mesurable de $(X, \mathcal{T}, \mu)$ vers $(\mathbf{R}, \overline{\mathcal{B}(\mathbf{R})}, f_*\mu)$ i.e. l'image réciproque des ensembles $f_*\mu$ -mesurables sont dans $\mathcal{T}$. Pour cette raison, il est préférable de travailler avec des tribus complètes, dès l'instant où l'on considère des espaces mesurés, afin de ne pas avoir à distinguer entre fonctions et applications mesurables.

Espaces de Lebesgue et isomorphisme

Définition d'un isomorphisme

Deux espaces mesurés $(X, \mathcal{T}, \mu)$ et $(Y, \mathcal{S}, \nu)$ sont dits isomorphes si on peut trouver deux ensembles mesurables $X_0 \subset X$, $Y_0 \subset Y$ tels que $\mu(X_0^c) = 0$, $\nu(Y_0^c) = 0$, ainsi qu'une application mesurable bijective $\varphi : X_0 \rightarrow Y_0$, dont la réciproque est mesurable, et qui vérifie : $\varphi_*\mu = \nu$.

Exemple

Soit μ une mesure de probabilité borélienne non-atomique, définie sur $[0, 1]$. Montrons que $([0, 1], \mathcal{B}([0, 1]), \mu)$ est isomorphe à $([0, 1], \mathcal{B}([0, 1]), \lambda)$. L'application φ est donnée par la fonction de répartition de μ :

$$\varphi(x) = \mu([0, x]).$$

Comme μ est sans atome, l'application φ est continue croissante surjective (fig.1).

– Montrons que $\varphi_*\mu = \lambda$. Soit $y \in [0, 1]$ et $x = \min\{x' \in [0, 1] \mid \varphi(x') = y\}$. On a $\varphi(x) = y$, $\varphi_*\mu([0, y]) = \mu(\varphi^{-1}([0, \varphi(x)])) = \mu([0, x]) = \varphi(x) = y = \lambda([0, y])$.

– Construisons un borélien $X_0 \subset [0, 1]$ en restriction duquel l'application φ est injective. Pour cela, remarquons que tout ouvert de $]0, 1[$ est union dénombrable disjointe d'intervalles ouverts. On peut donc trouver des $a_i, b_i \in]0, 1[$, $a_i < b_i$, tels que

$$]0, 1[\setminus \text{supp } \mu = \coprod_{i \in \mathbf{N}}]a_i, b_i[.$$

Posons $X_0 =]0, 1[\setminus (\cup [a_i, b_i]) = \text{supp } \mu \setminus \cup \{a_i, b_i\} \cup \{0, 1\}$ et $Y_0 = \varphi(X_0) =]0, 1[\setminus \cup \{\varphi(a_i)\}$. La fonction φ est alors une bijection strictement croissante entre X_0 et Y_0 . Comme $\varphi : X_0 \rightarrow Y_0$ et $\varphi^{-1} : Y_0 \rightarrow X_0$ sont strictement croissantes, elles sont boréliennes ; comme $\varphi_*\mu = \lambda$, elles sont mesurables.

Lemme de mesurabilité (version borélienne)

Soit X un espace topologique séparé, μ une mesure borélienne finie définie sur X , intérieurement régulière. Soit Y un espace métrique séparable, $\varphi : X \rightarrow Y$ une application borélienne. Soit $A \subset X$ un ensemble borélien ou μ -mesurable satisfaisant $\varphi(A) \cap \varphi(A^c) = \emptyset$. Alors $\varphi(A)$ est un ensemble mesurable relativement à $\varphi_*\mu$.

En particulier, si φ est injective, c'est un isomorphisme mesurable :

$$\varphi : (X, \overline{\mathcal{B}(X)}, \mu) \xrightarrow{\sim} (Y, \overline{\mathcal{B}(Y)}, \varphi_*\mu)$$

Preuve

Pour tout $\varepsilon > 0$, on peut trouver $K \subset A$ compact tel que $\mu(A \setminus K) < \varepsilon$. D'après le théorème de Lusin, on peut trouver $K' \subset K$ compact tel que $\mu(K \setminus K') < \varepsilon$ et $\varphi|_{K'}$ est continue. En particulier, $\varphi(K')$ est compact donc borélien, $\varphi_*\mu$ -mesurable.

On construit alors par récurrence des compacts K_i disjoints deux à deux, en restriction desquels φ est continue et dont l'union donne presque tout A : il existe $N \subset A$, $\mu(N) = 0$, tel que :

$$A = \coprod K_i \amalg N.$$

On peut appliquer le même raisonnement à A^c et écrire : $A^c = \coprod K'_i \amalg N'$.

Calculons la mesure de $\varphi(\cup K_i)$, qui est borélien :

$$\varphi_*\mu(\varphi(\cup K_i)) = \mu(\varphi^{-1}(\varphi(\cup K_i))) \geq \mu(\cup K_i) = \mu(A).$$

Le même raisonnement montre que : $\varphi_*\mu(\varphi(\cup K'_i)) \geq \mu(A^c)$.

Comme $\varphi(\cup K_i) \subset \varphi(A)$ et $\varphi(\cup K'_i) \subset \varphi(A^c)$, ces deux ensembles sont disjoints.

$$\varphi_*\mu(\varphi(\cup K_i) \amalg \varphi(\cup K'_i)) = \varphi_*\mu(\varphi(\cup K_i)) + \varphi_*\mu(\varphi(\cup K'_i)) \geq \mu(A) + \mu(A^c) = \varphi_*\mu(Y)$$

Par conséquent, $\varphi(\cup K_i) \cup \varphi(\cup K'_i)$ est de mesure totale. Son complémentaire est négligeable, il contient l'ensemble $\varphi(A) \setminus \varphi(\cup K_i)$, qui est donc mesurable. On en déduit que $\varphi(A)$ est mesurable (cf fig.2).

Remarques

- On a bien sûr $\varphi(X) \cap \varphi(X^c) = \emptyset$. Ceci montre que φ est presque surjective : $\varphi(X)$ est mesurable, $\varphi_*\mu(\varphi(X)^c) = 0$.
- La condition $\varphi(A) \cap \varphi(A^c) = \emptyset$ équivaut à l'égalité $\varphi^{-1}(\varphi(A)) = A$, ou encore à $A = \bigcup_{y \in \varphi(A)} \varphi^{-1}(\{y\})$. Ceci revient à dire que A est union de lignes de niveau de φ .

Définition d'un espace de Lebesgue

Un espace de Lebesgue $(X, \mathcal{T}, \mu)$ est un espace mesuré muni d'une mesure de probabilité μ définie sur une tribu complète $\mathcal{T}$, qui est isomorphe à $([0, 1], \overline{\mathcal{B}([0, 1])}, \lambda)$.

Théorème d'isomorphisme

Soit X un sous-ensemble borélien d'un espace métrique séparable complet, μ une mesure de probabilité borélienne non atomique définie sur X . Alors $(X, \overline{\mathcal{B}(X)}, \mu)$ est un espace de Lebesgue.

Preuve

D'après le théorème d'Oxtoby-Ulam, μ est une mesure intérieurement régulière sur X . Tout sous-ensemble d'un espace métrique séparable étant métrique séparable, on peut trouver une base dénombrable d'ouverts $\{U_i\}_{i \in \mathbf{N}}$ pour la topologie de X .

$$\begin{array}{ccc} \varphi_1 : X & \hookrightarrow & \{0, 1\}^{\mathbf{N}} \\ x & \mapsto & \{\mathbf{1}_{U_i}(x)\}_{i \in \mathbf{N}} \end{array} \quad \begin{array}{ccc} \varphi_2 : \{0, 1\}^{\mathbf{N}} & \hookrightarrow & [0, 1] \\ a_i & \mapsto & \sum \frac{a_i}{3^i} \end{array}$$

L'injection borélienne $\varphi_2 \circ \varphi_1$ établit un isomorphisme entre l'espace $(X, \overline{\mathcal{B}(X)}, \mu)$ et $([0, 1], \overline{\mathcal{B}([0, 1])}, \varphi_2 \circ \varphi_{1*}(\mu))$; cf fig.3. On est ramené à l'exemple précédent.

Lemme de mesurabilité (version mesurable)

Soit φ une application mesurable entre deux espaces de Lebesgue qui préserve la mesure et A un sous-ensemble mesurable satisfaisant $\varphi(A) \cap \varphi(A^c) = \emptyset$. Alors $\varphi(A)$ est mesurable. En particulier, si φ est injective, φ est un isomorphisme.

Preuve

On se ramène à $([0, 1], \overline{\mathcal{B}([0, 1])}, \lambda)$ par isomorphisme et on procède comme plus haut.

Voici une application de ces résultats :

Théorème de Stone-Weierstraß mesurable

Soit $(X, \mathcal{T}, \mu)$ un espace de Lebesgue, $p \in [1, \infty[$, et $f_n : X \rightarrow \mathbf{R}$ une suite de fonctions mesurables bornées qui sépare les points :

$$\forall x, y \in X, x \neq y, \exists n \in \mathbf{N} \text{ tel que } f_n(x) \neq f_n(y).$$

Alors l'algèbre engendrée par les fonctions f_n et les constantes, est dense dans $L^p(X, \mathcal{T}, \mu)$.

Preuve

L'hypothèse de séparation montre que l'application suivante est injective :

$$\begin{array}{ccc} \varphi : X & \hookrightarrow & \mathbf{R}^{\mathbf{N}} \\ x & \mapsto & \{f_i(x)\}_{i \in \mathbf{N}} \end{array}$$

Elle réalise donc un isomorphisme de $(X, \mathcal{T}, \mu)$ sur $(\mathbf{R}^{\mathbf{N}}, \overline{\mathcal{B}(\mathbf{R}^{\mathbf{N}})}, \varphi_*\mu)$. L'ensemble $\varphi(X)$ est un sous-ensemble du compact $K = \prod_{i \in \mathbf{N}} [-\|f_i\|, \|f_i\|]$; la mesure $\varphi_*\mu$ est supportée par K . L'application φ induit donc une isométrie inversible de $L^p(X, \mathcal{T}, \mu)$ sur $L^p(K, \overline{\mathcal{B}(K)}, \varphi_*\mu)$, qui fait correspondre aux fonctions f_i les projections sur les coordonnées. Ces projections sont des fonctions continues sur K qui séparent les points de K . Par le théorème de Stone-Weierstraß usuel, l'algèbre engendrée par ces projections est uniformément dense dans $C(K)$, et $C(K)$ est dense dans $L^p(K, \overline{\mathcal{B}(K)}, \varphi_*\mu)$, cf annexes.

figure 1 : Isomorphisme entre $[0,1]$ et $[0,1/3] \cup [2/3,1]$,
munis de la mesure de Lebesgue renormalisée

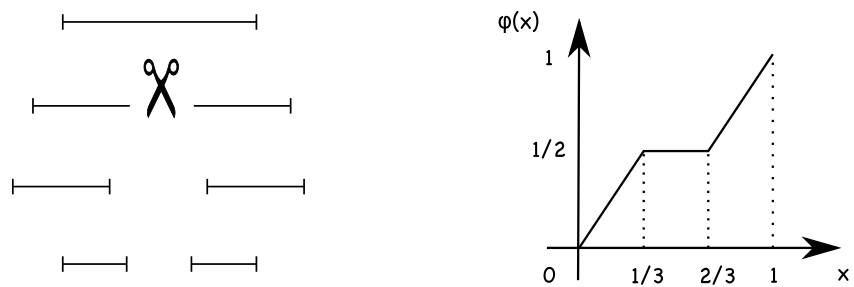


figure 2 : Image d'un ensemble borélien

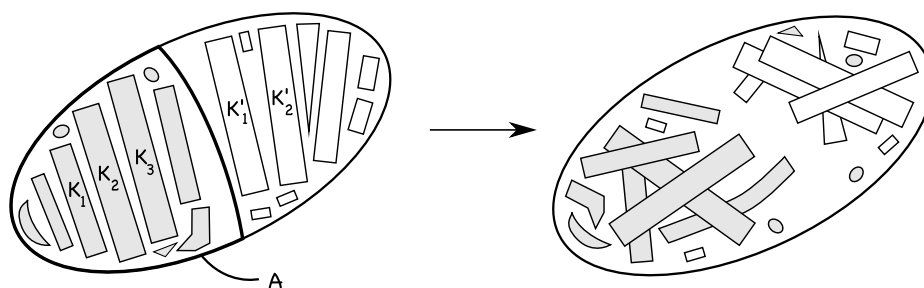
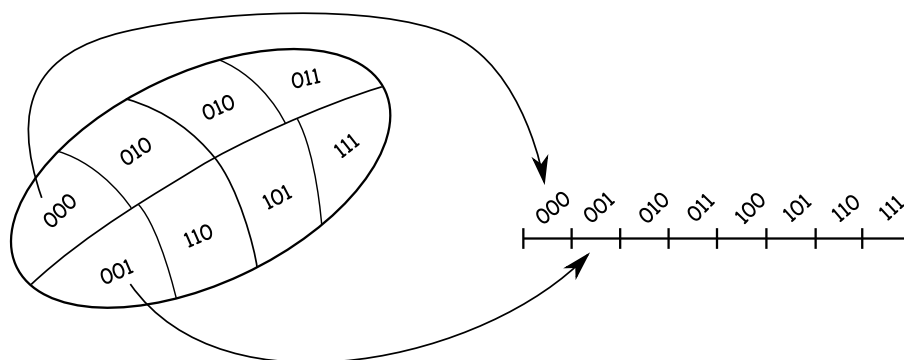


figure 3 : Construction de l'isomorphisme



Exercices

exercice 1 :

Soit A, B deux sous-ensembles boréliens de $[0, 1]$, $\varphi : A \rightarrow B$ une application croissante de A dans B . Montrez que φ est borélienne.

exercice 2 :

Soit $(X, \mathcal{T}, \mu)$ et $(Y, \mathcal{S}, \nu)$ deux espaces de Lebesgue, $\varphi : X \rightarrow Y$ une application mesurable et $A \in \mathcal{T}$.

- Montrez qu'on peut trouver un ensemble $A' \subset A$ tel que $\mu(A \setminus A') = 0$ et $\varphi(A') \in \mathcal{S}$.
- L'ensemble $\varphi(A \setminus A')$ est-il nécessairement mesurable ?

exercice 3 :

– Donnez un exemple d'ensemble $N \subset [0, 1]$ négligeable pour la mesure de Lebesgue, qui peut être mis en bijection avec $\mathbf{R}$. Construisez un sous-ensemble M de $[0, 1]$ qui n'est pas mesurable pour la mesure de Lebesgue, et qui est en bijection avec $\mathbf{R}$.

– On définit une fonction φ comme suit : φ est égale à l'identité sur $[0, 1] \setminus N$, et φ est une bijection de N sur M . Montrez que $\varphi : [0, 1] \rightarrow [0, 1]$ est une application mesurable qui préserve la mesure de Lebesgue.

– Montrez que l'image d'un ensemble négligeable, par une application mesurable entre deux espaces de Lebesgue qui préserve la mesure, n'est pas forcément négligeable.

exercice 4 :

Soit X un espace topologique séparé, à base dénombrable d'ouverts, μ une mesure de probabilité borélienne non atomique intérieurement régulière. Montrez que $(X, \overline{\mathcal{B}(X)}, \mu)$ est un espace de Lebesgue.

exercice 5 :

Soit $(X, \mathcal{T}, \mu)$ un espace de Lebesgue, $\{B_i\}_{i \in \mathbf{N}}$ une famille dénombrable d'éléments de $\mathcal{T}$. Supposons que pour tout x, y dans X , $x \neq y$, on peut trouver $i \in \mathbf{N}$ tel que $x \in B_i$, $y \notin B_i$. Montrez que les B_i et les ensembles négligeables engendrent $\mathcal{T}$.

exercice 6 :

Soit $f : [0, 1] \rightarrow \mathbf{R}$ et λ la mesure de Lebesgue sur $[0, 1]$. Calculez $f_*\lambda$ quand :

- l'application f est constante,
- l'application f est C^1 injective,
- l'application f est affine par morceaux.

exercice 7 :

Soit B un borélien dans un espace métrique complet, muni d'une mesure borélienne finie sans atome. Montrez que pour tout $t \in [0, \mu(B)]$, on peut trouver un ensemble borélien $A \subset B$ tel que $\mu(A) = t$.

exercice 8 :

Soit μ une mesure borélienne finie définie sur un espace métrique séparable X . Soit A un borélien de X satisfaisant la propriété suivante : pour tout $B \subset A$ mesurable, $\mu(B) = 0$ ou $\mu(A \setminus B) = 0$. Montrez qu'on peut trouver $x \in A$ tel que $\mu(A \setminus \{x\}) = 0$.

exercice 9 :

Les espaces de Lebesgue que nous avons considérés dans cette partie sont des espaces probabilisés *sans atomes*. On peut aussi définir des espaces de Lebesgue de mesure finie avec atomes : ce sont les espaces mesurés isomorphes à $[0, a] \cup \mathbf{N}$ muni de la mesure $\lambda_{[0, a]} + \sum_{i \in \mathbf{N}} a_i \delta_i$, avec $a, a_i \in \mathbf{R}_+$ et $a + \sum a_i < +\infty$.

- Classez ces espaces à isomorphisme près.
- Montrez que le lemme de mesurabilité reste vrai pour ces espaces.
- Montrez que le théorème de Stone-Weierstraß est encore vrai pour ces espaces.
- Montrez que tout borélien dans un espace métrique séparable complet, muni d'une mesure borélienne finie, est isomorphe à un tel espace.

Commentaires

Les espaces de Lebesgue présentés ci-dessus sont des espaces probabilisés sans atomes. On peut aussi définir des espaces de Lebesgue σ -fini avec atomes : un tel espace est isomorphe à l'union d'un intervalle de la forme $[0, a[$, $a \in [0, +\infty[$, muni de sa mesure de Lebesgue, et d'un nombre au plus dénombrable de points de mesure strictement positive. Un tel espace est déterminé à isomorphisme près par la suite des masses des atomes et par a . Le théorème d'isomorphisme s'étend sans difficulté : *Tout borélien d'un espace métrique séparable complet, muni d'une mesure borélienne σ -finie, est un espace de Lebesgue σ -fini avec atomes.*

Un espace métrique séparable peut toujours être plongé dans un espace métrique séparable complet ; par exemple sa complétion. Mais rien ne garantit qu'il soit un sous-espace borélien de sa complétion. Si c'est le cas, on dit que cet espace topologique est un *espace borélien standard*. Un tel espace devient un espace de Lebesgue si on le munit d'une mesure de probabilité borelienne.

On peut donner une version plus forte du lemme de mesurabilité : *l'image d'un borélien par une application borélienne injective, définie entre deux espaces boréliens standards, est un borélien*. Ce résultat est plus délicat à établir que les lemmes donnés plus haut ; la preuve transite en général par la théorie des ensembles analytiques. Des preuves raisonnables se trouvent par exemple dans Cohn [Co93] ch7, ou Dudley [Du02] ch 13. Sans hypothèse d'injectivité, on a le résultat suivant : considérons une application borélienne définie entre deux espaces boréliens standards ; alors l'image d'un borélien est *universellement mesurable*, c'est-à-dire mesurable relativement à toute mesure borélienne finie définie sur l'espace d'arrivée.

Le théorème de Lusin admet une version purement topologique : soit X un espace métrique séparable et $f : X \rightarrow \mathbf{R}$ une fonction borélienne ; alors on peut trouver un sous-ensemble G_δ dense en restriction duquel f est continue.

Les compacts donnés par le théorème de Lusin ne sont pas connexes en général, contrairement à ce que pourrait laisser croire la figure 2. On peut montrer que pour un ensemble G_δ dense d'applications boréliennes $f : [0, 1] \rightarrow \mathbf{R}$ (dans L^p par exemple, $p \in [0, \infty[$), il n'existe pas d'ouvert sur lequel f est continue.

Soit $f : [0, 1] \rightarrow [0, 1]$ une application borélienne. En général l'image réciproque d'un ensemble Lebesgue mesurable n'a pas de raison d'être Lebesgue mesurable. On a vu dans les exercices que c'est le cas si la mesure $f_*\lambda$ est égale à λ . Plus généralement, c'est le cas si f satisfait la propriété suivante : un borélien de $[0, 1]$ est négligeable si et seulement si son image réciproque par f est négligeable. On dit alors que f est *non singulière*.

Voici deux types d'espaces probabilisés qui ne sont pas des espaces de Lebesgue :

- les espaces trop gros ; par exemple un produit non dénombrable de cercles, muni de sa mesure de Haar.
- les espaces trop petits ; par exemple la restriction de la mesure de Lebesgue à un sous-ensemble non mesurable de $[0, 1]$.

Il existe d'autres notions d'espaces probabilisés “sympathiques” : espaces standards, classes compactes de Marczewski, espaces parfaits de Gnedenko et Kolmogorov, espaces lusiniens etc. cf par exemple M. M. Rao [Ra05]. Toutes ces notions se ramènent à supposer peu ou prou que toute sous-tribu séparable est une tribu de boréliens, pour une topologie *ad hoc* relativement à laquelle X est presque σ -compact. Ces notions sont équivalentes entre elles ; leur but est de récupérer le théorème de désintégration énoncé dans la suite. La terminologie “espace de Lebesgue” est introduite par V. A. Rokhlin [Rok52]. Le lemme de mesurabilité apparaît dans un article de J. Haezendonck (1973).

Il existe des versions du théorème de Stone-Weierstraß mesurable pour les mesures de Baire sur les espaces localement compacts ; cf R. H. Farrell , S. Cater ou R. J. Nagel.

Soit $(X, \mathcal{T}, \mu)$ un espace de Lebesgue. Notons $L^0(X)$ l'ensemble des fonctions mesurables à valeurs réelles, muni de la topologie de la convergence en probabilité. Cette topologie correspond à la distance :

$$d(f, g) = \inf_{\varepsilon > 0} \{ \varepsilon + \mu(\omega \mid d(f(\omega), g(\omega)) > \varepsilon) \}.$$

Le théorème de Stone-Weierstrass mesurable est encore vrai dans cet espace, et il n'est pas nécessaire de supposer les fonctions bornées dans l'énoncé du théorème. La preuve est inchangée.

Les espaces de Lebesgue jouissent d'un certain nombre de propriétés, qui ne sont pas vraies pour tous les espaces probabilisés. Le théorème de Stone-Weierstraß est un exemple. Un autre exemple est donné par le résultat suivant : soit $(X, \mathcal{T}, \mu)$ un espace de Lebesgue ; tout morphisme de σ -algèbre de $\mathcal{T}$ dans $\mathcal{T}$ est de la forme $A \mapsto F^{-1}A$, où $F : X \rightarrow X$ est une application mesurable.

On peut chercher des équivalents de nature topologique au théorème d'isomorphisme. Il est par exemple possible d'associer une représentation symbolique à chacun des points d'un espace métrique compact K : on peut toujours construire une surjection continue $\varphi : \{0, 1\}^{\mathbf{N}} \rightarrow K$, si bien que l'espace K est homéomorphe au quotient de $\{0, 1\}^{\mathbf{N}}$ par la relation d'équivalence $x \sim y \leftrightarrow \varphi(x) = \varphi(y)$.

Dans certains cas spécifiques, on peut construire des isomorphismes entre espaces mesurés qui sont continus. Par exemple, deux mesures de probabilité sans atomes, de support total, définies sur le cube $[0, 1]^n$, ne chargeant pas le bord du cube, se déduisent l'une de l'autre par le biais d'un homéomorphisme (Oxtoby-Ulam, 1941). Ce résultat est faux sur $\{0, 1\}^{\mathbf{N}}$; sur cet espace, le sous-ensemble de $\mathbf{R}$ donné par les mesures associées aux ouverts fermés (i.e. aux cylindres) est un invariant non trivial d'homéomorphisme. Le problème de la classification des mesures sur $\{0, 1\}^{\mathbf{N}}$ à homéomorphisme près reste ouvert à ce jour.

Spectre des systèmes dynamiques

His way had therefore come full circle, or rather had taken the form of an ellipse or a spiral, following as ever no straight unbroken line, for the rectilinear belongs only to Geometry and not to Nature and Life.

Hermann Hesse

Pour classer les systèmes dynamiques mesurés à isomorphisme près, on peut regarder leur action sur l'espace des fonctions de carré intégrable à valeurs complexes. Cet espace est un espace de Hilbert et l'action obtenue est une isométrie. Si deux systèmes sont isomorphes, les isométries de L^2 associées sont conjuguées.

Soit $(X, \mathcal{T}, \mu)$ un espace probabilisé et $T : X \rightarrow X$ une transformation qui préserve la mesure μ . On considère l'opérateur $U(f) = f \circ T$ défini sur l'espace $L^2(X, \mathbf{C})$ des classes de fonctions de carré intégrable à valeurs complexes, muni du produit hilbertien

$$\langle f, g \rangle = \int_X f \bar{g} d\mu.$$

Les invariants de conjugaison de cet opérateur sont des invariants d'isomorphisme pour T . Ils sont appelés *invariants spectraux* du système. L'invariant de conjugaison le plus simple de U est donné par ses valeurs propres et leurs multiplicités. Les vecteurs propres associés sont appelés *fonctions propres* du système dynamique, ce sont les fonctions qui satisfont l'égalité $f \circ T = \lambda f$ presque partout pour une certaine valeur propre $\lambda \in \mathbf{C}$, cf fig.1 et 2.

Les vecteurs propres d'une isométrie définie sur un espace de Hilbert de dimension finie forment toujours une base hilbertienne de l'espace et on peut diagonaliser l'opérateur dans cette base, ce qui permet d'identifier cet opérateur à conjugaison près. Il n'en va pas de même en dimension infinie. Nous allons voir que les opérateurs associés aux systèmes dynamiques mélangeants n'ont pas de vecteurs propres, à part les fonctions constantes. L'approche spectrale ne donne donc pas d'information sur ces systèmes du point de vue de la conjugaison.

À l'opposé, les translations du tore préservant le volume possèdent une base hilbertienne de fonctions propres, on dit qu'elles sont à *spectre discret*. On va montrer qu'une transformation ergodique à spectre discret est caractérisée à isomorphisme près par l'ensemble de ses valeurs propres, et qu'elle est conjuguée à une translation sur un groupe abélien compact, muni de sa mesure de Haar.

Rappelons que tout groupe topologique compact admet une unique mesure de probabilité invariante par translation, appelée *mesure de Haar*. Une preuve courte de l'unicité est proposée en exercice pour les groupes abéliens compacts qui sont considérés dans ce chapitre.

Voici quelques exemples de groupes abéliens compacts : $\mathbf{Z}/n\mathbf{Z}$, le cercle unité $S^1 \simeq \mathbf{R}/2\pi\mathbf{Z}$, l'anneau des entiers p-adiques $\mathbf{Z}_p$, les produits de tels espaces comme les tores de dimension finie $(S^1)^n$ ou infinie $(S^1)^{\mathbf{N}}$. Ces exemples sont métrisables, la topologie produit sur $(S^1)^{\mathbf{N}}$ est induite par la distance

$$d(\{x_i\}_{i \in \mathbf{N}}, \{y_i\}_{i \in \mathbf{N}}) = \sum_i \frac{1}{2^i} d_{S^1}(x_i, y_i)$$

qui est invariante par translation. De manière générale, les groupes abéliens compacts sont métrisables si et seulement si ils sont séparables. Un exemple de groupe compact non séparable est donné par un produit infini non dénombrable de cercles. On va se limiter au cas séparable dans ce chapitre. On peut montrer que ces groupes s'identifient aux sous-groupes fermés de l'ensemble $(S^1)^{\mathbf{N}}$. Une façon simple de construire de tels sous-groupes est d'imposer des relations linéaires entre les coordonnées des éléments de $(S^1)^{\mathbf{N}}$.

Valeurs propres des systèmes dynamiques

Définition

Soit $(X, \mathcal{T}, \mu)$ un espace mesuré fini, $T : X \rightarrow X$ une application mesurable qui préserve μ . Le nombre complexe $\lambda \in \mathbf{C}$ est valeur propre de T s'il existe une fonction $f : X \rightarrow \mathbf{C}$ de carré intégrable non nulle satisfaisant l'égalité $f \circ T = \lambda f$ presque partout. La fonction f est une fonction propre associée à λ .

Proposition

- Les valeurs propres sont de module égal à un.
- Si la transformation est ergodique, le module des fonctions propres est constant et les valeurs propres sont de multiplicité un.
- Deux fonctions propres associées à des valeurs propres différentes sont orthogonales entre elles.

Preuve

Soient f et λ tels que $f \circ T = \lambda f$. Prenons le module et intégrons cette égalité.

$$\int |f| d\mu = \int |f \circ T| d\mu = |\lambda| \int |f| d\mu.$$

Ceci implique $|\lambda| = 1$. La fonction $|f|$ est donc invariante par T , constante presque partout si T est ergodique.

Soient f_1 et f_2 deux fonctions propres associées à la valeur propre λ . Le quotient $\frac{f_1}{f_2}$ est une fonction propre associée à la valeur propre 1, il est constant si la transformation est ergodique. L'espace propre associé à λ est alors de dimension un.

Enfin, si f_1 et f_2 sont deux fonctions propres associées à des valeurs propres distinctes λ_1 et λ_2 ,

$$\int f_1 \overline{f_2} d\mu = \int f_1 \circ T \overline{f_2 \circ T} d\mu = \lambda_1 \overline{\lambda_2} \int f_1 \overline{f_2} d\mu.$$

Le nombre complexe $\lambda_1 \overline{\lambda_2} = \lambda_1 / \lambda_2$ est différent de un, les fonctions f_1 et f_2 sont orthogonales. Ceci termine la preuve.

Les fonctions constantes sont toujours fonctions propres associées à la valeur propre 1. La transformation T est ergodique s'il n'y a pas d'autres fonctions propres associées à la valeur propre 1 c'est-à-dire si 1 est valeur propre simple de l'opérateur $f \mapsto f \circ T$. Il est possible qu'il n'y ait pas d'autres valeurs propres.

Définition

Une transformation $T : X \rightarrow X$ définie sur un espace mesuré fini $(X, \mathcal{T}, \mu)$ et préservant la mesure, est dite faiblement mélangeante relativement à μ si toutes ses fonctions propres sont constantes.

Proposition

Une transformation mélangeante relativement à une mesure de probabilité μ est faiblement mélangeante relativement à cette mesure.

Preuve

Soit f et λ tels que $f \circ T = \lambda f$ et $g \in L^2$ une fonction d'intégrale nulle. D'après la propriété de mélange,

$$\int f \overline{g} d\mu = \lambda^{-n} \int f \circ T^n \overline{g} d\mu \xrightarrow{n \rightarrow \infty} 0.$$

La fonction f est orthogonale à toutes les fonctions d'intégrale nulle, elle est donc constante. Ceci termine la preuve.

La multiplication $x \mapsto 2x$, définie sur $\mathbf{R}/\mathbf{Z}$, est mélangeante relativement à la mesure de Lebesgue, elle ne possède pas d'autres valeurs propres que 1. À l'opposé, la translation $x \mapsto x + \alpha$ sur $\mathbf{R}/\mathbf{Z}$ admet comme valeurs propres les nombres complexes $e^{2i\pi k\alpha}$, $k \in \mathbf{Z}$, associés aux fonctions propres $x \mapsto e^{2i\pi kx}$. Ces fonctions forment une base de L^2 , on dit que le spectre est discret.

Définition

Un système est à spectre discret s'il existe une base hilbertienne de L^2 constituée de fonctions propres de la transformation.

Il est possible de classer à isomorphisme près les systèmes à spectre discret. On va se restreindre aux systèmes définis sur des espaces de Lebesgue.

Théorème (Halmos, Von Neumann)

Un système dynamique probabilisé, défini sur un espace de Lebesgue sans atome $(X, \mathcal{T}, \mu)$, ergodique et à spectre discret, est isomorphe à une translation sur un groupe métrique abélien compact.

Preuve

On note $(\lambda_k)_{k \in \mathbf{N}}$ les valeurs propres de T , et f_k les fonctions propres associées, normalisées de façon à être de module un. Définissons une application φ , à valeurs dans l'espace $(S^1)^{\mathbf{N}}$ des suites de nombres complexes de module un, comme suit :

$$\begin{aligned} f : X &\rightarrow (S^1)^{\mathbf{N}} \\ x &\mapsto (f_k(x))_{k \in \mathbf{N}} \end{aligned}$$

Vérifions que l'application f est presque injective, c'est-à-dire injective après restriction à un sous-ensemble de mesure totale. Soit $\varphi : X \rightarrow [0, 1]$ un isomorphisme provenant de la structure d'espace de Lebesgue. Il est presque injectif. Décomposons φ sur la base orthonormée des f_k . Quitte à extraire, la série obtenue converge presque partout vers φ , si bien que deux points de X pour lesquels tous les f_k coïncident, ont même image par φ . Le résultat s'ensuit.

La multiplication coordonnées par coordonnées munit l'espace $(S^1)^{\mathbf{N}}$ d'une structure de groupe métrique abélien compact. Posons $\lambda = (\lambda_k)_{k \geq 0}$. D'après le lemme de mesurabilité, l'application f réalise un isomorphisme entre X et $(S^1)^{\mathbf{N}}$ qui conjugue l'application T à la translation $S_\lambda(x) = \lambda x$:

$$f \circ T = (f_k \circ T)_{k \geq 0} = (\lambda_k f_k)_{k \geq 0} = \lambda f = S_\lambda \circ f$$

L'application S_λ est ergodique relativement à la mesure image $\nu = f_*\mu$. Soit y un point du support de ν dont l'orbite est dense dans ce support :

$$\text{supp } \nu = \overline{\{y\lambda^k \in (S^1)^{\mathbf{N}} \mid k \in \mathbf{Z}\}} = y\overline{\lambda^{\mathbf{Z}}}$$

La translation $S_{y^{-1}}$ commute avec S_λ . L'application $S_{y^{-1}} \circ f$ réalise un isomorphisme de X sur le groupe $G = \overline{\lambda^{\mathbf{Z}}}$ et conjugue T à la restriction de S_λ à G , cf fig. 3. La mesure image de ν par $S_{y^{-1}}$, notée λ_G , est invariante par S_λ , donc par l'action du groupe $\lambda^{\mathbf{Z}}$ et par densité, par l'action du groupe G . On a obtenu un isomorphisme de $(X, \mathcal{T}, \mu)$ sur $(G, \mathcal{B}(G), \lambda_G)$ qui conjugue T à S_λ .

Remarque

Un groupe métrique compact possède une unique mesure de probabilité invariante par translation, appelée mesure de Haar. Une preuve est proposée dans les exercices. Un système ergodique à spectre discret est donc déterminé à isomorphisme près par l'ensemble de ses valeurs propres.

Figure 1 : fonction propre

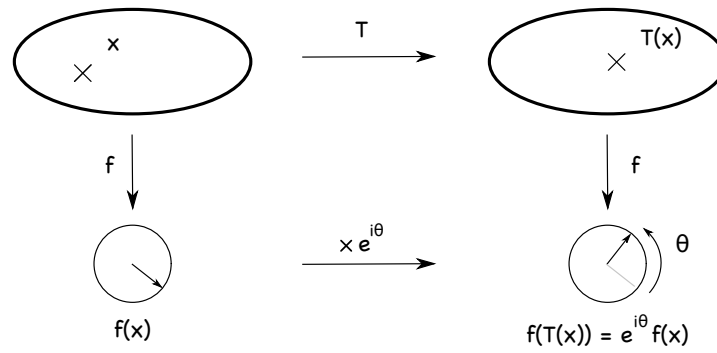


Figure 2 : l'angle donné par une fonction propre augmente de la même quantité à chaque itération

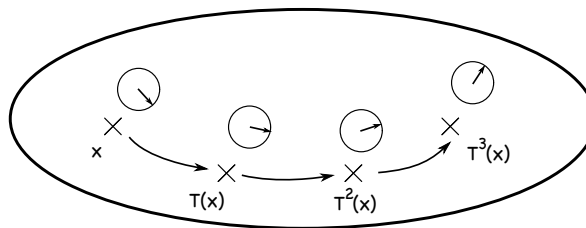
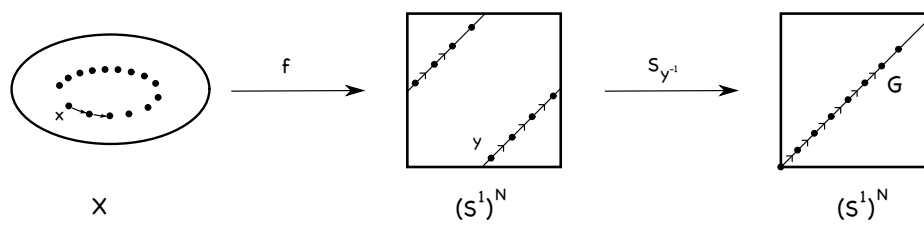


Figure 3 : construction de la conjugaison



Exercices

exercice 1 :

Montrer que les valeurs propres d'une transformation forment un sous-groupe de S^1 .

exercice 2 :

Calculer les valeurs propres et les fonctions propres d'une translation sur le tore $\mathbf{T}^d = \mathbf{R}^d / \mathbf{Z}^d$ donnée par $T_\theta : x \mapsto x + \theta$, $\theta \in \mathbf{T}^d$.

exercice 3 :

Montrer que deux systèmes dynamiques mesurés isomorphes ont les mêmes valeurs propres. En déduire que deux translations irrationnelles T_θ et $T_{\theta'}$ sur $\mathbf{R}/\mathbf{Z}$, $\theta, \theta' \notin \mathbf{Q}$, sont isomorphes si et seulement si $\theta = \pm\theta' \bmod \mathbf{Z}$.

exercice 4 :

Montrer que si T est une transformation faiblement mélangeante, il en va de même de ses itérés T^n , $n > 0$.

exercice 5 :

Montrer que tous les itérés positifs T^k , $k > 0$, d'une transformation T sont ergodiques si et seulement si 1 est la seule valeur propre de T racine de l'unité et elle est de multiplicité un. Une telle transformation est parfois dite *totalelement ergodique*.

exercice 6 :

Montrer qu'un système à spectre discret est inversible. Montrer qu'il est conjugué à son inverse.

exercice 7 :

Soit $\lambda \in (S^1)^{\mathbf{N}}$; montrer qu'il existe une suite $n_k \rightarrow +\infty$ telle que $\lambda^{n_k} \rightarrow 1$.

Indication : considérer la suite λ^n et extraire, ou utiliser un théorème de récurrence.

En déduire que $\overline{\lambda^{\mathbf{N}}} = \overline{\lambda^{\mathbf{Z}}}$.

Étant donné une isométrie agissant sur un espace métrique compact, montrer que l'orbite positive d'un point quelconque est dense dans l'orbite totale du point et que l'isométrie est surjective.

exercice 8 :

Soit G un groupe métrique abélien compact muni d'une mesure de probabilité invariante par les translations de G et $f : G \rightarrow \mathbf{C}$ une fonction intégrable.

– Montrer que l'ensemble $\{g \in G \mid f(gx) = f(x) \text{ pour presque tout } x \in G\}$ est un sous-groupe fermé de G .

Indication : approcher f par une fonction continue.

– Soit $\lambda \in (S^1)^{\mathbf{N}}$ et $G = \overline{\lambda^{\mathbf{Z}}} \subset (S^1)^{\mathbf{N}}$. Montrer que $S_\lambda : x \mapsto \lambda x$ est ergodique relativement à toute mesure de probabilité sur G invariante par translation.

– En déduire qu'il existe une unique mesure de probabilité sur G invariante par translation (cas particulier du théorème d'unicité de la mesure de Haar).

– Soit G un groupe métrique abélien compact et T une translation sur G . Montrer que si T est transitive, elle est ergodique relativement à la mesure de Haar de G .

Commentaires

Le théorème de classification des systèmes à spectre discret est dû à P. Halmos et J. Von Neumann (1942). Si l'espace sous-jacent n'est pas un espace de Lebesgue, on peut démontrer que le système est isomorphe à une translation sur un groupe abélien compact, mais la topologie du groupe n'est plus métrisable. La preuve générale fait appel à la théorie des caractères sur les groupes abéliens compacts, théorie qui généralise les séries de Fourier sur $\mathbf{R}/\mathbf{Z}$.

Le mélange faible admet les caractérisations suivantes : T est faiblement mélangeante si et seulement si pour tout $f, g \in L^2$, $\left| \int f \circ T^k g \, d\mu - \int f \, d\mu \int g \, d\mu \right|$ converge vers 0 le long d'une suite d'entiers k de densité un, ou de manière équivalente si

$$\frac{1}{N} \sum_{k=0}^{N-1} \left| \int f \circ T^k g \, d\mu - \int f \, d\mu \int g \, d\mu \right| \rightarrow 0.$$

Cette dernière propriété est souvent prise comme définition du mélange faible. L'équivalence avec la définition que nous avons donnée en terme de fonctions propres est habituellement démontrée en utilisant le théorème spectral. Le mélange faible est aussi équivalent à la propriété suivante, appelée ergodicité double : pour tout A, B de mesure non nulle, on peut trouver $n > 0$ tel que $\mu(A \cap T^{-n}A) > 0$ et $\mu(A \cap T^{-n}B) > 0$; voir le livre de Furstenberg "Recurrence in Ergodic Theory and Combinatorial Number Theory".

Le mélange faible est aussi défini pour les flots : une fonction $f \in L^2$ est fonction propre associée à la valeur propre $e^{i\theta}$ pour un flot $\{\varphi_t\}$ si pour tout $t \in \mathbf{R}$, on a l'égalité $f \circ \varphi_t = e^{i\theta t} f$ pp. Un flot est faiblement mélangeant si toutes ses fonctions propres sont constantes presque partout. On caractérise cette propriété en terme de moyenne comme suit : pour tout $f, g \in L^2$,

$$\frac{1}{T} \int_0^T \left| \int f \circ \varphi_t g \, d\mu - \int f \, d\mu \int g \, d\mu \right| dt \rightarrow 0$$

De manière remarquable, un flot $\{\varphi_t\}$ est faiblement mélangeant si et seulement si pour tout $t \neq 0$, les transformations $x \mapsto \varphi_t(x)$ sont ergodiques. Ce résultat n'a pas d'analogue pour les transformations : une rotation irrationnelle sur le cercle est à spectre discret mais tous ses itérés sont ergodiques.

La notion de mélange faible pour les flots dépend du paramétrage. Tout flot ergodique, dont l'ensemble des points périodiques est négligeable, peut être reparamétrisé de façon à ce qu'il devienne faiblement mélangeant (Chacon 1966) et même mélangeant (Kocergin 1973).

Il existe un analogue topologique à la notion de mélange faible. Une transformation est *topologiquement faiblement mélangeante* si toutes ses fonctions propres continues sont constantes. Un système faiblement mélangeant relativement à une mesure de support total est topologiquement faiblement mélangeant, mais on peut produire des exemples de transformations topologiquement faiblement mélangeantes qui sont à spectre discret.

Pour les transformations continues définies sur un espace compact X , il existe aussi une notion de spectre topologique discret : le sous-espace des fonctions propres continues est dense dans l'ensemble des fonctions continues. On montre alors qu'on peut trouver une distance sur X pour laquelle la transformation est une isométrie si elle est transitive. Ceci est détaillé dans le livre de P. Walters ("An introduction to ergodic theory", 5.5).

Tous les exemples de transformations dans ces notes qui sont faiblement mélangeantes sont fortement mélangeantes. Pour obtenir des exemples de transformations non mélangeantes qui sont faiblement mélangeantes, on peut se placer du point de vue catégorique : étant donnée une transformation, on se demande si le mélange faible est une propriété typique au sens de Baire, dans l'espace de toutes les mesures de probabilité invariantes, muni de la topologie faible. C'est le cas pour les transformations uniformément hyperboliques comme par exemple l'automorphisme du tore associé à la matrice $\begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$ (K. Sigmund 1972). Pour ces systèmes, les mesures faiblement mélangeantes forment un sous-ensemble G_δ -dense tandis que celles qui sont fortement mélangeantes sont contenues dans le complémentaire d'un ensemble G_δ -dense. Il existe donc beaucoup de mesures invariantes, singulières relativement à la mesure de Lebesgue, pour lesquelles la transformation est faiblement mélangeante sans être fortement mélangeante.

Pour des exemples explicites de transformations faiblement mélangeantes relativement à une mesure absolument continue par rapport à Lebesgue, mais pas fortement mélangeantes, on peut se référer à une construction décrite dans le livre de W. Parry ("Topics in ergodic theory", chapitre 5.7.) due à S. Kakutani et J. Von Neuman, et obtenue par une méthode de "cut and stack".

Les échanges d'intervalles forment une autre classe de transformations qui préservent une mesure absolument continue et qui ne sont jamais fortement mélangeantes relativement à cette mesure (A. Katok, 1980). A. Avila et G. Forni montrent en 2007 que la plupart de ces transformations sont faiblement mélangeantes.

Décomposition ergodique

The title of Rokhlin's paper [On the fundamental ideas of measure theory] seems to suggest that measurable partitions are the main object of measure theory. But probably this would seem very doubtful to most analysts ("all my life I have worked with the Lebesgue integral, and this is the first time I have heard about measurable partitions")

D. V. Anosov

Lorsqu'un système n'est pas ergodique, il est possible de décomposer l'espace ambiant en plusieurs morceaux, de façon à ce que la transformation soit ergodique en restriction à chacun des morceaux. On parle alors de *partition en composantes ergodiques*. Ces composantes peuvent être en nombre non dénombrable, mais la partition obtenue satisfait tout de même une certaine propriété de régularité : il est possible de l'approcher par des partitions ayant un nombre fini de morceaux.

Stricto sensu, construire la partition en composantes ergodiques n'est pas difficile. De manière générale, à toute σ -algèbre on peut associer une partition, de telle sorte que les fonctions mesurables relativement à cette σ -algèbre sont exactement les fonctions constantes sur les éléments de la partition (cf annexe). Il suffit donc de considérer la partition associée à la σ -algèbre des fonctions invariantes par la transformation ; un ensemble mesurable est invariant mod 0 si et seulement si c'est une union d'éléments de cette partition, à un ensemble négligeable près.

Pour être en mesure de parler de l'ergodicité de la transformation en restriction à chacune de ses composantes ergodiques, il ne suffit pas de partitionner l'espace en ces composantes, il faut aussi obtenir une mesure invariante sur chacune des composantes, à partir de la mesure définie sur l'espace tout entier (cf fig. 1).

Ces mesures sont obtenues par *désintégration*. Une telle désintégration est possible dès l'instant où on travaille avec une partition ξ *mesurable* au sens suivant : il existe une suite de partitions ξ_n dont les éléments sont mesurables, en nombre fini, et qui engendre ξ de la façon suivante : $\xi(x) = \bigcap_{n \in \mathbf{N}} \xi_n(x)$, pp x . On peut sans perte de généralité supposer dans cette définition que les partitions ξ_n ont chacune deux éléments B_n et B_n^c , auquel cas la condition de mesurabilité peut s'exprimer en terme de la suite $\{B_n\}$ seulement (cf fig.2) :

$$\xi(x) = \bigcap_{\substack{n \in \mathbf{N} \\ \text{t.q. } x \in B_n}} B_n \bigcap_{\substack{n \in \mathbf{N} \\ \text{t.q. } x \in B_n^c}} B_n^c$$

La notion de décomposition ergodique est une notion purement mesurable. Lorsqu'on travaille avec des mesures définies sur des espaces topologiques ou des espaces métriques, il est intéressant d'avoir une définition géométrique des composantes ergodiques. Pour cela, on peut s'inspirer de l'argument de Hopf. Plutôt que d'apparier les points dont les trajectoires sont asymptotiques, on peut apparier les points dont les sommes de Birkhoff, associées à n'importe quelle fonction continue (ou lipschitzienne) bornée, sont asymptotiques. Ce faisant on obtient une réalisation borélienne des composantes ergodiques, qui ne dépend que de la topologie de l'espace. La mesure sélectionne alors les composantes des points qui sont typiques, du point de vue du théorème ergodique de Birkhoff. Et lorsqu'on définit les composantes à l'aide des fonctions lipschitziennes, l'argument de Hopf devient une conséquence immédiate du théorème de décomposition ergodique.

Pour exprimer les composantes ergodiques de manière topologique, on va se placer sur des *espaces Borel standard* : on dira qu'un espace topologique X est Borel standard s'il est homéomorphe à un sous-ensemble borélien d'un espace métrique complet. Un espace Borel standard muni d'une mesure de probabilité borélienne est bien sûr un espace de Lebesgue, et tout espace de Lebesgue est isomorphe à un espace Borel standard muni d'une mesure de probabilité (après complétion).

Désintégration

Une partition $\{\xi(x)\}$ est dite mesurable si on peut trouver des ensembles mesurables B_n , chacun union d'éléments de la partition, tels que, pp x , $\xi(x) = \bigcap_{B_n \ni x} B_n \bigcap_{B_n^c \ni x} B_n^c$

Théorème :

Soit $(X, \mathcal{T}, \mu)$ un espace de Lebesgue et ξ une partition mesurable. Alors il existe une unique famille de mesures de probabilité $\mu_{\xi(x)}$ portées par les éléments de la partition, et satisfaisant, pour tout $A \in \mathcal{T}$:

- l'ensemble A est $\mu_{\xi(x)}$ -mesurable pour μ -presque tout x ;
- la fonction $x \mapsto \mu_{\xi(x)}(A)$ est μ -intégrable et son intégrale vaut $\mu(A)$:

$$\mu(A) = \int \mu_{\xi(x)}(A) d\mu(x).$$

Preuve :

On peut supposer que X est de la forme $\{0, 1\}^{\mathbb{N}}$. Notons par $\mathcal{C}$ l'algèbre des unions finies de cylindres de $\{0, 1\}^{\mathbb{N}}$ et par $\mathcal{B}$ la tribu engendrée par $\mathcal{C}$. L'algèbre $\mathcal{C}$ satisfait le critère de Kolmogorov : toute suite décroissante d'éléments de cette famille, dont l'intersection est vide, est vide à partir d'un certain rang ; elle est donc de mesure nulle relativement à toute mesure finiment additive définie sur $\mathcal{C}$. Par conséquent toute mesure finiment additive définie sur $\mathcal{C}$ admet une unique extension σ -additive à $\mathcal{B}$. Il suffit donc de définir les $\mu_{\xi(x)}$ sur l'algèbre engendrée par les cylindres.

Soit $\hat{\xi} \subset \mathcal{T}$ la tribu des ensembles qui sont unions d'éléments de la partition. Supposons la relation intégrale vérifiée. Considérons $C \in \mathcal{C}$ et $A \in \hat{\xi}$. Comme A est constante sur les éléments de ξ , on doit avoir $\mu(A \cap C) = \int_A \mu_{\xi(x)}(C) d\mu$. D'après les propriétés de l'espérance conditionnelle, ceci est équivalent à l'égalité :

$$\forall C \in \mathcal{C}, \mu - pp x, \quad \mu_{\xi(x)}(C) = E(\mathbf{1}_C | \hat{\xi})(x)$$

La fonction $E(\mathbf{1}_C | \hat{\xi})$ n'est à priori définie que presque partout. Choisissons un représentant qui est constant sur les éléments de la partition. Les éléments de $\mathcal{C}$ étant en nombre dénombrable, on peut trouver un ensemble Ω , $\mu(\Omega^c) = 0$, pour lequel les quantités $E(\mathbf{1}_C | \hat{\xi})(x)$ sont définies pour tout $C \in \mathcal{C}$ et pour tout $x \in \Omega$. Quitte à restreindre à nouveau Ω , on peut supposer que la fonction $C \mapsto E(\mathbf{1}_C | \hat{\xi})(x)$ est finiment additive, ce qui donne la mesure $\mu_{\xi(x)}$ recherchée.

La relation intégrale est satisfaite pour $A \in \mathcal{C}$. Montrons que c'est encore vrai si $A \in \mathcal{T}$. Comme tout ouvert est union croissante d'élément de $\mathcal{C}$, elle est encore vraie pour les ouverts, et par passage au complémentaire, pour les fermés. Soit $A \in \mathcal{B}$, pour tout $\varepsilon > 0$, on peut trouver un fermé F et un ouvert U tels que $\mu(U \setminus F) < \varepsilon$ et $F \subset A \subset U$. On a donc :

$$\mu(A) - \varepsilon \leq \mu(F) = \int \mu_{\xi(x)}(F) d\mu(x) \leq \int \mu_{\xi(x)}(U) d\mu(x) = \mu(U) \leq \mu(A) + \varepsilon$$

Ceci démontre la relation recherchée pour $A \in \mathcal{B}$. Il reste à la vérifier pour $A \in \mathcal{T}$ μ -négligeable. Un tel ensemble est contenu dans un $B \in \mathcal{B}$ négligeable. La relation montre que B est $\mu_{\xi(x)}$ -négligeable pour μ -presque tout x . Il en va donc de même pour A . Ceci implique $\int \mu_{\xi(x)}(A) d\mu = 0$.

Il reste à montrer que $\mu_{\xi(x)}$ est supporté par $\xi(x)$. Soit B_n la suite d'ensembles intervenant dans la définition de ξ . On a, pour μ -presque tout x , et pour tout $n \in \mathbb{N}$, $\mu_{\xi(x)}(B_n) = E(\mathbf{1}_{B_n} | \hat{\xi})(x) = \mathbf{1}_{B_n}(x)$. Par conséquent, les B_n et B_n^c qui contiennent x sont de mesure totale pour $\mu_{\xi(x)}$, et il en va de même pour $\xi(x)$, qui est une intersection dénombrable de tels ensembles.

Remarque : Si $\mathcal{B} \subset \mathcal{T}$ est une tribu engendrée par un nombre dénombrable d'éléments, alors pour presque tout x , les éléments de $\mathcal{B}$ sont $\mu_{C(x)}$ -mesurables. Cependant, en général, les $\mu_{C(x)}$ ne sont pas définis sur toute la tribu $\mathcal{T}$.

Décomposition ergodique

Soit X un espace métrique et $x \in X$. La *composante ergodique* de x est définie par :

$$C(x) = \{y \in X \mid \forall f \in C_b(X), \frac{1}{n}S_n(f)(x) - \frac{1}{n}S_n(f)(y) \longrightarrow 0\}$$

Les ensembles $C(x)$ sont T -invariants et partitionnent l'espace X (cf exemple fig.3).

Proposition

Soit X un espace métrique, $T : X \rightarrow X$ une application borélienne et $x \in X$. Il existe au plus une mesure de probabilité borélienne invariante portée par $C(x)$, c'est-à-dire telle que $C(x)$ est mesurable et de complémentaire négligeable. Cette mesure est notée ν_x lorsqu'elle existe. Dans ce cas, ν_x est ergodique et on a :

$$\forall y \in C(x), \forall f \in C_b(X), \frac{1}{n}S_n(f)(y) \rightarrow \int f d\nu_x.$$

Preuve

Soit $f \in C_b(X)$ et $n_i \in \mathbf{N}$ une suite telle que $\frac{1}{n_i}S_{n_i}(f)(x)$ converge ; on note l sa limite. Comme $\frac{1}{n_i}S_{n_i}(f)(y)$ converge aussi vers l pour tout $y \in C(x)$, on a pour toute probabilité μ invariante, supportée par $C(x)$:

$$\int f d\mu = \int_{C(x)} f d\mu = \int_{C(x)} \frac{1}{n_i}S_{n_i}(f)(y) d\mu(y) = l \mu(C(x)) = l$$

Deux probabilités invariantes portées par $C(x)$ donnent même valeur à l'intégrale de f , elles sont donc égales. Si de plus une telle probabilité ν_x existe, alors $\int f d\nu_x$ est la seule valeur d'adhérence possible pour la suite $\frac{1}{n}S_n(f)(y)$, qui est donc convergente.

Théorème

Soit X un espace Borel standard, $T : X \rightarrow X$ une application borélienne et μ une probabilité invariante par T . Alors la partition $\{C(x)\}_{x \in X}$ est mesurable. L'ensemble des x pour lesquels $C(x)$ porte une probabilité invariante est de μ -mesure totale et on a, pour toute fonction f borélienne positive, et g μ -intégrable invariante :

$$\int_X f g d\mu = \int_X \left(\int_{C(x)} f d\nu_x \right) g(x) d\mu(x)$$

Preuve

Soit $\{f_k\}$ une famille dénombrable de fonctions continues, qui permet d'approcher de manière croissante les fonctions indicatrices d'ouverts. Le théorème ergodique de Birkhoff donne un ensemble $\tilde{\Omega}$ de mesure totale pour lequel $\frac{1}{n}S_n(f_k)(x)$ converge vers $Pf_k(x)$ pour tout $k \in \mathbf{N}$. On pose, pour ces x :

$$\begin{aligned} \tilde{C}(x) &= \{y \in \tilde{\Omega} \mid \forall k \in \mathbf{N}, \frac{1}{n}S_n(f_k)(x) - \frac{1}{n}S_n(f_k)(y) \longrightarrow 0\} \\ &= \{y \in \tilde{\Omega} \mid \forall k \in \mathbf{N}, Pf_k(x) = Pf_k(y)\} \end{aligned}$$

La partition $\tilde{C}$ est mesurable car ses éléments sont intersection d'ensembles de la forme $\{x \mid Pf_k(x) \in [r_1, r_2]\}$, $m \in \mathbf{N}$, $r_1, r_2 \in \mathbf{Q}$. Elle est T -invariante. Il en va donc de même pour les mesures $\mu_{\tilde{C}(x)}$, par unicité de la désintégration. Comme $\frac{1}{n}S_n(f_k)(y)$ converge vers $Pf_k(x)$ pour tout $y \in \tilde{C}(x)$, on a $\int f_k d\mu_{\tilde{C}(x)} = Pf_k(x)$.

Soit $\Omega = \{x \in \tilde{\Omega} \mid \forall k, \frac{1}{n}S_n(f_k)(x) \rightarrow \int f_k d\mu_{\tilde{C}(x)}\}$. La formule de désintégration montre que Ω^c est $\mu_{\tilde{C}(x)}$ -négligeable pour μ -pp x , et le théorème du portemanteau montre que $\tilde{C}(x) \cap \Omega$ est égal à $C(x) \cap \Omega$. La partition C est donc mesurable et pour μ -pp x , $C(x)$ porte une mesure de probabilité invariante : $\nu_x = \mu_{\tilde{C}(x)} = \mu_{C(x)}$.

Enfin, de la relation $\int f_k d\mu_{C(x)} = Pf_k(x)$, pour pp $x \in X$, on en déduit la relation intégrale recherchée pour les f_k , puis, par approximation, pour tous les f .

figure 1 : Désintégration

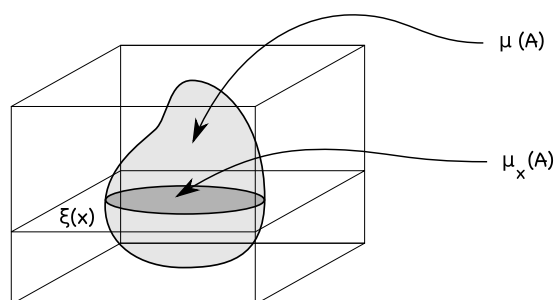


figure 2 : Désintégration de $f(x,y)dxdy$ sur $[0,1]^2$ le long des horizontales

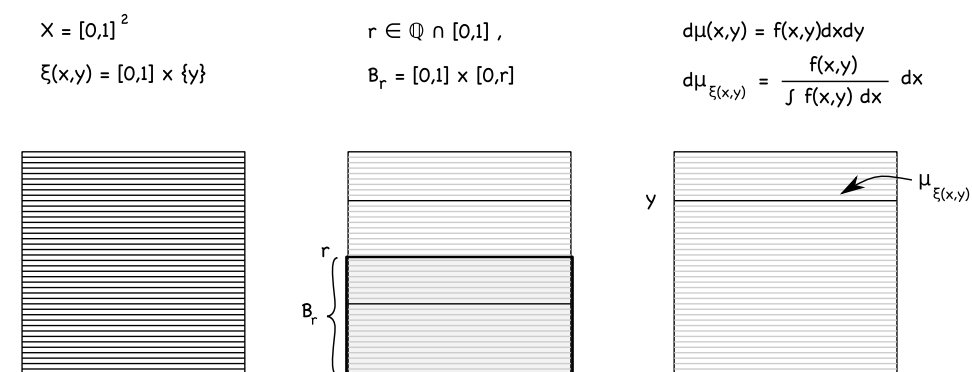
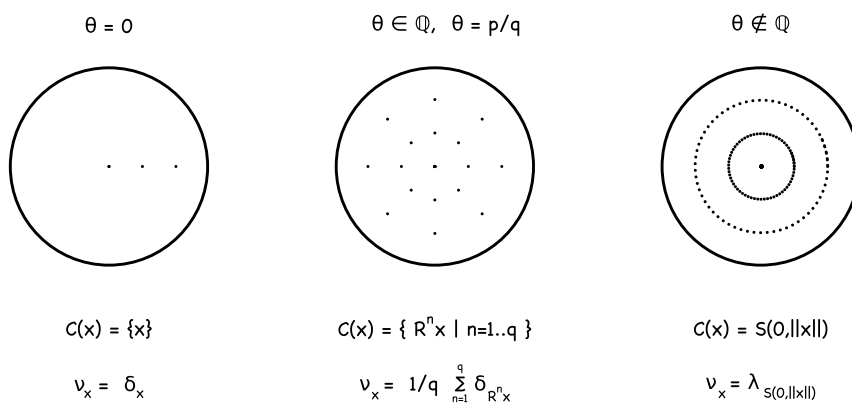


figure 3 : Décomposition ergodique d'une rotation R d'angle $2\pi\theta$ sur le disque



Exercices

exercice 1 :

Soit $(X, \mathcal{T}, \mu)$ un espace de Lebesgue et ξ une partition mesurable. Vérifiez que, pour tout f mesurable, et pour presque tout $x \in X$, $E(f \mid \hat{\xi})(x) = \int f \, d\mu_{\xi(x)}$. Exprimez les propriétés usuelles de l'espérance conditionnelle à l'aide des $\mu_{\xi(x)}$.

exercice 2 :

Soit $(X, \mathcal{T}, \mu)$ un espace de Lebesgue, $T : X \rightarrow X$ une application mesurable qui préserve μ et ξ une partition mesurable qui satisfait $\xi(T(x)) = T(\xi(x))$ pour presque tout x . Montrez que pour presque tout x , $T_*\mu_{\xi(x)} = \mu_{\xi(T(x))}$.

exercice 3 :

Donnez un exemple d'espace de Lebesgue $(X, \mathcal{T}, \mu)$ et de partition mesurable ξ tels que μ soit non-atomique, mais les $\mu_{\xi(x)}$ sont toutes atomiques.

exercice 4 :

Soit $(X, \mathcal{T}, \mu)$ un espace de Lebesgue et ξ une partition mesurable. Vérifiez que pour presque tout x , les espaces X et $\xi(x)$, munis de la mesure $\mu_{\xi(x)}$ et de la tribu des ensembles $\mu_{\xi(x)}$ -mesurables, sont des espaces de Lebesgue (avec atomes).

exercice 5 :

Soit X un espace Borel standard et d une métrique sur X qui engendre la topologie. Soit $BL(X)$ l'ensemble des fonctions lipschitziennes bornées. Montrez que les composantes ergodiques coïncident à un ensemble négligeable près avec la partition :

$$C'(x) = \{y \in X \mid \forall f \in BL(X), \frac{1}{n}S_n(f)(x) - \frac{1}{n}S_n(f)(y) \longrightarrow 0\}$$

Montrez que c'est encore vrai si on remplace $BL(X)$ par un sous-ensemble de $C_b(X)$ contenant une famille dénombrable $\mathcal{D}$ déterminante pour la topologie étroite.

exercice 6 :

Soit X un espace Borel standard, $T : X \rightarrow X$ une application borélienne et μ une mesure borélienne finie T -invariante. Montrez que pour μ -presque tout x , $C(x)$ est un espace Borel standard.

exercice 7 :

Soit X un espace Borel standard, $T : X \rightarrow X$ une application borélienne. Montrez que l'ensemble :

$$\left\{ x \in X \mid \frac{1}{n} \sum_{k=1}^n \delta_{T^k x} \text{ converge étroitement vers une mesure portée par } C(x) \right\}$$

est de mesure totale pour toute mesure borélienne finie invariante. Est-il de mesure totale pour toute mesure borélienne invariante ? Cet ensemble peut-il être vide ?

Remarque : les points de cet ensemble sont souvent appelés points génériques.

exercice 8 :

Soit X un espace métrique, $T : X \rightarrow X$ une application borélienne. On considère un ensemble borélien $\Omega \subset X$ tel que pour tout $x \in \Omega$, $C(x)$ porte une probabilité invariante ν_x . Montrez que pour toute mesure borélienne invariante finie μ et toute fonction f borélienne positive, $\int_{\Omega} f d\mu = \int_{\Omega} \left(\int f \, d\nu_x \right) d\mu(x)$.

exercice 9 :

Soit X un espace métrique séparable, $T : X \rightarrow X$ une application borélienne. Montrez que toute mesure finie invariante ergodique est supportée par une composante ergodique.

exercice 10 :

Soit X un espace métrique, $T : X \rightarrow X$ une application borélienne et μ une mesure finie T -invariante. Montrez que tout ensemble mesurable T -invariant coïncide, presque partout, avec un ensemble qui est union de composantes ergodiques.

Commentaires

La construction des mesures conditionnelles $\mu_{\xi(x)}$ repose sur le critère de Kolmogorov. Si X est localement compact, on peut aussi se baser sur le théorème de représentation de Riesz et donner une preuve qui ne fait pas appel à un modèle symbolique ; cf par exemple le livre de Furstenberg, “*recurrence in ergodic theory and combinatorial number theory*” [Fu81].

Lorsque μ est une mesure produit définie sur un espace produit, et ξ est la partition en horizontales ou verticales, le théorème de désintégration se réduit au théorème de Fubini. On retrouve donc la pathologie associée à ce genre de situation. Par exemple, la mesurabilité d’un ensemble A relativement à une probabilité μ ne peut se déduire de la mesurabilité de A relativement à toutes les désintégrations $\mu_{\xi(x)}$. Un contre-exemple fameux est donné par Sierpinski en 1920 : modulo l’axiome du continu, Sierpinski construit un sous-ensemble de $[0, 1] \times [0, 1]$ qui intersecte toute droite en au plus deux points, mais dont la mesure extérieure, au sens de Lebesgue, est égale à un.

Les ensembles $C(x)$ ne sont en général pas boréliens. Si X est métrique compact, ou si X est métrique précompact et que l’on demande à ce que la convergence dans la définition des $C(x)$ ait lieu pour les fonctions uniformément continues uniquement, alors les $C(x)$ sont des ensembles boréliens. La partition $C(x)$ est même borélienne, en restriction à l’ensemble $\Omega = \{x \mid \forall f, \frac{1}{n}S_n(f)(x) \text{ converge}\}$: on peut trouver des boréliens B_n tels que pour tout $x \in \Omega$, $C(x) \cap \Omega$ coïncide avec l’intersection de tous les B_n et B_n^c qui contiennent x . Ceci découle de l’existence d’un sous-ensemble dénombrable dense pour la topologie uniforme, dans l’ensemble des fonctions uniformément continues.

L’usage est de désigner sous le terme de “composantes ergodiques” la classe mod 0 de la partition $\{C(x)\}$ qui a été construite plus haut, ce qui fait dépendre les composantes de la mesure μ considérée. Dans les applications, il est parfois préférable de disposer d’une version borélienne de ces composantes, indépendante de la mesure. La version présentée plus haut est inspirée de l’argument de Hopf. On aurait pu prendre : $C(x) = \{y \in X \mid \forall f \in C_b, \exists n_i \text{ tel que } \frac{1}{n_i}S_{n_i}(f)(x) - \frac{1}{n_i}S_{n_i}(f)(y) \rightarrow 0\}$, ce qui autorise l’emploi du théorème ergodique L^2 plutôt que le théorème ergodique presque partout. Ou encore : $C(x) = \{y \in X \mid \forall f \in BL(X), \frac{1}{n}S_n(f)(x) - \frac{1}{n}S_n(f)(y) \rightarrow 0\}$, où $BL(X)$ est l’espace des fonctions bornées Lipschitziennes, relativement à une distance compatible avec la topologie de X . Ce choix fait dépendre les composantes de la distance choisie. Une autre possibilité est donnée par : $C(x) = \{y \in X \mid \rho(\frac{1}{n}\sum_{k=1}^n \delta_{T^k x}, \frac{1}{n}\sum_{k=1}^n \delta_{T^k y}) \rightarrow 0\}$, où ρ est la distance de Prokhorov. Dans ce cas, toutes les composantes ergodiques sont des espaces Borel standard.

Il existe plusieurs approches au théorème de décomposition ergodique. On peut utiliser un théorème abstrait de Choquet : *tout point dans un convexe compact, contenu dans un espace vectoriel topologique localement convexe, est barycentre d’une mesure portée par les points extrémaux du compact*. La décomposition ergodique est obtenue en appliquant ce résultat à $\mathcal{M}^1(X)$. Cette approche ne donne pas de description explicite des composantes ergodiques. Une autre approche, due à Krylov, Bogolyoubov, se base sur le concept de point générique. Ces points $x \in X$ sont ceux pour lesquels $\frac{1}{n}\sum_{k=1}^n f(T^k x)$ converge, pour tout $f \in C_b(X)$. Lorsque X est localement compact, le théorème de Riesz permet d’associer à chacun de ces points une mesure conditionnelle ν_x , ce qui permet de définir la composante ergodique comme l’ensemble des y tels que $\nu_x = \nu_y$.

Le théorème de décomposition ergodique est encore vrai pour les actions de groupes dénombrables sur des espaces Borel standard. Cette extension est due à V. S. Varadarajan (1963), une preuve se trouve dans le livre de E. Glasner, “*ergodic theory via joinings*” [Gl03]. Il existe aussi des théorèmes de décomposition ergodique pour des actions quasi-invariantes, ou dans le cadre de la mesure infinie. On peut consulter le livre de K. Schmidt, “*Cocycles on ergodic transformation groups*”, à ce sujet.

Convergence faible dans un espace de Hilbert

Un espace de Hilbert H est un espace vectoriel muni d'un produit scalaire tel que la norme associée au produit scalaire est complète. On prend comme notation $\langle , \rangle$ pour le produit scalaire et $\|f\| = \sqrt{\langle f, f \rangle}$ pour la norme.

On peut munir H de deux topologies :

- la *topologie forte* : une suite f_n d'éléments de H converge fortement vers $f \in H$ si $\|f_n - f\| \xrightarrow{n \rightarrow \infty} 0$
- la *topologie faible* : une suite f_n d'éléments de H converge faiblement vers $f \in H$ si $\langle f_n, g \rangle \xrightarrow{n \rightarrow \infty} \langle f, g \rangle$ pour tout $g \in H$.

Théorème

Une suite fortement convergente est faiblement convergente.

C'est une conséquence de l'inégalité suivante :

Inégalité de Cauchy-Schwarz

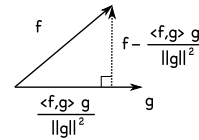
$$\forall f, g \in H, \quad |\langle f, g \rangle| \leq \|f\| \|g\|$$

preuve

C'est une conséquence du théorème de Pythagore :

$$\|f\|^2 = \|\langle f, \frac{g}{\|g\|} \rangle \frac{g}{\|g\|}\|^2 + \|f - \langle f, \frac{g}{\|g\|} \rangle \frac{g}{\|g\|}\|^2$$

Remarquons qu'on a égalité ssi f est proportionnel à g .



Théorème

Une suite faiblement convergente est bornée.

Preuve

Il suffit de montrer qu'il existe $g \in H$ et $\varepsilon > 0$ tel que :

$$\sup \{ \langle f_n, h \rangle \mid n \in \mathbf{N}, h \in B(g, \varepsilon) \} < \infty$$

car alors l'égalité suivante montre que $\|f_n\|$ est majorée indépendamment de n :

$$\|f_n\| = \frac{1}{\varepsilon} (\langle f_n, g + \varepsilon \frac{f_n}{\|f_n\|} \rangle - \langle f_n, g \rangle)$$

Supposons qu'on ne puisse pas trouver g et ε comme plus haut. On construit par récurrence une suite croissante d'entiers n_k et une suite de boules fermées B_k décroissante pour l'inclusion, dont le diamètre tend vers 0, et telles que pour tout $h \in B_k$, $\langle f_{n_k}, h \rangle \geq 2^k$.

Soit g_∞ l'unique point appartenant à tous les B_k . La majoration $\langle f_{n_k}, g_\infty \rangle \geq 2^k$ contredit la convergence faible de la suite f_n .

La convergence faible correspond à la convergence "coordonnées par coordonnées" :

Théorème

soit $\mathcal{D}$ une famille d'éléments de H qui engendre un sous-espace vectoriel dense de H ; pour montrer que f_n converge faiblement vers f , il suffit de vérifier la convergence $\langle f_n, g \rangle \xrightarrow{n \rightarrow \infty} \langle f, g \rangle$ pour $g \in \mathcal{D}$.

Preuve

Soit $\varepsilon > 0$, $g \in H$, $g_k \in \text{Vect}(\mathcal{D})$ tel que $\|g_k - g\| \xrightarrow{n \rightarrow \infty} 0$. On a la majoration :

$$|\langle f_n - f, g \rangle| \leq |\langle f_n - f, g_k \rangle| + \|f_n - f\| \|g_k - g\|$$

Comme la suite $\|f_n - f\|$ est bornée, on peut trouver un $k \in \mathbf{N}$ tel que le dernier terme de cette inégalité soit inférieur à ε pour tout n . Pour cette valeur de k , la suite $\langle f_n - f, g_k \rangle$ tend vers 0 ; on peut donc majorer $|\langle f_n - f, g \rangle|$ par 2ε dès que n est assez grand.

Voici deux autres résultats relatifs à la topologie faible.

Théorème de Banach-Alaoglu

Soit H un espace de Hilbert, $\{f_n\}_{n \in \mathbb{N}}$ une suite bornée d'éléments de H . Alors on peut trouver $f \in H$ et une sous-suite $f_{n_1}, \dots, f_{n_k} \dots$ telle que :

$$f_{n_k} \rightharpoonup f \quad \text{quand} \quad k \rightarrow +\infty.$$

Preuve

Soit F le sous-espace vectoriel fermé engendré par les f_n , et $\{v_m\}$ une base hilbertienne de F . Comme $\langle v_1, f_n \rangle$ est une suite bornée, on peut trouver une sous-suite $f_{n'_i}$ telle que $\langle v_1, f_{n'_i} \rangle$ soit convergente. De cette sous-suite, on peut extraire une sous-suite telle que $\langle v_2, f_{n''_i} \rangle$ converge, et ainsi de suite. En prenant le k^{ieme} terme de la k^{ieme} sous-suite, on construit une suite f_{n_k} telle que, pour tout $m \in \mathbb{N}$, la suite $\langle v_m, f_{n_k} \rangle$ est convergente. Notons c_m sa limite. Soit $N \in \mathbb{N}$.

$$\|f_{n_k}\|^2 \geq \sum_{m=0}^N |\langle v_m, f_{n_k} \rangle|^2 \xrightarrow{k \rightarrow +\infty} \sum_{m=0}^N |c_m|^2$$

La suite $\|f_{n_k}\|$ étant bornée, cette dernière quantité est convergente, et on obtient un élément de H bien défini en posant $f = \sum c_i v_i$. Cet élément vérifie : $\forall m \in \mathbb{N}$, $\langle v_m, f_{n_k} \rangle \xrightarrow{k \rightarrow +\infty} \langle v_m, f \rangle$. On a donc $\langle h, f_{n_k} \rangle \rightarrow \langle h, f \rangle$, si $h \in F$. Si $h \in F^\perp$, on a $\langle h, f_{n_k} \rangle = 0$ et $\langle h, f \rangle = 0$. Ceci démontre le théorème.

Théorème de Banach-Saks

Soit H un espace de Hilbert, $\{f_n\}_{n \in \mathbb{N}}$ une suite d'éléments de H telle que $f_n \rightharpoonup f$ faiblement. Alors il existe une sous-suite $f_{n_1}, \dots, f_{n_k} \dots$ telle que :

$$\frac{1}{m} \sum_{k=1}^m f_{n_k} \xrightarrow{m \rightarrow +\infty} f \quad \text{fortement.}$$

Preuve

On peut supposer $f = 0$. Construisons une sous-suite f_{n_k} satisfaisant :

$$\sum_{i < j} |\langle f_{n_i}, f_{n_j} \rangle| < +\infty$$

Supposons construits $f_{n_1} \dots f_{n_k}$ telles que si $i < j$, $|\langle f_{n_i}, f_{n_j} \rangle| < \frac{1}{2^j}$. Pour tout $i \leq k$, on sait que $\langle f_{n_i}, f_n \rangle \xrightarrow{n \rightarrow +\infty} 0$. On peut donc trouver n_{k+1} tel que $\forall i \leq k$, $|\langle f_{n_i}, f_{n_{k+1}} \rangle| < \frac{1}{2^{k+1}}$. Au final, $\sum_{i < j} |\langle f_{n_i}, f_{n_j} \rangle| < \sum \frac{1}{2^j} \leq 2$.

$$\left\| \frac{1}{m} \sum_{k=1}^m f_{n_k} \right\|^2 = \frac{1}{m^2} \sum_{k=1}^m \|f_{n_k}\|^2 + \frac{2}{m^2} \sum_{\substack{k=1 \\ i < j}}^m \langle f_{n_i}, f_{n_j} \rangle \leq \frac{1}{m} \sup \|f_{n_k}\|^2 + \frac{4}{m^2}$$

Cette quantité tend bien vers 0.

Remarques

- On en déduit que tout ensemble convexe fortement fermé est faiblement fermé.
- Soit $(X, \mathcal{T}, \mu)$ un espace mesuré. Toute suite de L^2 qui converge en norme L^2 admet une sous-suite qui converge presque partout. Si H est de la forme $L^2(X)$, on peut donc trouver m_i tel que : $\frac{1}{m_i} \sum_{k=1}^{m_i} f_{n_k} \xrightarrow{i \rightarrow +\infty} f$ presque partout.

Calcul différentiel

Considérons un flot défini sur un ouvert de $\mathbf{R}^n$ ou sur une variété M . A quelle condition ce flot préserve-t-il la mesure de Lebesgue ? Le théorème suivant montre qu'il suffit que la divergence du champ de vecteur $X(x) = \frac{d}{dt}\varphi_t(x)|_{t=0}$ soit nulle.

Théorème

Soit φ_t un flot C^2 défini sur un ouvert U de $\mathbf{R}^d$, X le champ de vecteur associé et $f : U \rightarrow \mathbf{R}$ une application intégrable, nulle hors d'un compact de U . Alors :

$$\frac{d}{dt} \int f(\varphi_t(x)) \, dx \Big|_{t=t_0} = \int f(\varphi_{t_0}(x)) \operatorname{div} X(x) \, dx$$

Preuve

En utilisant un raisonnement de densité, on peut se restreindre au cas où f est C^1 à support compact, localisé dans un rectangle R . Quitte à remplacer f par $f \circ \varphi_{t_0}$, on peut supposer de plus $t_0 = 0$. Dérivons sous le signe somme :

$$\frac{d}{dt} \int_R f(\varphi_t(x)) \, dx = \int_R \langle \nabla f(x), X(x) \rangle \, dx$$

Puis effectuons une intégration par partie :

$$\begin{aligned} \int_R \frac{\partial f}{\partial x_i} X_i \, dx &= \int_R \frac{\partial(f X_i)}{\partial x_i} \, dx - \int_R f \frac{\partial X_i}{\partial x_i} \, dx \\ \int \cdots \int \left(\int \frac{\partial(f X_i)}{\partial x_i} \, dx_i \right) dx_1 \cdots dx_n &= \int \cdots \int [f X_i]_{a_i}^{b_i} dx_1 \cdots dx_n = 0 \end{aligned}$$

en tenant compte du fait que la fonction f s'annule sur le bord de R .

Corollaire (Liouville)

Soit φ_t un flot C^1 un flot défini sur une variété différentielle, et X le champ de vecteur associé. Si la divergence de X est identiquement nulle, alors le flot préserve la mesure de Lebesgue.

Montrer qu'une transformation C^1 sur une variété compacte est bijective n'est pas toujours évident. Le résultat suivant permet de simplifier les calculs. On peut par exemple l'utiliser pour montrer l'invérissibilité de l'application dérivée d'Anosov étudiée au chapitre "Un attracteur étrange".

Théorème

Soit M une variété différentielle compacte connexe et $f : M \rightarrow M$ une application C^1 dont le jacobien $|\det d_x f|$ ne s'annule pas sur M . Alors le cardinal des fibres $\operatorname{Card}(f^{-1}(y))$ est fini, indépendant de y .

Preuve

Soit $y \in M$. Appliquons le théorème d'inversion locale : pour tout $x \in f^{-1}(y)$, on peut trouver un ouvert U_x tel que f soit un difféomorphisme de U_x sur son image. En particulier $f^{-1}(y)$ intersecte U_x en x seulement, ce qui montre que $f^{-1}(y)$ est un ensemble isolé. Comme il est compact, il est de cardinal fini.

Fixons $y_0 \in M$ et choisissons les ouverts U_x , $x \in f^{-1}(y_0)$, de façon à ce qu'ils aient tous la même image V et qu'ils soient disjoints deux-à-deux. L'application $y \mapsto \operatorname{Card}(f^{-1}(y))$ est alors constante sur l'ouvert $V \setminus f\left(\bigcup_x U_x\right)^c$.

Ceci montre que l'application $y \mapsto \operatorname{Card}(f^{-1}(y))$ est localement constante, donc continue. Comme M est connexe, elle est constante.

Remarque : On a démontré un peu plus ; le cardinal des fibres est localement constant sur l'ensemble des valeurs y qui satisfont $\det d_x f \neq 0$ pour tout $x \in f^{-1}(y)$. Ces valeurs sont dites régulières.

Topologie et mesure

Séparabilité

Définition

Un espace métrique X est séparable s'il contient une partie dénombrable dense.

Un espace topologique est à base dénombrable si on peut trouver un ensemble dénombrable d'ouverts tel que tout ouvert puisse s'écrire comme union d'éléments de cet ensemble.

Remarquons que dans ce cas tout ouvert peut s'écrire comme union dénombrable d'éléments de la base ; en l'occurrence comme l'union de tous les éléments de la base qu'il contient.

Théorème

Un espace métrique est séparable si et seulement s'il est à base dénombrable.

Ceci implique que tout sous-ensemble d'un espace métrique séparable est séparable. En effet, ce sous-ensemble admet une base dénombrable d'ouverts, qui est obtenue en restreignant la base de X au sous-ensemble.

Preuve

Si X est à base dénombrable, il suffit de prendre un point dans chaque élément de la base. L'ensemble obtenu rencontre tous les ouverts, il est donc dense.

Supposons X séparable ; soit $\mathcal{D}$ une partie dénombrable dense. Montrons que l'ensemble $\{B(y, r) \mid y \in \mathcal{D}, r \in \mathbf{Q}\}$ est une base d'ouverts. Soit U un ouvert et $x \in U$. On peut trouver un point $y \in \mathcal{D}$ tel que $d(y, x) < d(y, U^c)$. Soit $r \in \mathbf{Q}$ tel que $d(y, x) < r < d(y, U^c)$. La relation $x \in B(y, r) \subset U$ montre que U est bien union de boules ouvertes, à rayon rationnel, centrées en $\mathcal{D}$.

Support d'une mesure

Définition

Soit X un espace métrique, μ une mesure borélienne. Un point $x \in X$ appartient au support de la mesure μ si tout voisinage de x est de mesure strictement positive. Le support est noté par $\text{supp } \mu$.

Théorème

- Le support est un ensemble fermé.
- Si X est séparable, alors $\mu((\text{supp } \mu)^c) = 0$.
- Soit $A \subset X$ tel que $\mu(A^c) = 0$. Alors $\text{supp } \mu \subset \bar{A}$.
- Si la mesure μ est finie, son support est séparable : on peut trouver un sous-ensemble dénombrable du support, qui est dense dans le support.

Preuve

On a l'égalité :

$$(\text{supp } \mu)^c = \bigcup_{\substack{U \text{ ouvert} \\ \mu(U)=0}} U$$

L'union peut être restreinte à une base d'ouverts ; ceci démontre les deux premiers points. Soit $x \in \text{supp } \mu$ et U un ouvert contenant x . Comme $\mu(U) > 0$ et $\mu(A^c) = 0$, $\mu(U \cap A) > 0$. L'ouvert U rencontre A , ce qui montre que $x \in \bar{A}$; i.e. $\text{supp } \mu \subset \bar{A}$.

Pour tout $k, n \in \mathbf{N}^*$, on considère les ensembles A qui satisfont les deux conditions :

- $\forall x, y \in A, x \neq y$ implique $d(x, y) > 2/k$;
- $\forall x \in A, \mu(B(x, 1/k)) > 1/n$.

Ces ensembles ont un cardinal fini, majoré par $n\mu(X)$. Pour chaque k, n , on choisit un tel ensemble $A_{k,n}$ de cardinal maximal. L'union de tous ces ensembles, pour $k, n \in \mathbf{N}^*$, forme une partie dénombrable dense de $\text{supp } \mu$.

En effet, si $x \in \text{supp } \mu$ et $k > 0$, on peut trouver $n \in \mathbf{N}^*$ tel que $\mu(B(x, 1/k)) > 1/n$. Par maximalité de $A_{k,n}$, $d(x, y) < \frac{2}{k}$ pour un certain $y \in A_{k,n}$, ou alors $x \in A_{k,n}$.

Densité dans les espaces L^p

Soit X un espace topologique, on note $\mathcal{B}(X)$ sa tribu des boréliens. Soit μ une mesure borélienne définie sur X . La complétion de $\mathcal{B}(X)$ relativement à μ sera notée $\overline{\mathcal{B}(X)}$, ses éléments sont les ensembles μ -mesurables.

$$\overline{\mathcal{B}(X)} = \{A \subset X \mid \exists B', B'' \in \mathcal{B}(X) \text{ tels que } B' \subset B \subset B'' \text{ et } \mu(B'' \setminus B') = 0\}$$

Théorème de régularité extérieure

Soit X un espace métrique, μ une mesure borélienne finie et $A \subset X$ un ensemble μ -mesurable. Alors pour tout $\varepsilon > 0$, on peut trouver un ouvert U contenant A tel que $\mu(U \setminus A) < \varepsilon$.

On dit que la mesure est *extérieurement régulière*. Ceci implique que tout ensemble μ -mesurable peut s'écrire sous la forme $(\cap U_i) \setminus N$, avec U_i suite d'ouverts de X et N ensemble négligeable.

Preuve

Posons $\mathcal{T} = \{A \text{ mesurable} \mid \forall \varepsilon > 0, \exists U \text{ ouvert tel que } A \subset U \text{ et } \mu(U \setminus A) < \varepsilon \\ \forall \varepsilon > 0, \exists V \text{ ouvert tel que } A^c \subset V \text{ et } \mu(V \setminus A^c) < \varepsilon\}$

Les ouverts sont dans $\mathcal{T}$: le complémentaire de tout ouvert U peut s'écrire comme intersection des ouverts $V_n = \{x \in X \mid d(x, U^c) < \varepsilon\}$, la mesure des V_n tend donc vers celle de U^c . Les ensembles négligeables sont aussi dans $\mathcal{T}$. Il suffit maintenant de montrer que $\mathcal{T}$ est une tribu. L'invariance par passage au complémentaire est évidente. Soit A_i des ensembles mesurables dans $\mathcal{T}$ et U_i des ouverts tels que $A_i \subset U_i$ et $\mu(U_i \setminus A_i) < \varepsilon/2^i$. L'union des U_i approche bien l'union des A_i :

$$\cup A_i \subset \cup U_i, \quad \mu(\cup U_i \setminus \cup A_i) \leq \sum \mu(U_i \setminus A_i) < \varepsilon$$

Pour les complémentaires, on peut trouver n tel que $\mu(\bigcap_1^n A_i \setminus \bigcap_1^\infty A_i) < \varepsilon/2$ et on utilise l'inclusion :

$$\bigcap_1^n V_i \setminus \bigcap_1^n A_i^c \subset \bigcup_1^n (V_i \setminus A_i^c)$$

Théorème de densité

Soit X un espace métrique, μ une mesure borélienne finie et $1 \leq p < \infty$. Alors toute fonction $f \in L^p(X, \mu)$ peut être approchée, en norme L^p , par une suite de fonctions lipschitziennes bornées appartenant à L^p .

Preuve

Comme les fonctions qui sont combinaisons linéaires de fonctions indicatrices sont denses dans les L^p , on peut se restreindre au cas où f est une fonction indicatrice, $f = \mathbf{1}_A$. Soit U un ouvert contenant A tel que $\mu(U \setminus A) < \varepsilon$.

$$\|\mathbf{1}_U - \mathbf{1}_A\|_p = \mu(U \setminus A)^{1/p} < \varepsilon^{1/p}$$

Il suffit maintenant d'approcher $\mathbf{1}_U$ par une suite de fonctions lipschitziennes bornées. Posons $f_k = \min(1, kd(x, U^c))$. Ces fonctions sont lipschitziennes :

$$\forall x, y \in X, \quad |f_k(x) - f_k(y)| \leq k d(x, y)$$

La suite f_k converge vers $\mathbf{1}_U$ simplement, et les fonctions $|\mathbf{1}_U - f_k|$ sont majorées uniformément par 1. Ceci assure la convergence des f_k vers $\mathbf{1}_U$ en norme L^p .

Remarque

Ce théorème est encore vrai si X est métrique séparable et μ est localement finie.

Régularité intérieure

On s'intéresse maintenant à une propriété plus forte que la régularité extérieure.

Définition

Une mesure borélienne μ définie sur X est dite *intérieurement régulière* si pour tout borélien $A \subset X$ de mesure finie, et tout $\varepsilon > 0$, on peut trouver un compact $K \subset A$ tel que $\mu(A \setminus K) < \varepsilon$.

Théorème d'Oxtoby-Ulam

Toute mesure borélienne finie, définie sur un espace métrique séparable complet, est intérieurement régulière.

Preuve

Soit $\{x_i\}$ une suite dense dans X et $n \in \mathbf{N}^*$. La famille de boules fermées $\overline{B}(x_i, 1/n)$, $i \in \mathbf{N}$, recouvre X si bien qu'on peut trouver $N \in \mathbf{N}$, dépendant de n , tel que

$$\mu\left(X \setminus \bigcup_{i=0}^N \overline{B}(x_i, 1/n)\right) < \varepsilon/2^n.$$

Posons

$$K' = \bigcap_{n \in \mathbf{N}^*} \bigcup_{i=0}^N \overline{B}(x_i, 1/n)$$

Cet ensemble est fermé et précompact, il est donc compact. On a de plus l'inégalité : $\mu(K^c) < \varepsilon$. Soit maintenant $A \subset X$ un borélien. Par régularité extérieure, on peut trouver $F \subset A$ fermé tel que $\mu(A \setminus F) < \varepsilon$. Posons $K = K' \cap F \subset A$. Cet ensemble est compact et $\mu(A \setminus K) < 2\varepsilon$. Le théorème est démontré.

Théorème de Lusin

Soit X un espace topologique séparé, μ une mesure finie intérieurement régulière, Y un espace métrique séparable. Soit $f : X \rightarrow Y$ une fonction mesurable (i.e. l'image réciproque de tout borélien est μ -mesurable). Alors pour tout $A \subset X$ μ -mesurable, pour tout $\varepsilon > 0$, on peut trouver un compact $K \subset A$ tel que $\mu(A \setminus K) < \varepsilon$ et $f|_K$ est continue.

Preuve

Fixons $n \in \mathbf{N}^*$. Soit E_i une famille dénombrable disjointe de boréliens de diamètre plus petit que $2/n$ qui recouvre Y . On peut construire les E_i par récurrence à partir d'une suite $\{y_i\}_{i \in \mathbf{N}^*}$ dense dans Y en posant : $E_i = B(y_i, 1/n) \setminus \bigcup_{j < i} E_j$.

Par régularité, pour chaque i , on peut trouver un borélien A'_i et un compact K_i tel que :

$$K_i \subset A'_i \subset A \cap f^{-1}(E_i), \quad \mu(A \cap f^{-1}(E_i) \setminus K_i) < \varepsilon/2^{i+n}.$$

On a alors $\mu(A \setminus \bigcup K_i) < \varepsilon/2^n$ si bien qu'il existe $N \in \mathbf{N}$, dépendant de n , tel que :

$$\mu\left(A \setminus \bigcup_{i=0}^N K_i\right) < \varepsilon/2^n.$$

Définissons une fonction f_n sur $K'_n = \bigcup_{i=0}^N K_i$ en posant $f_n(x) = y_i$ pour tout $x \in K_i$. Les compacts K_i sont disjoints, la fonction f_n est donc continue. Elle vérifie de plus $d(f_n(x), f(x)) < 1/n$ sur K'_n . La suite f_n converge vers f uniformément sur $\bigcap_n K'_n$, ce qui montre que f est continue sur ce compact. Ceci termine la preuve.

Exercices

exercice 1 :

Donner un exemple d'espace métrique localement compact qui n'est pas séparable.

exercice 2 :

Soit X un espace métrique séparable. Montrer qu'il existe une suite de fonctions continues $f_n : X \rightarrow \mathbf{R}$ qui sépare les points : pour tous $x, y \in X$ distincts, il existe n tel que $f_n(x) \neq f_n(y)$. En déduire qu'on peut trouver une injection continue de X dans $\mathbf{R}^{\mathbf{N}}$.

exercice 3 :

Donner une suite explicite de compacts $K_n \subset [0, 1] \setminus \mathbf{Q}$ telle que $\lambda(K_n)$ tend vers 1 quand n tend vers $+\infty$.

exercice 4 :

Soit X un espace topologique séparé, muni d'une mesure borélienne finie. On dit que X est *presque σ -compact*, s'il existe une suite de compacts K_i telles que $\mu(X \setminus \bigcup_{i \in \mathbf{N}} K_i) = 0$. Montrer l'équivalence entre les propriétés suivantes :

- l'espace X est presque σ -compact,
- la mesure μ est intérieurement régulière,
- tout borélien de X est presque σ -compact.

Le terme utilisé pour cette propriété dans la littérature de langue anglaise est "tight".

exercice 5 :

Montrez que tout espace métrique muni d'une probabilité borélienne intérieurement régulière est *presque séparable* : il admet un sous-ensemble séparable de mesure totale.

exercice 6 :

Montrez que $\mathbf{R}^{\mathbf{N}}$ n'est pas σ -compact, mais qu'il est presque σ -compact, relativement à toute mesure finie borélienne.

exercice 7 :

Montrez que le théorème de Oxtoby-Ulam est encore vrai si au lieu de supposer la mesure finie, on la suppose σ -finie : il existe une suite de boréliens B_i de mesure finie tels que $X = \bigcup_{i \in \mathbf{N}} B_i$.

exercice 8 :

Une tribu $\mathcal{T}$ est dite complète si on a la propriété suivante : pour tout $B \in \mathcal{T}$ tel que $\mu(B) = 0$ et pour tout $A \subset B$, A appartient à $\mathcal{T}$. Soit X un espace topologique. Montrez que $\overline{\mathcal{B}(X)}$ est complète.

exercice 9 :

Soit $(X, \mathcal{T}, \mu)$ un espace mesuré avec $\mathcal{T}$ complète, Y un espace topologique, et $f : X \rightarrow Y$ une fonction telle que l'image réciproque de tout borélien est dans $\mathcal{T}$. Montrez que l'image réciproque de tout ensemble $f_*\mu$ -mesurable est dans $\mathcal{T}$.

Soient X, Y deux espaces topologiques, μ, ν deux mesures boréliennes définies sur X et Y , et $\varphi : X \rightarrow Y$ une application borélienne telle que $\varphi_*\mu = \nu$. Montrez que φ est mesurable : l'image réciproque d'un ensemble ν -mesurable est μ -mesurable.

exercice 10 :

Soit $\varphi : [0, 1] \rightarrow [0, 1]$ une application mesurable au sens suivant : l'image réciproque d'un ensemble borélien est Lebesgue-mesurable. Montrez qu'on peut trouver un sous-ensemble borélien $X_0 \subset [0, 1]$ tel que $\varphi(X_0)$ est borélien et $\varphi : X_0 \rightarrow \varphi(X_0)$ est borélienne. *Indic: utiliser le théorème de Lusin.*

Partitions mesurables et σ -algèbres

Définition d'une partition mesurable

Soit $(X, \mathcal{T}, \mu)$ un espace probabilisé. Une partition ξ de X est la donnée d'un ensemble de parties de X , disjointes deux à deux, recouvrant X . L'élément de la partition qui contient le point $x \in X$ est noté $\xi(x)$.

La partition est dite mesurable s'il existe une famille dénombrable d'ensembles mesurables $\{B_n\}$, chacun des B_n étant union d'éléments de la partition, et tel que : $\forall C_1, C_2 \in \xi$ tels que $C_1 \neq C_2$, $\exists n \in \mathbf{N}$ satisfaisant :

$$C_1 \subset B_n \text{ et } C_2 \subset B_n^c \quad \text{ou} \quad C_1 \subset B_n^c \text{ et } C_2 \subset B_n.$$

On convient d'identifier deux partitions ξ et η s'il existe un ensemble mesurable $\Omega \subset X$, de mesure pleine, tel que $\xi(x) \cap \Omega = \eta(x) \cap \Omega$, $\forall x \in X$.

Les éléments d'une partition mesurable sont parfois appelés les *atomes* de la partition. Ce sont des ensembles mesurables ; il suffit de remarquer que les ensembles B_n qui interviennent dans la définition de la partition déterminent complètement cette partition : $\xi(x) = \bigcap_{B_n \ni x} B_n \bigcap_{B_n^c \ni x} B_n^c$.

On dira qu'une sous σ -algèbre $\mathcal{A}$ de $\mathcal{T}$ est *complète* si elle contient les ensembles μ -mesurables négligeables relativement à la mesure μ . Il s'agit d'une complétion relative. Toute sous σ -algèbre $\mathcal{A} \subset \mathcal{T}$ admet une complétion unique $\bar{\mathcal{A}}$: c'est la σ -algèbre engendrée par $\mathcal{A}$ et les ensembles μ -négligeables. De manière équivalente, c'est l'ensemble des parties $\tilde{A} \in \mathcal{T}$ pour lesquelles il existe un ensemble $\Omega \in \mathcal{T}$ de mesure totale, et un ensemble $A \in \mathcal{A}$ satisfaisant : $\tilde{A} \cap \Omega = A \cap \Omega$.

On va maintenant associer à chaque partition une σ -algèbre, et réciproquement.

σ -algèbre associée à une partition

À la partition ξ , on associe la complétion de la σ -algèbre engendrée par les ensembles mesurables saturés par ξ :

$$\hat{\xi} = \overline{\{A \in \mathcal{T} \mid A = \bigcup_{x \in A} \xi(x)\}}$$

Remarquons qu'une fonction mesurable est $\hat{\xi}$ -mesurable si et seulement si elle coïncide presque partout avec une fonction constante sur les atomes de la partition ; il suffit d'approcher la fonction par des combinaisons linéaires d'indicatrices pour s'en convaincre.

Le lemme suivant va permettre de mettre en bijection partitions et σ -algèbres.

Lemme

Soit $(X, \mathcal{T}, \mu)$ un espace de Lebesgue, ξ une partition mesurable de X et B_n les ensembles qui interviennent dans sa définition. Alors la σ -algèbre $\hat{\xi}$ est engendrée par les B_n et les ensembles négligeables.

Preuve :

Soit $\varphi : X \rightarrow \{0, 1\}^{\mathbf{N}}$ la fonction définie par $\varphi(x) = \{1_{B_n}(x)\}$. Remarquons que l'image réciproque d'un borélien par φ est mesurable. De plus, l'image réciproque de la σ -algèbre des ensembles $\varphi_*\mu$ -mesurables est contenue dans la σ -algèbre engendrée par les B_n et les ensembles négligeables. Il s'agit donc de montrer qu'elle contient la σ -algèbre $\{A \in \mathcal{T} \mid A = \bigcup_{x \in A} \xi(x)\}$.

Considérons un ensemble $A \in \mathcal{T}$; on a l'égalité $\varphi^{-1}\varphi(A) = \bigcup_{x \in A} \xi(x)$. Par conséquent, si $A = \bigcup_{x \in A} \xi(x)$, on obtient : $\varphi^{-1}\varphi(A) = A$. Le lemme de mesurabilité montre que $\varphi(A)$ est $\varphi_*\mu$ -mesurable, ce qui termine la démonstration.

Partition associée à une σ -algèbre

On dira qu'une σ -algèbre complète $\mathcal{A} \subset \mathcal{T}$ est *séparable* si c'est la complétion d'une σ -algèbre engendrée par une famille dénombrable de parties.

Lemme

Soit $(X, \mathcal{T}, \mu)$ un espace de Lebesgue. Toute σ -algèbre complète $\mathcal{A} \subset \mathcal{T}$ est séparable.

Preuve :

L'espace $L^1(X, \mathcal{T}, \mu)$ est séparable. Toute partie d'un espace métrique séparable étant séparable, l'ensemble $\{\mathbf{1}_A \mid A \in \mathcal{A}\}$ est séparable pour la norme L^1 . Soit $\{\mathbf{1}_{A_n}\}$ une partie dénombrable dense ; montrons que les A_n engendrent $\mathcal{A}$. Pour tout $A \in \mathcal{A}$, on peut trouver une suite n_k telle que $\mathbf{1}_{A_{n_k}}$ converge presque partout vers $\mathbf{1}_A$. La différence symétrique de A et de $\overline{\lim A_{n_k}}$ est donc de mesure nulle. l'ensemble A est dans la complétion de la σ -algèbre engendrée par les A_n .

Soit $\mathcal{A}$ une sous σ -algèbre complète de $\mathcal{T}$, et $\{B_n\}$ un ensemble dénombrable de parties tels que $\mathcal{A}$ soit engendré par les B_n et les ensembles négligeables. On associe à $\mathcal{A}$ la partition $\xi_{\mathcal{A}}$ dont les atomes sont donnés par la formule :

$$\xi_{\mathcal{A}}(x) = \bigcap_{B_n \ni x} B_n \bigcap_{B_n^c \ni x} B_n^c$$

Cette partition est mesurable ; les B_n séparent bien les éléments de la partition.

Lemme

La définition de la partition $\xi_{\mathcal{A}}$ ne dépend pas du choix des B_n .

Preuve :

Soit B_n un ensemble de parties et $\langle B_n \rangle$ la σ -algèbre engendrée. On a l'égalité :

$$\bigcap_{B_n \ni x} B_n \bigcap_{B_n^c \ni x} B_n^c = \bigcap_{\substack{A \ni x \\ A \in \langle B_n \rangle}} A$$

Pour voir cela, on remarque que $\langle B_n \rangle_{x,y} = \{A \in \langle B_n \rangle \mid x \in A \leftrightarrow y \in A\}$ est une σ -algèbre qui contient les ensembles B_n , si $x \in B_n \leftrightarrow y \in B_n, \forall n$.

Soit B_n et B'_n deux familles dénombrables dont $\mathcal{A}$ est la complétion. Pour chaque n , il existe des ensembles $A_n^1, A_n^2 \in \langle B_n \rangle$ tel que $A_n^1 \subset B'_n \subset A_n^2$ et $\mu(A_n^2 - A_n^1) = 0$. De la même façon, il existe des ensembles $A_n'^1, A_n'^2 \in \langle B'_n \rangle$ tel que $A_n'^1 \subset B_n \subset A_n'^2$ et $\mu(A_n'^2 - A_n'^1) = 0$. Les partitions associées aux B_n et aux B'_n coïncident sur $\Omega = (\cup A_n^2 - A_n^1)^c \cap (\cup A_n'^2 - A_n'^1)^c$.

Facteurs et partitions

Définition d'un facteur

Soit $(X, \mathcal{T}, \mu)$ un espace de Lebesgue. Un facteur de $(X, \mathcal{T}, \mu)$ est la donnée d'un espace de Lebesgue $(Y, \mathcal{S}, \nu)$ et d'une application mesurable $\varphi : X \rightarrow Y$ satisfaisant : $\varphi_* \mu = \nu$.

Le lemme de mesurabilité montre que l'application φ est presque surjective. On dira que φ est la projection associée au facteur.

A tout facteur de $(X, \mathcal{T}, \mu)$, on peut associer une partition de X par la formule $\xi = \{\varphi^{-1}(\{y\}) \mid y \in Y\}$. Cette partition est mesurable ; il suffit de prendre pour ensembles B_n les images réciproques d'une famille dénombrable de parties de Y séparant les points.

Réciproquement, à toute partition mesurable de $(X, \mathcal{T}, \mu)$, on peut associer un facteur en quotientant X par la relation d'équivalence $x \sim y$ ssi $y \in \xi(x)$. L'espace quotient est noté X/ξ et la projection canonique $\pi : X \rightarrow X/\xi$. Cet espace est muni de la tribu $\pi_*\mathcal{T} = \{A \subset X/\xi \mid \pi^{-1}A \in \mathcal{T}\}$ et de la mesure $\pi_*\mu$.

Lemme

L'espace $(X/\xi, \pi_\mathcal{T}, \pi_*\mu)$ est un espace de Lebesgue.*

Preuve

On vérifie immédiatement que $\pi_*\mathcal{T}$ est complète relativement à $\pi_*\mu$. Notons par B_n les ensembles intervenant dans la définition de la partition ξ . L'application $\varphi : X \rightarrow \{0, 1\}^{\mathbf{N}}$ définie par $\varphi(x) = \{\mathbf{1}_{B_n}(x)\}$ passe au quotient et donne une application mesurable injective $\bar{\varphi} : (X/\xi, \pi_*\mathcal{T}, \pi_*\mu) \rightarrow \{0, 1\}^{\mathbf{N}}$. Le lemme de mesurabilité montre que l'image des éléments de $\pi_*\mathcal{T}$ sont des ensembles $\varphi_*\mu$ -mesurables. L'application $\bar{\varphi}$ est donc un isomorphisme :

$$\bar{\varphi} : (X/\xi, \pi_*\mathcal{T}, \pi_*\mu) \xrightarrow{\sim} (\{0, 1\}^{\mathbf{N}}, \overline{\mathcal{B}(\{0, 1\}^{\mathbf{N}})}, \varphi_*\mu).$$

On dira que deux facteurs sont *isomorphes* s'il existe un isomorphisme entre les facteurs qui commute aux deux projections.

Considérons un facteur $\varphi : (X, \mathcal{T}, \mu) \rightarrow (Y, \mathcal{S}, \nu)$ et notons ξ la partition associée. Remarquons que l'application φ passe au quotient pour donner une injection mesurable de $(X/\xi, \pi_*\mathcal{T}, \pi_*\mu)$ dans $(Y, \mathcal{S}, \nu)$. Comme ces deux espaces sont des espaces de Lebesgue, cette injection est un isomorphisme.

On en déduit qu'il y a une bijection entre les facteurs et les partitions mesurables.

σ -algèbres et algèbres de fonctions

Soit $(X, \mathcal{T}, \mu)$ un espace probabilisé ; on note $L^0(X, \mathcal{T}, \mu)$ l'algèbre des fonctions mesurables à valeurs réelles. La convergence en probabilité fait de cet espace un espace métrique complet.

Il existe une bijection entre sous σ -algèbres complètes de $\mathcal{T}$ et sous-algèbres fermées de $L^0(X, \mathcal{T}, \mu)$. Cette bijection est donnée par :

$$\begin{array}{ccc} \mathcal{A} & \longrightarrow & L^0(X, \mathcal{A}, \mu) \\ \{A \in \mathcal{T} \mid \mathbf{1}_A \in \mathbf{A}^0\} & \longleftarrow & \mathbf{A}^0 \end{array}$$

L'égalité : $\mathcal{A} = \{A \in \mathcal{T} \mid \mathbf{1}_A \in L^0(X, \mathcal{A}, \mu)\}$ se démontre en approchant les fonctions $\mathcal{A}$ -mesurables par des combinaisons linéaires de fonctions indicatrices d'éléments de $\mathcal{A}$.

L'égalité : $\mathbf{A}^0 = L^0(X, \{A \mid \mathbf{1}_A \in \mathbf{A}^0\}, \mu)$ revient à montrer que si $f \in \mathbf{A}^0$, alors $\mathbf{1}_{f^{-1}(I)} \in \mathbf{A}^0$ pour tout intervalle ouvert $I \subset \mathbf{R}$. Ceci provient du fait suivant, laissé en exercice : on peut trouver une suite de polynômes P_n définis sur $\mathbf{R}$ qui converge simplement vers $\mathbf{1}_I$.

Correspondance de Rokhlin

On déduit des lemmes précédents :

Théorème

Soit $(X, \mathcal{T}, \mu)$ un espace de Lebesgue. Il existe une bijection entre :

- les partitions mesurables de X ;
- les sous σ -algèbres complètes de $\mathcal{T}$;
- les sous-algèbres fermées de $L^0(X, \mathcal{T}, \mu)$;
- les facteurs de $(X, \mathcal{T}, \mu)$, à isomorphisme près.

References

- [A67] V. I. Arnold, A. Avez Problèmes ergodiques de la mécanique classique. Monographies Internationales de Mathématiques Modernes, No. 9 Gauthier-Villars, Éditeur, Paris 1967 ii+243 pp.
- [Co93] D. L. Cohn Measure theory. Reprint of the 1980 original. Birkhäuser Boston, Inc., Boston, MA, 1993. x+373 pp.
- [Du02] R. M. Dudley Real analysis and probability. Revised reprint of the 1989 original. Cambridge Studies in Advanced Mathematics, 74. Cambridge University Press, Cambridge, 2002. x+555 pp.
- [DS88] N. Dunford, J. T. Schwartz Linear operators. Part I. General theory. With the assistance of William G. Bade and Robert G. Bartle. Reprint of the 1958 original. Wiley Classics Library. A Wiley-Interscience Publication. John Wiley and Sons, Inc., New York, 1988. xiv+858 pp.
- [Fe69] H. Federer Geometric measure theory. Die Grundlehren der mathematischen Wissenschaften, Band 153 Springer-Verlag New York Inc., New York 1969 xiv+676 pp.
- [Fu81] H. Furstenberg Recurrence in ergodic theory and combinatorial number theory. M. B. Porter Lectures. Princeton University Press, Princeton, N.J., 1981. xi+203 pp.
- [Gl03] E. Glasner Ergodic theory via joinings. Mathematical Surveys and Monographs, 101. American Mathematical Society, Providence, RI, 2003. xii+384 pp.
- [Irw01] M. C. Irwin Smooth dynamical systems. Reprint of the 1980 original. With a foreword by R. S. MacKay. Advanced Series in Nonlinear Dynamics, 17. World Scientific Publishing Co., Inc., River Edge, NJ, 2001. xii+259 pp.
- [Ka92] S. Katok Fuchsian groups. Chicago Lectures in Mathematics. University of Chicago Press, Chicago, IL, 1992. x+175 pp.
- [HK95] B. Hasselblatt, A. Katok Introduction to the modern theory of dynamical systems. With a supplementary chapter by Katok and Leonardo Mendoza. Encyclopedia of Mathematics and its Applications, 54. Cambridge University Press, Cambridge, 1995. xviii+802 pp.
- [Kr95] U. Krengel Ergodic theorems. With a supplement by Antoine Brunel. de Gruyter Studies in Mathematics, 6. Walter de Gruyter & Co., Berlin, 1985. viii+357 pp.
- [La89] D. F. Lawden Elliptic functions and applications. Applied Mathematical Sciences, 80. Springer-Verlag, New York, 1989. xiv+334 pp.
- [Ma87] R. Mañé Ergodic theory and differentiable dynamics. Translated from the Portuguese by Silvio Levy. Ergebnisse der Mathematik und ihrer Grenzgebiete (3) 8. Springer-Verlag, Berlin, 1987. xii+317 pp.
- [Pa69] W. Parry Entropy and generators in ergodic theory. W. A. Benjamin, Inc., New York-Amsterdam 1969 xii+124 pp
- [PJ82] J. Palis, W. de Melo Geometric theory of dynamical systems. An introduction. Translated from the Portuguese by A. K. Manning. Springer-Verlag, New York-Berlin, 1982. xii+198 pp.
- [RN65] F. Riesz, B. Sz.-Nagy Leçons d'analyse fonctionnelle. Quatrième édition. Académie des Sciences de Hongrie Gauthier-Villars, Editeur-Imprimeur-Libraire, Paris; Akadémiai Kiadó, Budapest 1965 viii+490 pp.
- [Ra05] M. M. Rao Conditional measures and applications. Second edition. Pure and Applied Mathematics (Boca Raton), 271. Chapman & Hall/CRC, Boca Raton, FL, 2005. xxxiv+483 pp.

- [Rob99] C. Robinson Dynamical systems. Stability, symbolic dynamics, and chaos. Second edition. Studies in Advanced Mathematics. CRC Press, Boca Raton, FL, 1999. xiv+506 pp.
- [Rok52] V. A. Rokhlin On the fundamental ideas of measure theory. *Amer. Math. Soc. Translation* 1952, (1952). n°71, 55 pp.
- [Sh78] M. Shub Stabilité globale des systèmes dynamiques. Astérisque, 56. Société Mathématique de France, Paris, 1978. iv+211 pp.
- [Vi80] M. F. Vigneras Arithmétique des algèbres de quaternions. Springer L. Notes Vol 800 (1980).