Systèmes Dynamiques

Patrick Bernard

22 décembre 2017

Version très préliminaire des notes de cours.
Quelques références conseillées :
Coudènes, *Théorie ergodique et systèmes dynamiques / Ergodic Theory ans Dynamical Systems*
Mañé, *Ergodic Theory and Differentiable Dynamics*
Milnor, *Dynamics : Introductory Lectures*
Parry, *Topics in ergodic theory*
Pollicott et Yuri, *Dynamical Systems and Ergodic Theory*

Table des matières

1 Introduction, points fixes et périodiques.

Il y a plusieurs contextes mathématiques que l'on peut qualifier de systèmes dynamiques. Dans tous les cas, il y a un espace X qui décrit les états possible du système, et un temps $\mathbb{T}$. On considérera les cas $\mathbb{T} = \mathbb{Z}, \mathbb{N}, \mathbb{R}, \mathbb{R}^+$.

Le cas $\mathbb{T} = \mathbb{N}$. On se donne une application $\varphi : X \longrightarrow X$, et on étudie les suites définies par récurrence $x_{n+1} = \varphi(x_n)$. Une telle suite est déterminée par sa condition initiale x_0, on dit que c'est l'orbite du point x_0. Dit autrement, on étudie la suite des applications φ^n (φ itérée n fois), pour laquelle $x_n = \varphi^n(x_0)$. On a bien sur $\varphi^{n+m} = \varphi^n \circ \varphi^m$.

Si l'application φ est inversible, on peut définir φ^n pour tout $n \in \mathbb{Z}$, la relation ci-dessus restant vérifiée. C'est le cas $\mathbb{T} = \mathbb{Z}$.

Dans le cas $\mathbb{T} = \mathbb{R}$, on considère un flot $\varphi^t(x) = \varphi(t, x) : \mathbb{R} \times X \longrightarrow X$ vérifiant la relation $\varphi^{t+s} = \varphi^t \circ \varphi^s$. L'une des manières les plus fréquentes de définir un tel flot est de se donner un champ de vecteurs $V(x)$ sur X. Ceci a un sens si X est une variété, ou un espace $\mathbb{R}^d$. On considère alors l'équation différentielle $x'(t) = V(x(t))$. Si V est assez régulier (C^1 ou même Lipschitz), alors pour toute condition initiale x_0, il y a une et une seule solution maximale de cette équation vérifiant $x(0) = x_0$. Cette solution est définie sur un intervalle $]T_-, T_+[$ contenant 0. On peut rassembler toutes ces solutions en une application

$$\varphi : \mathbb{R} \times X \supset U \longrightarrow X$$

telle que, pour chaque $x \in X$, L'application $U_x \ni t \longmapsto \varphi(t, x)$ est la solution maximale de l'équation, où $U_x = \{t : (t, x) \in U\}$. Le domaine de définition U est un ouvert de $\mathbb{R} \times X$, et l'application φ est C^r si le champ V est C^r. On dit que le champ V est complet si $U = \mathbb{R} \times X$, c'est à dire si toutes les solutions maximales de l'équation différentielle sont définies sur $\mathbb{R}$. Alors, $\varphi^t(x) = \varphi(t, x)$ est un flot sur X. On considérera principalement des champs complets. On rappelle quelques résultats utiles à ce propos :

Proposition 1.1. *Tout champ de vecteurs à support compact est complet. En particulier, tout champ de vecteurs sur une variété compacte est complet.*

Si V est un champ de vecteurs, il existe une fonction f, strictement positive et lisse, telle que fV est complet.

Si x_0 est un point tel que le temps d'existence T_+ de la solution maximale est fini, alors l'image $x([0, T^+[) \subset X$ de l'orbite n'est pas relativement compacte.

Dans tous les cas, on considère donc une action du (semi)-groupe $\mathbb{T}$ sur l'espace X.

On parle de semi-flot lorsque $\mathbb{T} = \mathbb{R}^+$. Même si on considérera plus rarement ce cas, il apparaît de manière naturelle dans les cas suivants :

Dans l'études des équations aux dérivées partielles d'évolution, par exemple de type parabolique. De telles équations induisent souvent un semi-flot sur un espace X de dimension infinie.

Dans l'étude des flots, on est souvent amené à considérer des parties positivement invariantes, c'est à dire des parties $Y \subset X$ pour lesquelles $\varphi^t(Y) \subset Y$ pour tout $t \geqslant 0$ (mais pas forcément pour $t < 0$). La famille des applications $\varphi^t_{|Y}$ est alors un semi-flot. Si le flot φ^t provient d'un champ de vecteurs complet V et que Y est un ouvert de X, le champ $V_{|Y}$ est complet vers le futur, mais pas vers le passé.

Nature des orbites

On dit que x est un point fixe de l'application φ si $\varphi(x) = x$. On dit que x est un point périodique si il existe $n \geqslant 1$ tel que $\varphi^n(x) = x$. L'ensemble des entiers n pour lesquels $\varphi^n(x) = x$ est l'ensemble des multiple d'un entier $T > 0$, que l'on appelle la période minimale de x. Les points fixe sont donc les orbites périodiques dont la période minimale est égale à 1. Il existe trois types d'orbites pour une application φ :

Les orbites périodiques.

Les orbites injectives.

Les orbites préperiodiques (x n'est pas périodique, mais son orbite contient un point périodique). Si φ est inversible, toute orbite préperiodique est périodique.

Considérons maintenant le cas d'un flot engendré par un champ de vecteurs complet V de classe C^1. On dit que x est un point fixe (ou un point critique) si $V(x) = 0$, ou de manière équivalente si $\varphi^t(x) = x$ pour tout t. On dit que x est un point périodique si il existe $t > 0$ tel que $\varphi^t(x) = x$. L'ensemble des périodes est une sous-groupe fermé de $\mathbb{R}$, c'est donc ou bien $\mathbb{R}$ (cas d'un point fixe) ou bien $T\mathbb{Z}$ pour un réel $T > 0$, qui est appelé la période minimale de x.

Il y a trois types d'orbites pour un flot :

Les points fixes (dits aussi points singuliers).

Les orbites périodiques (de période minimale strictement positive). Si $x(t)$ est une orbite de période $T > 0$, alors la courbe x engendre un plongement du cercle $\mathbb{R}/T\mathbb{Z}$ dans X. Plus précisément, en notant $\pi : \mathbb{R} \longrightarrow \mathbb{R}/T\mathbb{Z}$, il existe un plongement $\theta : \mathbb{R}/T\mathbb{Z} \longrightarrow X$ tel que $x = \pi \circ \theta$.

Les orbites injectives. Dans ce cas, l'application $t \longmapsto \varphi^t(x)$ est une immersion injective de $\mathbb{R}$ dans X. En effet, $x'(t) = V(x(t))$, et ce vecteur est non-nul pour tout t, sinon l'orbite serait un point fixe.

Dans un semi-flot, il peut aussi exister des orbites prépériodiques.

Conjugaisons Soit $\phi : X \longrightarrow X'$ une bijection. Si $\psi = \phi^{-1} \circ \varphi \circ \phi$, alors ϕ envoie les orbites de φ sur les orbites de ψ. L'existence d'une telle bijection ϕ implique que les dynamiques de φ et ψ ont certaines similarités (d'autant plus si ϕ a certaines propriétés de régularité). On dit que ϕ conjugue φ et ψ.

On dit que ϕ conjugue les flots φ^t et ψ^t elle conjugue les applications φ^t et ψ^t pour tout t.

Dans le cas où X et X' sont des variétés, et où les flots φ^t et ψ^t sont engendrés par des champs de vecteurs V et W, le difféomorphisme ϕ est une conjugaison si et seulement si

$$W(\phi(x)) = d\phi_x \cdot V(x)$$

pour tout $x \in X$ (on dit que W est l'image directe de V par ϕ).

Une façon d'étudier une système dynamique est de le conjuguer à un système plus simple. C'est souvent difficile globalement, mais la question est aussi intéressante au niveau local. On verra par exemple qu'au voisinage d'un point fixe hyperbolique, un système est conjugué à son linéarisé par un homéomorphisme. En ce qui concerne les points réguliers (points non fixes) des champs de vecteurs, le problème de conjugaison locale est facile :

Proposition 1.2. *Soit V un champ de vecteurs C^r sur la variété X, et soit x_0 un point régulier de V, c'est à dire que $V(x_0) \neq 0$. Il existe alors un difféomorphisme local $\phi : X \supset U \longrightarrow U' \subset \mathbb{R}^d$, où U est un ouvert contenant x_0 et U' est un ouvert de $\mathbb{R}^d$, qui conjugue V à un champ de vecteurs constant V' sur U'. On peut même demander que $V' = (1, 0, \dots, 0)$ et que $U' =]-1, 1[\times B$, où B est une boule de $\mathbb{R}^{d-1}$. On dit alors que U est une boite de flot locale de V.*

◄ Soit $f : \mathbb{R}^{d-1} \longrightarrow X$ une application telle que $f(0) = x_0$ et $df_{x_0} \cdot \mathbb{R}^{d-1}$ est un supplémentaire de $V(x_0)$ dans $T_{x_0}M$. On considère alors l'application

$$\psi : \mathbb{R} \times \mathbb{R}^{d-1} \ni (t, y) \longmapsto \varphi^t(f(y)) \in X$$

où φ est le flot de V. On a

$$d\psi_{(y,t)} \cdot (1, 0) = \partial_t \psi(y, t) = V(\psi(y, t))$$

c'est à dire que ψ conjugue le champ constant $(1, 0)$ et le champ V. On constate que

$$d\psi_{(0,0)} \cdot (s, v) = v + sV(x_0),$$

est un isomorphisme linéaire de $\mathbb{R} \times \mathbb{R}^{d-1}$ dans $T_{x_0}M$, donc ψ est un difféomorphisme local. Si B est un petite boule ouverte centrée en 0 dans $\mathbb{R}^{d-1}$ et I un petit intervalle ouvert contenant 0, alors ψ est un difféomorphisme de $U' := B \times I$ sur son image U, qui est un ouvert de X. L'inverse $\phi : U \longrightarrow U'$ est le difféomorphisme cherché. ►

On peut obtenir un résultat un peu moins local :

Proposition 1.3. *Soit V un champ de vecteurs C^r sur la variété X, et soit $x(t) : [0, T] \longrightarrow X$ un segment d'orbite injectif. Il existe alors un tube de flot contenant $x([0, T])$. Plus précisément, il existe un intervalle ouvert I contenant $[0, T]$, une boule ouverte B de $\mathbb{R}^{d-1}$, un voisinage ouvert U de $x([0, T])$ et un difféomorphisme $\phi : U \longrightarrow I \times B$ qui conjugue le champ V et le champ $(1, 0)$.*

◄ On poursuit la preuve précédente. Comme $\psi(t + s, y) = \varphi^t \circ \psi(s, y)$, on a $d\psi_{(t,0)} = d\varphi^t_{x_0} \circ d\psi_{(0,0)}$, cette application linéaire est donc un isomorphisme. On peut donc supposer, en choisissant I et B comme dans l'énoncé assez petits, que ψ est un difféomorphisme local sur $I \times B$. Montrons maintenant que l'on peut, quitte à diminuer I et B, le supposer injectif. C'est alors un difféomorphisme sur son image, ce qui termine la preuve.

Si ψ n'est injectif sur aucun voisinage $I \times B$, alors il existe deux suites $(t_n, x_n) \neq (t'_n, x'_n)$telles que $\psi(t_n, x_n) = \psi(t'_n, x'_n)$ et $t_n \longrightarrow t \in [0, T], t'_n \longrightarrow t' \in [0, T]$, $x_n \longrightarrow 0$, $x'_n \longrightarrow 0$. A la limite, par injectivité de l'orbite $x_{|[0,T]}$, on obitent que $t = t'$. Comme ψ est un difféomorphisme local au voisinage de $(t, 0)$, ceci implique que les suites (t_n, x_n) et (t'_n, x'_n) sont égales après un certain rang, une contradiction. ►

Il y a plusieurs façons de faire des liens entre les systèmes à temps continu et à temps discret.

La plus évidente consiste, étant donné un flot φ^t, à considérer l'application φ^T pour un temps T donné. La dynamique du flot φ^t et celle de l'application φ^T sont fortement reliées, on en verra plusieurs exemples.

Une autre méthode pour associer une application à un flot consiste à considérer l'application de retour associée à une section de Poincaré. Cette méthode à l'avantage de diminuer la dimension.

Une section de Poincaré est une sous variété Y de X, de codimension un, transverse au champ de vecteurs, c'est à dire que $T_yX = T_yY \oplus \mathbb{R}V(y)$ pour tout $y \in Y$. Pour $y \in Y$, on dit que $z \in Y$ est le premier retour de y dans Y si il existe $T > 0$ tel que $z = \varphi^T(y)$ et $\varphi^t(y) \notin Y$ pour tout $t \in]0, T[$. On dit que z est un premier retour régulier si de plus $\varphi^t(y) \notin \bar{Y}$ pour tout $t \in]0, T[$.

Proposition 1.4. *Soit Y une section de Poincaré pour le champ V de classe C^r. L'ensemble des points de Y admettant un premier retour régulier est un ouvert Y_1 de Y, et l'application $\psi : Y_1 \longrightarrow Y$ qui a chaque point de Y_1 associe son premier retour dans Y est C^r. C'est un difféomorphisme sur un ouvert Y_2 de Y.*

En général, Y_1 peut être vide. Il se peut même qu'aucun point de Y ne revienne dans Y. Toutefois, cette construction est souvent très utile. Même si ψ n'est pas une application de Y_1 dans lui-même, on peut souvent penser à ψ comme engendrant un système dynamique discret dont l'étude aide à comprendre le flot de V.
◄ Soit y un point qui admet un premier retour régulier $z = \varphi^T(y)$. Au voisinage de z, soit f une équation de Y, c'est à dire une fonction $f : X \longrightarrow \mathbb{R}$ de classe C^r telle que $df_z \neq 0$ et telle que $Y \subset f^{-1}(0)$. Pour chercher les retours dans Y des points voisins de y, on résout l'équation $g(t,x) := f \circ \varphi(t,x) = 0$ au voisinage de sa solution (T, y). On constate que $\partial_t g(T, y) = df_z \cdot V(z) \neq 0$. Il existe donc un voisinage U de y et une fonction $\tau : U \longrightarrow \mathbb{R}$, de classe C^r, telle que $\varphi^{\tau(x)}(x) \in Y$ pour tout $x \in U$. De plus, $\tau(x)$ est la seule solution de l'équation $g(t,x) = 0$ proche de T. L'application $\psi(x) := \varphi^{\tau(x)}(x) : U \longrightarrow Y$ est C^r, elle associe à chaque point $x \in U$ un de ses retours dans Y.

Il reste à montrer que, si U est un voisinage de y assez petit, alors c'est bien le premier retour et qu'il est régulier. Il faut donc montrer que $\varphi^t(x) \notin \bar{Y}$ pour tout $x \in U$ et tout $t \in]0, \tau(x)[$. Supposons, par contradiction, que cette propriété ne soit satisfaite sur aucun voisinage U de y. Il existe alors une suite x_n tendant vers y, et pour chaque n un temps $t_n \in]0, \tau(x_n)[$ tel que $\varphi^{t_n}(x_n) \in Y$. En extrayant une sous-suite, on peut supposer que la suite t_n converge vers une limite $s \in [0, T]$, pour laquelle $\varphi^s(y) \in \bar{Y}$. Comme z est un retour régulier de y, on ne peut pas avoir $s \in]0, T[$. Si l'on a $s = T$, alors pour n grand t_n est une solution de l'équation $g(t_n, x_n) = 0$ qui est proche de T, mais différente de $\tau(x_n)$, ce qui contredit la conclusion du théorème des fonctions implicites. Une application similaire du théorème des fonctions implicites au voisinage de $(0, y)$ montre que 0 est localement l'unique solution de l'équation $\varphi^t(x) \in Y$. On ne peut donc pas avoir $s = 0$.

On peut renverser le temps en considérant le champ $-V$. Y est une section de Poincaré pour $-V$. Notons Y_2 l'ensemble des points de Y qui admettent un premier retour régulier pour $-V$. On montre comme ci-dessus que Y_2 est un ouvert de Y et que l'application de retour ϕ est C^r. Il est clair que z est un premier retour régulier de y pour V si et seulement si y est un premier retour régulier de z pour $-V$. On a donc $Y_2 = \psi(Y_1)$ et $\phi = \psi^{-1}$. ►

Définition 1.5. *Soit φ une application C^1 de la variété X. On dit que le point fixe x de φ est non dégénéré si le linéarisé $L = d\varphi_x$ n'admet pas 1 pour valeur propre.*

Le linéarisé L est un endomorphisme de l'espace tangent $T_x X$. Le système linéarisé $v_{n+1} = Lv_n$ est une approximation du système φ au voisinage du point fixe x.

Proposition 1.6. *Soit $\varphi_\mu(x) = \varphi(\mu, x) : \mathbb{R} \times X \longrightarrow X$ une application C^1, que l'on voit comme une famille d'applications C^1 de X. Soit x_0 un point fixe non dégénéré de φ_0. Alors, il existe un voisinage U de x_0 dans X, un intervalle ouvert I contenant 0, et une application $I \ni \mu \longmapsto x_\mu \in U$, de classe C^1, telle que x_μ est, pour chaque $\mu \in I$, le seul point fixe de φ_μ dans U.*

◄ C'est le théorème des fonctions implicites pour l'équation $F(\mu, x) := \varphi(\mu, x) - x = 0$ (on se place dans une carte en x_0 pour écrire cette différence). L'hypothèse de non-degenerescence est en effet exactement l'inversibilité de $\partial_x F(0, x_0)$. ►

Le cas des orbites périodiques est similaire. On dit que x est un point périodique non dégénéré de période n si c'est un point fixe non dégénéré de φ^n. Un point périodique non dégénéré de période n est isolé (parmi les points périodiques de période n) et survit à une perturbation de l'application.

Dans les cas des flots, il est utile de distinguer les cas des points fixes et le cas des orbites périodiques.

Définition 1.7. *Soit V un champ de vecteurs C^1 de la variété X. On dit que le point fixe x_0 de V est non dégénéré si le linéarisé $L = dV_{x_0}$ est inversible.*

Si $X = \mathbb{R}^d$, le champ de vecteurs V est simplement une application $V : \mathbb{R}^d \longrightarrow \mathbb{R}^d$, on définit alors le linéarisé par la formule $L = dV_{x_0}$, c'est un endomorphisme de $\mathbb{R}^d$. On a alors $(d\varphi^t)_{x_0} = e^{tL}$ pour tout t, ceci explique pourquoi la valeur propre 1 dans le cas des applications correspond à la valeur propre 0 pour le champ de vecteurs. Dans le cas où X est une variété, il n'est pas clair que le linéarisé L est bien défini en tant qu'endomorphisme de $T_x X$ (en fait, il n'est bien défini que lorsque $V(x) = 0$). Une façon de le définir est de poser $L = \frac{d}{dt}_{|t=0} d\varphi^t_x$ en remarquant que $d\varphi^t_x$ est bien un endomorphisme de $T_x M$ pour tout t (car $\varphi^t(x) = x$). Plus classiquement, on peut calculer le linéarisé L en exprimant le champ de vecteurs V dans des cartes. Soient W^1 et W^2 les expressions de V dans deux cartes différentes

Proposition 1.8. *Soit $V_\mu(x) = V(\mu, x) : \mathbb{R} \times \mathbb{R}^d \longrightarrow \mathbb{R}^d$ une application C^1, que l'on voit comme une famille de champs de vecteurs C^1 de $X = \mathbb{R}^d$. Soit x_0 un point fixe non dégénéré de V_0. Alors, il existe un voisinage U de x_0 dans X, un intervalle ouvert I contenant 0, et une application $I \ni \mu \longmapsto x_\mu \in U$, de classe C^1, telle que x_μ est, pour chaque $\mu \in I$, le seul point fixe de V_μ dans U.*

◄ C'est le théorème des fonctions implicites pour l'équation $V(\mu, x) = 0$. ►

Considérons maintenant le cas d'une orbite périodique d'un champ de vecteurs V, de période minimale $T > 0$. Les points de cette orbite sont des points fixes de l'application φ^T. Il faut noter cependant que ce ne sont jamais des points fixes non dégénérés. En effet, comme chacun des points de l'orbite périodique du flot est un point fixe de φ^T, ils ne sont pas isolés. Plus directement, l'hypothèse de non dégénerescence n'est pas satisfaite en raison de l'égalité suivante :

$$d\varphi_x^T \cdot V(x) = V(x).$$

$V(x)$ est donc un vecteur propre de $d\varphi_x^T$ associé à la valeur propre 1. Pour démontrer l'égalité, on observe que $\varphi^T \circ \varphi^t(x) = \varphi^t(x)$ pour tout t, et on dérive par rapport à t en $t = 0$.

Il est utile pour décrire cette situation d'introduire une section de poincaré locale en x, c'est à dire un petit disque plongé $Y \subset X$, centré en x et transverse à V. Comme l'orbite O de x est une partie compacte de X on peut supposer que $\bar{Y} \cap O = \{x\}$, c'est à dire que T est le temps de premier retour de x dans Y, et que ce retour est régulier. Il existe alors un voisinage Y_1 de x dans Y est une application de premier retour régulière $\psi : Y_1 \longrightarrow Y$, qui vérifie $\psi(x) = x$.

Propriété 1.9. *La multiplicité de* 1 *en tant que valeur propre de* $d\varphi_x^T$ *est exactement un de plus que la multiplicité de* 1 *en tant que valeur propre de* $d\psi_x$.

On dit que x est une orbite périodique non dégénérée si c'est un points fixe non dégénéré de ψ, c'est à dire si la multiplicité de 1 en tant que valeur propre de $d\varphi_x^T$ est égale à 1.
◄ On a $\varphi^T(x) = \varphi^{T-\tau(x)} \circ \psi(x)$ pour tout $x \in Y$. On a déjà vu que $d\varphi_x^T \cdot V(x) = V(x)$. Dans une base de $T_x X$ constituée de $V(x)$ et d'une base de $T_x Y$, on a la décomposition par blocs

$$d\varphi_x^T = \begin{bmatrix} 1 & -\partial_x \tau(x) \\ 0 & d\psi_x \end{bmatrix}. \blacktriangleright$$

Proposition 1.10. *Soit* $V_\mu(x) = V(\mu, x) : \mathbb{R} \times \mathbb{R}^d \longrightarrow \mathbb{R}^d$ *une application* C^1*, que l'on voit comme une famille de champs de vecteurs* C^1 *de* $X = \mathbb{R}^d$*. Soit* x_0 *un point périodique non dégénéré de* V_0 *de période minimale* $T > 0$*. Alors, il existe un voisinage* U *de* x_0 *dans* X*, un intervalle ouvert* I *contenant* 0*, un intervalle ouvert* J *contenant* T*, et une application* $I \ni \mu \longmapsto (x_\mu, T_\mu) \in U \times J$*, de classe* C^1*, telle que* x_μ *est, pour chaque* $\mu \in I$*, un point périodique de période minimale* T_μ *de* V_μ*. De plus, l'orbite de* x_μ *est la seule orbite périodique de* V_μ *qui entre dans* U *et dont la période minimale appartient à* J.

◄ Soit Y une section de Poincaré locale en x_0. Le point x_0 est un point fixe non dégénéré de l'application de premier retour ψ. Montrons que l'application de retour $\psi_\mu(x) = \psi(\mu, x)$ et le temps de retour $\tau(\mu, x)$ sont C^1.

On peut considérer l'application $V(\mu, x)$ comme une champ de vecteurs $\tilde{V}$ sur $\mathbb{R} \times \mathbb{R}^d$ dont la première composante est nulle. Son flot est alors $\tilde{\varphi}^t(\mu, x) = (\mu, \varphi_\mu^t(x))$, où φ_μ^t est le flot du champ V_μ sur $\mathbb{R}^d$. Si I est un intervalle ouvert contenant 0, alors $\tilde{Y} := I \times Y$ est une section de poincaré locale pour $\tilde{V}$ en $(0, x_0)$, qui est une orbite périodique de période T. Les fonctions $\tau(\mu, x)$ et $\psi(\mu, x)$ sont les temps et les applications de retour associées à la section $\tilde{Y}$. De plus, si J est un petit intervalle contenant T, pour tout $(\mu, y) \in I \times Y$, $\tau(\mu, y)$ est l'unique temps $t \in J$ pour lequel $\varphi^t(\mu, y) \in \tilde{Y}$.

Comme x_0 est un point fixe non dégénéré pour ψ_0, il existe une application $\mu \longrightarrow x_\mu$ et un voisinage W de x_0 dans Y tel que x_μ est le seul point fixe de ψ_μ dans W, c'est alors un point périodique de période $T_\mu = \tau(\mu, x_\mu)$ pour V_μ.

Il existe un voisinage U de x dans X qui a la propriété que toute orbite de U passe par W (on peut prendre par exemple une petite boite de flot). La propriété d'unicité de l'énoncé en découle : Si $x \in U$ est périodique pour V_μ, de période S contenue dans J, alors l'orbite de x intersecte W en un point y qui est lui aussi S-périodique. Il vérifie donc $\varphi_\mu^S(y) = y$, ce qui implique que $y = \psi(y)$ et donc $y = x_\mu$. ►

Considérons une dernière situation, celle d'un champ de vecteur laissant invariante une fonction h. Si $h : X \longrightarrow \mathbb{R}$ est une fonction C^1 sur X, on dit que h est invariante par le champ de vecteurs V si $h \circ \varphi^t = h$ pour tout t. Ceci est équivalent à demander que la fonction $dh \cdot V$ soit identiquement nulle. En effet,

$$\partial_t(h \circ \varphi^t(x)) = dh_{\varphi(x)} \cdot V(\varphi^t(x)) = (dh \cdot V)(\varphi^t(x)).$$

On rencontre souvent des fonctions invariantes, c'est par exemple le cas de l'énergie dans de nombreux systèmes issus de la physique.

Considérons une valeur régulière e de la fonction h, de sorte que $h^{-1}(e)$ est une sous-variété de X. Considérons maintenant un point périodique x_e dans $h^{-1}(e)$, de période minimale $T > 0$.

L'orbite de x_e ne peut pas être non dégénérée au sens ci-dessus. Considérons en effet une section de Poincaré locale Y en x_e et l'application de retour ψ. Le point x_e est un point régulier pour la restriction de h à Y. Notons $\ell := d(h_{|Y})_{x_e}$, c'est une forme linéaire sur $T_{x_e}Y$. Comme $h\circ\psi = h$, on a $\ell \circ L = \ell$, où $L = d\psi_{x_e}$. Ceci implique que 1 est valeur propre de L^*, donc de L.

Proposition 1.11. *Supposons que* 1 *est valeur propre simple de* $d\psi_{x_e}$. *Alors, il existe un intervalle* I *contenant* e *et une application* $(x(a), T(a)) : I \longrightarrow X$ *telle que, pour chaque* $a \in I$, *le point* $x(a)$ *est périodique de période* $T(a)$ *et contenu dans* $h^{-1}(a)$.

L'application

$$I \times \mathbb{S}^1 \ni (a, \theta) \mapsto \varphi^{\theta T(a)}(x(a)) \in X$$

est un plongement du cylindre dans X *dont l'image est invariante.*

◄ Donnons d'abord une démonstration légèrement informelle. L'orbite de x_e est non dégénérée en tant qu'orbite du champ V sur la variété $h^{-1}(e)$. Pour a voisin de e, la restriction V_a de V à $h^{-1}(a)$ est une perturbation de V_e. On applique alors le théorème précédent qui nous donne une orbite périodique x_a de V_a. On est toutefois pas exactement dans le cadre du théorème précédent car l'espace $h^{-1}(a)$ dépend du paramètre a. Même si cette preuve peut être rendue rigoureuse, il est plus simple d'utiliser une section de Poincaré.

On considère une section de Poincaré locale Y en x_e et l'application de retour ψ. On a $h\circ\psi = h$. On veut montrer qu'il existe un intervalle I contenant e et un courbe régulière $a \mapsto x_a$ de points fixes de ψ tels que $h(x_a) = a$. On se place dans une carte locale en x_e pour laquelle la fonction h est la dernière coordonnée. On a dont $Y = \mathbb{R}^{d-2} \times \mathbb{R}$, et, en notant $y = (z, a)$ les points de ce produit, on a $h(z, a) = a$. Comme l'application de retour ψ préserve a, elle est de la forme $\psi(z, a) = (F(z, a), a)$ où $F : Y \longrightarrow \mathbb{R}^{d-2}$ est régulière. En notant $F_a(z) = F(z, a)$ et $x_e = (z_e, e)$, on a $F_e(z_e) = z_e$, et ce point fixe est non dégénéré pour l'application F_e. On déduit donc l'existence d'une courbe $a \mapsto z_a$ telle que $F_a(z_a) = z_a$. On pose alors $x_a = (z_a, a)$.

Concernant le dernier point, il faut d'abord remarquer que l'application $F(a, \theta) = \varphi^{\theta T(a)}(x(a))$ est bien définie sur $I \times \mathbb{S}^1$ car $x(a)$ est $T(a)$-périodique. De plus, elle est injective, et c'est un difféomorphisme local. Si on considère un intervalle ouvert J dont la fermeture est contenue dans I, alors l'application F restreinte au compact $\bar{J} \times S^1$ est donc un plongement topologique, c'est à dire un homéomorphisme sur son image. La restriction de F à $J \times S^1$ est donc un difféomorphisme local et un plongement topologique, c'est à dire un plongement. ►

2 Points fixes asymptotiquement stables des systèmes discrets

Notion de stabilité d'un point fixe. On se place dans le contexte d'une application continue $\varphi : X \longrightarrow X$ où X est un espace métrique.

Le point fixe x_0 est dit *Lyapounov stable* si, pour tout voisinage U de x_0, il existe un voisinage V de x_0 tel que $\varphi^n(V) \subset U$ pour tout $n \in \mathbb{N}$.

Le point fixe x_0 est dit *asymptotiquement stable* si il est Lyapounov stable et si, de plus, il existe un voisinage W de x_0 tel que toute orbite partant dans W converge vers x_0.

Le bassin de x_0 est l'ensemble des points x tels que $\varphi^n(x) \longrightarrow x_0$ en $+\infty$. Si x_0 est un point fixe asymptotiquement stable, alors son bassin est un voisinage ouvert de x_0.

La stabilité de Lyapounov est une hypothèse nécessaire (le point fixe ci-dessous n'est pas asymptotiquement stable).

Les contractions (applications de constante de Lipschitz strictement inférieure à 1) fournissent des exemples de points asymptotiquement stables :

Théorème 2.1. *Soit X un espace métrique complet et φ une contraction. Alors φ admet un unique point fixe x_0. Ce point fixe est asymptotiquement stable et son bassin est X entier.*

Quand on s'intéresse au comportement asymptotique, ces dynamiques sont les plus simples que l'on puisse imaginer : toutes les orbites convergent vers x_0.
◄ On considère n'importe quel point y_0, et la suite $y_n = \varphi^n(y_0)$ associée. On a alors $d(y_{n+1}, y_n) \leqslant bd(y_n, y_{n-1})$, où b est la constante de Lipschitz de φ. Ceci implique que

$$d(y_n, y_m) \leqslant \sum_{i \geqslant n} b^i d(y_0, y_1) \leqslant \frac{b^n}{1-b} d(y_0, y_1)$$

pour tout $m \geqslant n$. On en déduit que la suite y_n est de Cauchy, et donc qu'elle a une limite x_0. En passant à la limite dans l'égalité $y_{n+1} = \varphi(y_n)$, on obtient que $x_0 = \varphi(x_0)$: x_0 est un point fixe. Si z est une autre condition initiale, alors $d(z_n, x_0) \leqslant b^n d(z, x_0) \longrightarrow 0$: toutes les orbites convergent vers x_0. Pour vérifier la stabilité de Lyapounov de x_0, il suffit de constater que les boules centrées en x_0 sont invariantes par φ. ►

Exercice 2.1. *Soit φ une application continue de X. Supposons que il existe N tel que φ^N est une contraction. Posons*

$$d_1(x, y) = d(x, y) + ad(\varphi(x), \varphi(y)) + \cdots + a^{N-1} d(\varphi^{N-1}(x), \varphi^{N-1}(y)).$$

Pour a convenablement choisi, montrer que d_1 est une distance sur X, équivalente à d, que (X, d_1) est complet, et que φ est une contraction pour d_1. Conclure qu'il existe un unique point fixe de φ, qui est asymptotiquement stable et attire X entier.

On peut poursuivre l'étude des contractions en s'intéressant au problème de conjugaison :

Proposition 2.2. *Soit X un espace métrique complet et soit φ un homéomorphisme de X qui est une contraction. Soit ψ un homéomorphisme qui est une contraction égale à φ en dehors d'une boule de rayon fini. Alors ψ est conjuguée à φ, c'est à dire qu'il existe un homéomorphisme h de X tel que $h \circ \varphi = \psi \circ h$.*

On peut penser à h comme à un changement de coordonnées qui transforme φ en ψ. Si l'on ne suppose pas que ψ est un homéomorphisme, on obtient quand même une application continue h, on dit que c'est une semi-conjugaison.
◄ Soit x_0 le point fixe de φ, y_0 le point fixe de ψ. Il existe $R > 0$ tel que $\psi = \varphi$ en dehors de la boule $B(x_0, R)$. Pour tout $x \neq 0$, on a $d(\varphi^{-1}(x), x_0) \geqslant d(x, x_0)/b$ donc $\varphi^{-n}(x)$ n'appartient pas à la boule $B(x_0, R)$ pour n assez grand. On déduit que la suite $\psi^n \circ \varphi^{-n}(x)$ se stabilise pour n grand. Posons $h(x) := \lim \psi^n \circ \varphi^{-n}(x)$ pour $x \neq x_0$, et $h(x_0) = y_0$. Montrons que h est continue. Pour tout $\epsilon > 0$, il existe N tel que $h = \psi^N \circ \varphi^{-N}$ en dehors de $B(x_0, \epsilon)$. La continuité de h dans le complémentaire de x_0 en découle. Il reste à montrer que $h(x_n) \longrightarrow y_0$ lorsque $x_n \longrightarrow x_0$. On note k_n le plus grand entier pour lequel $\varphi^{-k_n}(x_n) \in B(x_0, R)$. La suite k_n tend vers l'infini. Pour tout $i \geqslant 0$, on a $\psi^i \circ \varphi^{-i} \circ \varphi^{-k_n}(x) = \varphi^{-k_n}(x_n)$, donc $h(x_n) = \psi^{k_n} \circ \varphi^{-k_n}(x_n)$ donc $d(x_0, h(x_n)) \leqslant b^{k_n} R$ où $b =$ Lip (ψ). On déduit que h est continue.

On a $\psi \circ h(x_0) = \psi(y_0) = y_0 = h(x_0) = h(\varphi(x_0))$. Pour $x \neq x_0$, il existe N tel que $h(x) = \psi^n \circ \varphi^n(x)$ pour tout $n \geqslant N$. Pour $x \neq x_0$, on a

$$h(\varphi(x)) = \lim \psi^n \circ \varphi^{1-n}(x) = \psi(\lim \psi^{n-1} \circ \varphi^{1-n}(x)) = \psi \circ h(x).$$

On a montré que h est une semi-conjugaison. Dans le cas où ψ est un homéomorphisme, on pose de la même façon $g := \lim \varphi^n \circ \psi^{-n}$. C'est une application continue.

On a $g \circ h(x_0) = x_0$ et, pour $x \neq x_0$, il existe N tel que $h(x) = \psi^N \circ \varphi^{-N}(x)$ et $g(h(x)) = \varphi^N \circ \psi^{-N}(h(x))$, ce qui implique que $g \circ h(x) = x$. On montre de même que $h \circ g = id$. Donc h est un homéomorphisme. ►

Pour montrer que le point fixe x_0 d'une contraction est asymptotiquement stable, on a utilisé la fonction $f(x) := d(x, x_0)$. On a $f \circ \varphi \leqslant bf$, ce dont on conclut que $f(x_n) \longrightarrow 0$. On peut généraliser un peu cette méthode.

Théorème 2.3. *Supposons que X est localement compact et que x_0 est un point fixe de φ. Supposons de plus qu'il existe une fonction de Lyapounov stricte en x_0, c'est à dire une fonction continue $f : U \longrightarrow [0, \infty)$, où U est un voisinage de x_0, telle que les fonctions f et $f - f \circ \varphi$ sont strictement positives sur un voisinage pointé $V - \{x_0\}$ de x_0.*

Alors x_0 est asymptotiquement stable.

Quelques digressions sont utiles avant de démontrer ce résultat. On dit que f est une fonction de Lyapounov si $f \circ \varphi - f \leqslant 0$. Le point x est dit strict si $f \circ \varphi(x) < f(x)$, il est dit neutre en cas d'égalité. Le point x est dit totalement neutre si $f \circ \varphi^n(x) = x$ pour tout $n \in \mathbb{N}$. On dit que $a \in \mathbb{R}$ est une valeur (totalement) neutre de f si il existe un point (totalement) neutre x de f tel que $f(x) = a$.

On définit, pour tout point x, l'omega-limite $\omega(x) \subset X$ comme l'ensemble des points d'accumulation de la suite $\varphi^n(x), n \in \mathbb{N}$. Autrement dit, c'est l'ensemble des limites de sous-suites de la forme $\varphi^{n_k}(x)$, $n_k \longrightarrow \infty$. Dans le cas inversible, on définit $\alpha(x)$ comme l'ensemble des points d'accumulations en $-\infty$.

Lemme 2.4. *L'ensemble $\omega(x)$ est fermé et positivement invariant. Si l'espace X est compact, alors $\omega(x)$ est non-vide pour tout x et $\varphi(\omega(x)) = \omega(x)$.*

◄ Tout point y de $\omega(x)$ est la limite d'une suite de la forme $\varphi^{n_k}(x)$. Alors, $\varphi(y)$ est la limite de la suite $\varphi^{n_k+1}(x)$, il appartient donc à $\omega(x)$. On a

$$\omega(x) = \cap_{n \geqslant 0} \overline{\{\varphi^n(x), \varphi^{n+1}(x), \dots\}},$$

c'est donc une intersection de fermés, et, dans le cas où X est compact, une intersection décroissante de compacts non vides. Finalement, si $y = \lim \varphi^{n_k}(x)$ est un point de $\omega(x)$, alors toute valeur d'adhérence de la suite $\varphi^{n_k-1}(x)$ est une préimage de y contenue dans $\omega(x)$. ►

Proposition 2.5. *Si f est une fonction de Lyapounov, chaque ensemble limite $\omega(x)$ est contenu dans un niveau totalement neutre de f.*

◄ Si $\omega(x)$ est vide, il n'y a rien à démontrer. Sinon, on considère un point $y \in \omega(x)$ qui est donc la limite d'une sous-suite $\varphi^{n_k}(x)$. La suite $f \circ \varphi^n(x)$ étant décroissante, elle tend vers $a := f(y)$. L'ensemble $\omega(x)$ est donc contenu dans $f^{-1}(a)$. Comme c'est un ensemble positivement invariant, ses points sont donc totalement neutres. ►

Revenons à la démonstration du théorème.

◄ On considère la fonction $f : U \longrightarrow [0, \infty)$ et le voisinage $V \subset U$ tel que $f > 0$ et $f - f \circ \varphi > 0$ sur $V - \{x_0\}$.

On considère un voisinage compact $K \subset V$ de x_0. On considère un voisinage ouvert W de x_0 tel que $\varphi(\bar{W}) \subset K$. On pose $a = \min_{K-W} f$, on a $a > 0$. On pose $Y = \{x \in \bar{W}, f(x) \leqslant a\}$, qui est un voisinage compact de x_0. L'ensemble Y est positivement invariant. En effet, si $x \in Y$, alors $\varphi(x) \in K$ et $f \circ \varphi(x) < f(x) = a$. Ceci implique que x est dans W, donc dans Y. En appliquant la proposition à $\varphi_{|Y}$, on conclut que $\omega(x) \subset \{x_0\}$ pour tout $x \in Y$. Comme Y est compact, on a donc $\omega(x) = \{x_0\}$, c'est à dire que toutes les orbites de Y tendent vers x_0. Si U' est un voisinage de x_0 contenu dans U, on construit comme ci-dessus un voisinage de x_0 positivement invariant Y' contenu dans U', ce qui montre la stabilité de Lyapounov. ►

Exercice 2.2. *Supposons qu'il existe une fonction de Lyapounov continue $f : X \longrightarrow [0, \infty)$ telle que $f(x_0) = 0$ et $f > 0$ sur $X - \{x_0\}$. Supposons de plus que f est propre (c'est à dire que les ensembles $f^{-1}([0, a])$ sont compacts) et stricte en dehors de x_0. Montrer que x_0 est asymptotiquement stable de bassin X entier.*

Exercice 2.3. *Supposons qu'il existe une fonction de Lyapounov continue $f : X \longrightarrow [0, \infty)$ telle que $f(x_0) = 0$ et $f(x) \geqslant h(d(x_0, x))$, où $h : [0, \infty) \longrightarrow [0, \infty)$ est une fonction continue strictement croissante. Supposons de plus qu'il existe une fonction continue $\phi :]0, \infty) \longrightarrow]0, \infty)$ telle que $f \circ \varphi \leqslant f - \phi \circ f$. Montrer que x_0 est asymptotiquement stable de bassin X entier.*

Le théorème admet une réciproque :

Théorème 2.6. *On suppose X localement compact. Soit x_0 un point fixe asymptotiquement stable, de bassin B. Alors il existe une fonction de Lyapounov continue $f : X \longrightarrow [0,1]$ qui a les propriétés suivantes : $f = 1$ sur $X - B$, $f \in]0,1[$ sur $B - \{x_0\}$, $f(x_0) = 0$, $f\circ\varphi - f < 0$ sur $B - \{x_0\}$.*

On commence par quelques lemmes :

Lemme 2.7. *On suppose X localement compact. Le point fixe x_0 est asymptotiquement stable si et seulement si il existe un voisinage U de x_0 qui converge uniformément vers x_0 au sens suivant : pour tout voisinage W de x_0 il existe N tel que $\varphi^n(U) \subset W$ pout tout $n \geqslant N$.*

◄ Supposons que U existe. Il est alors clair que toutes les orbites de U tendent vers x_0. Fixons maintenant un voisinage W de x_0. Il existe N tel que $\varphi^n(U) \subset W$ pour tout $n \geqslant N$. D'autre part, la continuité de φ implique l'existence d'un voisinage $V \subset U$ de x_0 tel que $\varphi^n(V) \subset W$ pour tout $n \leqslant N$. On a donc $\varphi^n(V) \subset W$ pour tout $n \geqslant 0$, donc x_0 est Lyapounov stable.

Réciproquement, supposons que x_0 est asymptotiquement stable. Soit B le bassin de x_0, et U un voisinage compact de x_0 contenu dans B. Soit W un voisinage de x_0, et V un voisinage ouvert de x_0 tel que $\varphi^n(V) \subset W$ pour tout $n \geqslant 0$. Pour tout voisinage ouvert W de x_0, les préimages $\varphi^{-n}(V), n \in \mathbb{N}$ recouvrent B, donc elles recouvrent U. Par compacité, il existe N tel que les préimages $\varphi^{-n}(V), n \leqslant N$ recouvrent U. Pour tout $x \in U$, il existe donc $k \leqslant N$ tel que $\varphi^(x) \in V$, et alors $\varphi^n(x) \in W$ pour tout $n \geqslant k$, en particulier pour tout $n \geqslant N$. ►

Lemme 2.8. *Soit U un ouvert. Supposons que $\overline{\varphi(U)} \subset U$ et notons $A := \cap_{n\geqslant 0}\varphi^n(U)$ et $B := \cup_{n\geqslant 0}\varphi^{-n}(U)$. Alors il existe une fonction de Lyapounov continue $f : X \longrightarrow [0,1]$ qui vérifie les propriétés suivantes : $f^{-1}(0) = A$, $f^{-1}(1) = X - B$, $f\circ\varphi - f < 0$ sur $B - A$.*

◄ Pour chaque n, on considère ne fonction continue $g_n : X \longrightarrow [0,1]$ telle que $g_n^{-1}(0) = \overline{\varphi^n(U)}$ et $g_n^{-1}(1) = X - U$. On constate que $g_n\circ\varphi^n \leqslant g_n$, avec une inégalité stricte sur $\varphi^{-n}(U) - \overline{\varphi^n(U)}$. En effet, si $x \in U$, alors $g_n\circ\varphi^n(x) = 0 \leqslant g_n(x)$, et cette inégalité est stricte si $x \in U - \overline{\varphi^n(U)}$. Si $x \notin U$, alors $g_n(x) = 1 \geqslant g_n\circ\varphi^n(x)$, et cette inégalité est stricte pour $x \in \varphi^{-n}(U)$.

Pour chaque n, on pose $f_n := (g_n + \ldots + g_n\circ\varphi^{n-1})/n$. La fonction $f_n : X \longrightarrow [0,1]$ est continue, elle vérifie $f_n\circ\varphi - f_n = (g_n\circ\varphi^n - g_n)/n \leqslant 0$, avec inégalité stricte sur $\varphi^{-n}(U) - \overline{\varphi^n(U)}$. De plus, $f_n(x) = 1$ si et seulement si $g_n(x) = g_n\circ\varphi(x) = \cdots = g_n\circ\varphi^n(x) = 1$, c'est à dire si et seulement si $x \in X - \varphi^{-n}(U)$. D'autre part, $f_n(x) = 0$ si et seulement si $g_n(x) = g_n\circ\varphi(x) = \cdots = g_n\circ\varphi^n(x) = 0$, c'est à dire si et seulement si $x \in \overline{\varphi^n(U)}$.

On considère maintenant une suite $a_n, n \geqslant 0$, à terme strictement positifs, et telle que $\sum a_n = 1$. On pose $f := \sum a_n f_n$. Comme $f(x)$ est une somme de termes positifs, $f(x) = 0$ si et seulement si $f_n(x) = 0$ pour tout n, c'est à dire si et seulement si $x \in A$. De la même façon, on a $f(x) \leqslant \sum a_n = 1$ sauf si $g_n(x) = 1$ pour tout n, c'est à dire si $x \notin \cup\varphi^{-n}(U)$. Finalement, $f\circ\varphi - f = \sum a_n(f_n\circ\varphi - f_n) \leqslant 0$. Cette inégalité est stricte si et seulement si l'un des termes est strictement négatif, c'est à dire pour $x \in B - A$. ►

Concluons la preuve du théorème :

◄ Soit x_0 un point fixe asymptotiquement stable. Il existe alors un voisinage relativement compact U qui converge uniformément vers x_0. En particulier, il existe N tel que $\overline{\varphi^N(U)} \subset U$, $\cap\varphi^{nN}(U) = \{x_0\}$, et $\cup\varphi^{-nN}(U) = B$, le bassin de x_0.

Il existe alors une fonction de Lyapounov continue $g : X \longrightarrow [0,1]$ pour φ^N, telle que $g^{-1}(0) = \{x_0\}$, $g\circ\varphi^N - g < 0$ sur $B - A$, $g^{-1}(1) = X - B$.

La fonction $f := (g + g\circ\varphi + \cdots + g\circ\varphi^{N-1})/N$ vérifie alors $f\circ\varphi - f = (g\circ\varphi^N - g)/N \leqslant 0$, avec inégalité stricte sur $B - A$, ainsi que les autres propriétés demandées. ►

Si X est une variété, la fonction de Lyapounov peut être choisie de classe C^∞. C'est une conséquence du résultat suivant :

Proposition 2.9. *Soit X une variété C^∞ de dimension finie et $\varphi : X \longrightarrow X$ une application continue. Soit $f : X \longrightarrow [0,1]$ une fonction de Lyapounov continue. Supposons que les ensembles $A := f^{-1}(0)$, et $B := X - f^{-1}(1)$ son positivement invariants. Supposons finalement que f est stricte sur $B - A$. Alors il existe une fonction de Lyapounov g, de classe C^∞, telle que $g^{-1}(1) = X - B$, $g^{-1}(0) = A$ et g est stricte sur $B - A$.*

◄ Comme $f\circ\varphi < f$ sur l'ouvert $B - A$, il existe une fonction $\tilde{g} : B - A \longrightarrow]0,1[$ de classe C^∞ et telle que $f\circ\varphi < \tilde{g} < f$. Ceci implique que $\tilde{g}\circ\varphi < f\circ\varphi < \tilde{g}$, donc $\tilde{g}$ est une fonction de Lyapounov stricte sur $B - A$.

On prolonge $\tilde{g}$ par les valeurs 1 sur $X - B$ et 0 sur A. On obtient une fonction de Lyapounov continue sur X, qui est stricte et C^∞ sur $B - A$.

Il existe alors une fonction $\phi : [0,1] \longrightarrow [0,1]$ continue, strictement croissante, lisse sur $]0,1[$, et telle que $g := \phi \circ \tilde{g}$ est C^∞ (il faut prendre ϕ très plate en 0 et en 1). Cette fonction remplit toutes les conditions demandées. ►

Dans le cas où $X = \mathbb{R}^d$ (ou plus généralement une variété) et où φ est C^1, on considère le linéarisé $L = d\varphi_{x_0}$.

On note R le *rayon spectral* de L, c'est à dire le maximum des modules de ses valeurs propres complexes (L est un endomorphisme réel).

Théorème 2.10. *Si $R < 1$, alors x_0 est asymptotiquement stable. On dit que le point fixe est linéairement stable. Si $R > 1$, alors x_0 n'est pas Lyapounov stable (donc pas asymptotiquement stable).*

Comme le montre la preuve ci-dessous, il suffit de supposer que φ est différentiable en 0 (pas necessairement C^1).

On rappelle la formule du rayon spectrale : $R = \lim \|L^n\|^{1/n}$.

Nous baserons la preuve du théorème sur le lemme suivant, qui est important en lui-même :

Lemme 2.11. *Soit $L : B \longrightarrow B$ un endomorphisme de l'espace de Banach $(B, |.|)$, soit R le rayon spectral de L et $b > R$. Il existe une norme $\|.\|$ de Banach sur B, équivalente à la norme $|.|$, pour laquelle $\|L\| \leqslant b$. Si la norme $|.|$ est Hilbertienne, alors la norme $\|.\|$ l'est aussi.*

Si L est un isomorphisme de Banach, si r est l'inverse du rayon spectral de L^{-1} (c'est donc l'infimum des modules des éléments du spectre de L), et si $0 < a < r$, alors il existe une norme $[.]$ pour laquelle

$$a[v] \leqslant [Av] \leqslant b[v]$$

pour tout $v \in B$.

◄ Il existe N tel que $\|A^N\| \leqslant b^N$ (par la formule du rayon spectral). On pose

$$\|v\| := \left(\sum_{i=0}^{N-1} b^{-2i} |A^i v|^2 \right)^{1/2},$$

de sorte que

$$\|Av\|^2 = \sum_{i=0}^{N-1} b^{-2i} |A^{i+1} v|^2 = b^2 \sum_{i=1}^{N} b^{-2i} |A^i v|^2 = b^2 (\|v\|^2 - |v|^2 + b^{-2N} |A^N v|^2) \leqslant b^2 \|v\|^2.$$

Dans le cas où L est inversible, on choisit M tel que $\|A^{-M}\| \leqslant a^{-M}$ et on pose

$$[v]^2 := \sum_{i=0}^{M-1} a^{2i} \|A^{-i} v\|^2.$$

Par un calcul identique au précédent, on obtient que $[A^{-1}v] \leqslant a^{-1}[v]$ pour tout v, donc que $a[v] \leqslant [Av]$. Par ailleurs,

$$[Av]^2 = \sum_{i=0}^{M-1} a^{2i} \|A \circ A^{-i} v\|^2 \leqslant \sum_{i=0}^{M-1} a^{2i} b^2 \|A^{-i} v\| = b^2 [v]^2. \blacktriangleright$$

Montrons maintenant le théorème :

On suppose que $R(L) < 1$, et on choisit $b \in]R, 1[$. Il existe une norme $\|.\|$ pour laquelle $\|L\| < b$. Il existe alors un petit voisinage convexe de x_0 sur lequel $\|d\varphi\| \leqslant b$. Sur ce voisinage, l'application φ est donc b-Lipschitzienne. En particulier, on a

$$\|\varphi(x) - x_0\| \leqslant b\|x - x_0\|$$

ce dont on déduit facilement les conclusions voulues.

Réciproquement, supposons que $R > 1$ (et que l'espace est de dimension finie). On a alors une décomposition de l'espace $\mathbb{R}^d = E \oplus F$, où E et F sont invariants par L, où toutes les valeurs propres de $L_{|E}$ sont de module R, et où le rayon spectral R_F de $L_{|F}$ est strictement inférieur à R (une telle décomposition n'existe pas forcément en dimension infinie). On fixe a et $b > 1$ tels que $R_F < b < a < R$. En appliquant le Lemme, on trouve une norme Euclidienne $\|.\|_E$ sur E telle que $\|Lv\|_E \geqslant a\|v\|_E$ pour tout v dans E et un norme Euclidienne $\|.\|_F$ sur F telle que $\|Lv\|_F \leqslant b\|v\|_F$ pour tout

$v \in F$. On pose alors $\|v\|^2 = \|v_E\|_E^2 + \|v_F\|_F^2$, où v_E et v_F sont les composantes de v suivant E et F. C'est une norme Euclidienne sur $\mathbb{R}^d$, pour laquelle E et F son orthogonaux, et qui coïncide avec $\|.\|_E$ sur E et $\|.\|_F$ sur F.

Pour tout $\delta > 0$, il existe $\epsilon > 0$ tel que les inégalités suivantes sont satisfaites pour $x \in B(0, \epsilon)$:

$$\|\varphi_E(x)\| \geqslant a\|x_E\| - \delta\|x\| \quad , \quad \|\varphi_F(x)\| \leqslant b\|x_F\| + \delta\|x\|.$$

En choisissant δ assez petit pour que $b + 2\delta < a - 2\delta$, nous allons en conclure que tout voisinage de 0 contient un point dont l'orbite sort de $B(0, \epsilon)$, contredisant la stabilité de Lyapounov. On prend une condition initiale $x = x_E + x_F$ telle que $\|x_E\| \geqslant \|x_F\|$. Ceci implique que $\|\varphi_E(x)\| \geqslant (a - 2\delta)\|x_E\|$ et $\|\varphi_F(x)\| \leqslant b\|x_F\| + 2\delta\|x_E\| \leqslant (b + 2\delta)\|x_E\|$. En notant $x^n = x_E^n + x_F^n$ l'orbite de x, on obtient donc par récurrence que, tant que $x^n \in B(0, \epsilon)$, $\|x_E^n\| \geqslant \|x_F^n\|$ et $\|x_E^n\| \geqslant (a - 2\delta)^n\|x_E\|$. Comme $a - 2\delta > 1$, l'orbite sort de $B(0, \epsilon)$. ►

On peut se poser les mêmes questions dans le cas où X est un espace de Banach de dimension infinie. On remarque que la preuve ci-dessus n'utilise jamais la finitude de la dimension pour la partie stabilité, qui reste donc vraie en dimension infinie. La partie instabilité est un peu plus délicate. Pour que la preuve fonctionne, il faut supposer un peu plus sur le linéarisé afin de pouvoir décomposer l'espace comme nous l'avons fait dans la preuve. Il suffit par exemple de supposer l'existence d'un réel $R' \in [1, R[$ tel que le spectre du linéarisé ne contient aucun point de module R'. Cette hypothèse supplémentaire n'est pas nécessaire si φ est supposée C^2, mais il faut une autre démonstration.

Revenons à la questions de la conjugaison topologique au voisinage du point fixe :

Proposition 2.12. *Soit $\varphi : \mathbb{R}^d \longrightarrow \mathbb{R}^d$ une application continue fixant l'origine. Supposons que $L = d\varphi_0$ existe, est inversible, et de rayon spectral strictement inférieur à* 1. *Alors φ est localement conjuguée à L. Plus précisément il existe deux voisinages ouverts U et V de* 0 *et un homéomorphisme $h : U \longrightarrow V$ tel que $h \circ L = \varphi \circ h$ sur U.*

◄ On choisit une norme Euclidienne $|.|$ pour laquelle L est une contraction. On écrit $\varphi = L + \phi$, où $\phi(x) = o(|x|)$. On considère une fonction de troncature $f : \mathbb{R}^d \longrightarrow [0, 1]$ qui est égale à 1 sur la boule $B(0, 1)$ et à 0 hors de la boule $B(0, 2)$. En notant $f_\epsilon(x) = f(x/\epsilon)$, on considère l'application $\psi := L + f_\epsilon \phi$. On va montrer que c'est, pour $\epsilon > 0$ petit, un homéomorphisme et une contraction, ce qui implique la conclusion par la Proposition 2.2. Il suffit pour ceci de montrer que $\mathrm{Lip}(f_\epsilon \phi)$ peut être rendu arbitrairement petit. Il existe une fonction $\delta(\epsilon)$ qui tend vers 0 en 0 et a les propriétés suivantes : ϕ est $\delta(\epsilon)$-Lipschitz sur $B(0, 2\epsilon)$, et $|\phi| \leqslant \epsilon\delta(\epsilon)$ sur $B(0, 2\epsilon)$. On a alors

$$\mathrm{Lip}(f_\epsilon \phi) \leqslant \delta(\epsilon) + \delta(\epsilon)\mathrm{Lip}(f),$$

cette constante de Lipschitz tend donc vers 0 lorsque ϵ tend vers 0. ►

Exercice 2.4. *Soit φ et ψ deux homéomorphismes de X et Y, respectivement. Soient x_0 et y_0 des points fixes asymptotiquement stables de φ et ψ. Supposons que φ au voisinage de x_0 est topologiquement conjugué (c'est à dire conjugué par un homéomorphisme) à ψ au voisinage de y_0. Si B et B' sont les bassins d'attraction de x_0 et y_0, montrer que $\varphi_{|B}$ est topologiquement conjugué à $\psi_{|B'}$.*

En particulier, les bassins B et B' sont homéomorphes. Le bassin d'un point fixe linéairement stable d'un difféomorphisme est donc homéomorphe à $\mathbb{R}^d$.

Sous des hypothèses supplémentaires, on obtient une conjugaison plus régulière.

Proposition 2.13. *Soit φ un difféomorphisme C^{k+2} ($k \in \{1, \dots, \infty\}$) d'une variété X, fixant x_0. Supposons que les modules des valeurs propres du linéarisé $L = d\varphi_{x_0}$ sont contenues dans l'intervalle $[r, R]$, avec $R^2 < r < R < 1$. Alors, en notant B le bassin de x_0, l'application $\varphi_{|B} : B \longrightarrow B$ est globalement conjuguée à son linéarisé $L : T_{x_0}X \longrightarrow T_{x_0}X$ par un difféomorphisme C^k.*

En dimension 1, l'hypothèse de pincement est toujours satisfaite. C'est aussi le cas en dimension 1 complexe. Si $\varphi : \mathbb{C} \longrightarrow \mathbb{C}$ est une application holomorphe fixant 0 et si $|\varphi'(0)| < 1$ alors φ est localement conjuguée à son linéarisé par une conjugaison holomorphe, cette copnjugaison s'étend à l'ensemble du bassin d'attraction.
◄ A TERMINER

On fixe a et b tels que $b^2 < a < r < R < b$. On choisit une norme Euclidienne sur $T_{x_0}X$ telle que $|L| < b$ et $|L^{-1}| < 1/a$. On identifie un voisinage ouvert U de x_0 dans X à un voisinage V de 0 dans $T_{x_0}M$. On pose $h_n = L^{-n} \circ \varphi^n$. On va montrer que, au voisinage de x_0, la suite h_n converge uniformément sur U vers une limite continue $h : U \longrightarrow V$, qui vérifie automatiquement $h \circ \varphi = L \circ h$. Au voisinage de 0, φ est b-Lipschitz, et il existe une constante C telle que $|\varphi^{-1}(x) - L^{-1}(x)| \leqslant C|x|^2$. On a donc $|L^{-1}(\varphi^{n+1}(x)) - \varphi^{-1}(\varphi^{n+1}(x))| \leqslant Cb^{2n}|x|^2$. En appliquant L^{-n}, on déduit que $|h_{n+1}(x) - h_n(x)| \leqslant C(b^2/a)^n|x|^2$. Comme $b^2/a < 1$, cette inégalité implique la convergence uniforme de la suite h_n dans un voisinage de x_0, vers une limite continue h qui vérifie de plus $dh_0 = id$. Il existe donc un voisinage ouvert U de x_0 dans X tel que h_n converge, uniformément sur U, vers $h : U \longrightarrow V$ qui est un homéomorphisme, où V est un voisinage ouvert de 0 dans $T_{x_0}X$.

En fait, la convergence uniforme de h_n sur un voisinage de x_0 implique que cette convergence a lieu sur tout B, uniformément sur les compacts. En effet, si K est un compact de B et si U est un voisinage de x_0 sur lequel la convergence uniforme a lieu, il existe N tel que $\varphi^N(K) \subset U$. Ceci implique que $h_n \circ \varphi^N$ converge uniformément sur K, il en est donc de même de $h_{n+N} = L^{-N} \circ h_n \circ \varphi^N$. La limite h vérifie $h = L^{-N} \circ h \circ \varphi^N$, c'est donc un homéomorphisme de K sur son image. Si K' est un compact de $T_{x_0}X$, alors il existe N tel que $L^N(K') \subset V$, donc $h^{-1}(K') = \varphi^{-N}(h^{-1}(L^N(K')))$ est compact. L'application $h : B \longrightarrow T_{x_0}M$ est une bijection continue et propre, c'est donc un homéomorphisme.

Si $k \geqslant 1$,

On considère maintenant l'application $T\varphi(x, v) := (\varphi(x), d\varphi_x \cdot v)$. Elle fixe le point $(0, 0)$ et sa différentielle est $TL(v, w) = (Lv, Lw)$. Les valeurs propres de TL sont les mêmes que celles de L. On montre donc comme ci-dessus que la suite $H_n = (TL)^{-n} \circ (T\varphi)^n$ converge vers une limite continue H sur $B \times T_{x_0}X$, uniformément sur les compacts. Comme $H_n = Th_n$, on conclut que h est C^1 sur B, avec $Th = H$. Si $k > 1$, on montre de la même façon par récurrence que h est C^k.

Comme $dh_0 = id$, on conclut alors que h est un difféomorphisme local en 0, c'est à dire qu'il existe un voisinage U de 0

Dans le cas $k > 1$, on termine la preuve par récurrence. ►

La méthode ci-dessus est due à Nelson [1]. Il l'utilise pour démontrer qu'il existe une conjugaison régulière, sans hypothèse de pincement sur les valeurs propres de L, mais en supposant que $|\varphi - L| = O(|x|^r)$ avec r assez grand, ce résultat étant originalement du à Sternberg.

La régularité des conjugaisons est une question assez subtile, comme l'illustrent les exercices ci-dessous :

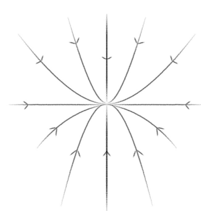
Exercice 2.5. *Considérons l'application $\varphi(x, y) = (x/2, ay + x^2)$ et son linéarisé $\phi(x, y) = (x/2, ay)$.*

Pour $a \neq 1/4$, montrer qu'il existe une conjugaison C^∞ entre φ et ϕ, qui est de la forme $h(x, y) = (x, y + bx^2)$.

Exercice 2.6. *Considérons l'application $\varphi(x, y) = (x/2, y/4 + x^2)$ et son linéarisé $\phi(x, y) = (x/2, y/4)$.*

Montrer qu'il existe une conjugaison entre φ et ϕ de la forme $h(x, y) = (x, y + ax^2 \ln x)$. Quelle est la régularité de h ?

Montrer qu'il n'existe pas de conjugaison C^2 au voisinage de 0.



1. E. Nelson, Topics in dynamics I : flows.

3 Points fixes asymptotiquement stables des champs de vecteurs

On considère maintenant le cas où X est une variété (ou juste $X = \mathbb{R}^d$) et où V est un champ de vecteurs C^1 et complet sur X. On note φ^t le flot. Les points fixes Lyapounov stables et asymptotiquement stables se définissent exactement comme dans le cas des applications :

Le point fixe x_0 est dit Lyapounov stable si, pour tout voisinage U de x_0, il existe un voisinage V de x_0 tel que $\varphi^t(V) \subset U$ pour tout $t \geqslant 0$.

Le point fixe x_0 est dit asymptotiquement stable si il est Lyapounov stable et si, de plus, il existe un voisinage W de x_0 tel que toute orbite partant dans W converge vers x_0.

Le bassin de x_0 est l'ensemble des points x tels que $\varphi^t(x) \longrightarrow x_0$ en $+\infty$. Si x_0 est un point fixe asymptotiquement stable, alors son bassin est un voisinage ouvert de x_0.

Propriété 3.1. *Le point fixe x_0 est Lyapounov (resp. asymptotiquement) stable pour V si et seulement si il existe $T > 0$ tel que x_0 est Lyapounov (resp. asymptotiquement) stable pour φ^T. De plus, le bassin de x_0 en tant que point fixe de φ^T est égal au bassin de x_0 en tant que point fixe de V.*

◄ Il est clair que les propriétés de stabilité pour V entreine les mêmes propriétés pour chacun des flots $\varphi^t, t > 0$.

Réciproquement, supposons que x_0 est Lyapounov stable pour φ^T, avec $T > 0$. Pour tout voisinage U de x_0, il existe un voisinage W tel que $\varphi^t(W) \subset U$ pour tout $t \in [0, T]$. Ceci découle de la compacité de $[0, T]$ et de la continuité de l'application $(t, x) \mapsto \varphi^t(x)$.

Comme x_0 est Lyaounov stable pour φ^T, il existe un voisinage W' tel que $\varphi^{nT}(W') \subset W$ pour tout $n \geqslant 0$. On déduit que $\varphi^t(W') \subset U$ pour tout $t \geqslant 0$.

Finalement, si l'orbite $x(t) = \varphi^t(x)$ converge vers x_0, il en est clairement de même de la suite $n \longmapsto x(nT)$. Réciproquement, supposons que la suite $x(nT)$ converge vers x_0. Pour tout voisinage U de x_0, on a vu qu'il existe un voisinage W pour lequel $\varphi^t(W) \subset U$ pour tout $t \in [0, T]$. Il existe N tel $x(nT) \in W$ pour tout $n \geqslant N$, ce qui implique que $x(t) \in U$ pour tout $t \geqslant NT$. ►

On manipulera dans la suite des inégalités différentielles, et on utilisera le résultat suivant (Lemme de Gronwall) :

Lemme 3.2. *Soit $f(t, x) : \mathbb{R} \times \mathbb{R} \longrightarrow \mathbb{R}$ une fonction mesurable. Soient $x(t)$ et $y(t) : [0, T[\longrightarrow \mathbb{R}$ des fonctions dérivables satisfaisant*

$$x'(t) < f(t, x(t)) \quad , \quad y'(t) = f(t, y(t))$$

pour tout $t \in [0, T[$ et $x(0) \leqslant y(0)$. Alors $x(t) < y(t)$ pour tout $t \in]0, T[$.

◄ Si la conclusion n'est pas satisfaite, alors il existe un intervalle maximal non vide $]0, s[, s < T$ sur lequel $x < y$ (dans le cas où $x(0) = y(0)$, ceci découle de l'inégalité $x'(0) < f(0, x(0)) = f(0, y(0)) = y'(0)$). Au temps s, on a alors $y(s) = x(s)$, et donc $y'(s) \leqslant x'(s)$. C'est une contradiction puisque

$$x'(s) < f(s, x(s)) = f(s, y(s)) = y'(s). \blacktriangleright$$

Nous utiliserons principalement le :

Corollaire 3.3. *Soit $x(t)$ une fonction dérivable telle que $x'(t) \leqslant bx(t)$ pour tout $t \geqslant 0$, avec $b \in \mathbb{R}$. Alors $x(t) \leqslant x(0)e^{bt}$ pour tout $t \geqslant 0$.*

◄ Pour $a > 0$, on pose $y_a(t) = (x(0) + at)e^{bt}$. On a $y_a'(t) = by_a(t) + ae^{bt}$. Le lemme précédent, appliqué à la fonction $f(t, x) = ax + ae^{bt} > ax$, implique que $x(t) < y_a(t)$ pour tout $t > 0$ et tout $a > 0$. À la limite $a \longrightarrow 0$ on déduit que $x(t) \leqslant x(0)e^{bt}$. ►

Plus généralement, on peut obtenir une version avec inégalités larges du Lemme sous des hypothèses de régularité supplémentaires :

Lemme 3.4. *Soit $f(t, x) : \mathbb{R} \times \mathbb{R} \longrightarrow \mathbb{R}$ une fonction localement Lipschitz. Soient $x(t)$ et $y(t) : [0, T[\longrightarrow \mathbb{R}$ des fonctions dérivables satisfaisant*

$$x'(t) \leqslant f(t, x(t)) \quad , \quad y'(t) = f(t, y(t))$$

pour tout $t \in [0, T[$ et $x(0) \leqslant y(0)$. Alors $x(t) \leqslant y(t)$ pour tout $t \in [0, T[$.

◄ On fixe $S \in]0, T[$. On considère la solution $y_a(t)$ de l'équation $y_a'(t) = f(t, y_a(t)) + a$ qui vérifie $y_a(0) = y(0)$. Par le théorème de Cauchy-Lipschitz, cette solution est définie au delà du temps S pour a petit, et la fonction $a \longmapsto y_a(S)$ est continue en $a = 0$. Pour chaque $a > 0$, on déduit du lemme précédent (appliqué à la fonction $f + a$) que $x(S) < y_a(S)$. À la limite $a \longrightarrow 0$, on déduit donc que $x(S) \leqslant y(S)$. ►

Pour donner un analogue des contractions en termes de champs de vecteurs, on commence par la remarque suivante :

Lemme 3.5. *Soit V un champ de vecteurs Lipschitz complet sur un espace de Hilbert E. Alors, pour tout $b \in \mathbb{R}$, il y a équivalence entre les propriétés suivantes :*
- *Le flot φ^t est e^{tb}-Lipschitz pour tout $t \geqslant 0$*
- *$\langle V(x) - V(y), x - y \rangle \leqslant b\|x - y\|^2$ pour tous x et y.*

◄ Un calcul simple montre que

$$\frac{d}{dt}\|\varphi^t(x) - \varphi^t(y)\|^2 = 2\langle V(\varphi^t(x)) - V(\varphi^t(y), \varphi^t(x) - \varphi^t(y)\rangle.$$

Si l'on suppose le premier point, alors $\|\varphi^t(x) - \varphi^t(y)\|^2 \leqslant e^{2tb}\|x - y\|^2$ pour tout $t \geqslant 0$, avec égalité en $t = 0$, donc

$$2\langle V(x) - V(y), x - y\rangle = \frac{d}{dt}_{|t=0}\|\varphi^t(x) - \varphi^t(y)\|^2 \leqslant 2b\|x - y\|^2.$$

Réciproquement, en supposant le second point, la fonction $f(t) := \|\varphi^t(x) - \varphi^t(y)\|^2$ vérifie l'inégalité $f'(t) \leqslant 2bf(t)$, ce qui entraine que $f(t) \leqslant f(0)e^{2bt}$ par le Lemme de Gronwall. ►

Proposition 3.6. *Soit V un champ de vecteur de l'espace de Hilbert E. Supposons qu'il existe $a > 0$ tel que*

$$\langle V(x) - V(y), x - y\rangle \leqslant -2a\|x - y\|^2$$

pour tous x, y. Alors Le champ V admet un unique point fixe, qui est asymptotiquement stable et attire toutes les orbites. Le champ V est positivement complet et ses flots $\varphi^t, t > 0$ sont des contractions.

◄ Soit $x(t) : [0, T[\longrightarrow E$ une orbite maximale. Montrons pour commencer que $T = +\infty$. Si T est fini, on écrit par exemple $\|x(S+t) - x(S)\| \leqslant e^{-aS}\|x(t) - x(0)\| \leqslant \|x(T-S) - x(0)\|$ pour tout $t \in [0, T-S[$. Comme la fonction $t \mapsto \|x(t) - x(0)\|$ tend vers 0 en 0, ceci implique que la courbe $x(t)$ a une limite en T, ce qui contredit la maximalité.

Le champ V est donc positivement complet, ses flots $\varphi^t, t > 0$ sont e^{-ta}-Lipschitz par les calculs ci-dessus. Soit x_0 l'unique point fixe de φ^1. C'est alors l'unique point fixe de $\varphi^{1/n}$ pour tout $n \in \mathbb{N}$. Autrement dit, l'orbite $x(t)$ associée est périodique de période $1/n$ pour tout n, c'est donc un point fixe : $V(x_0) = 0$. Il est alors évident que toutes les orbites convergent vers x_0. ►

Si f est une fonction de classe C^1, alors les points suivants sont évidemment équivalents :
- $f \circ \varphi^t \leqslant f$ pour tout $t \geqslant 0$.
- $df \cdot V \leqslant 0$.

On dit alors que f est une fonction de Lyapounov. Il n'est donc pas nécessaire de déterminer le flot pour vérifier que f est une fonction de Lyapounov.

Théorème 3.7. *Soit x_0 un point fixe du champ de vecteurs V.*

Si il existe une fonction de Lyapounov $f : U \longrightarrow [0, \infty)$, de classe C^1 sur un voisinage U de x_0, qui vérifie $f(x_0) = 0$ et telle que les fonctions f et $-df \cdot V$ sont strictement positives sur $U - \{x_0\}$, alors x_0 est asymptotiquement stable.

Réciproquement, si x_0 est asymptotiquement stable, de bassin B, alors il existe une fonction de Lyapounov $f : X \longrightarrow [0, 1]$, de classe C^∞, telle que $f(x_0) = 0$, $f \in]0, 1[$ sur $B - \{x_0\}$, $df \cdot V < 0$ sur $B - \{x_0\}$, et $f = 1$ sur $X - B$.

◄ Le premier énoncé découle du cas des applications.

Montrons le second énoncé. Le point x_0 est un point fixe asymptotiquement stable pour φ^1 dont le bassin est B. Il existe donc une fonction de Lyapounov $g : X \longrightarrow [0, 1]$, de classe C^∞, pour φ_1, telle que $g \circ \varphi^1 - g < 0$ sur $B - \{x_0\}$, $g^{-1}(0) = \{x_0\}$, $g^{-1}(1) = X - B$.

On pose alors $f = \int_0^1 g \circ \varphi^s ds$, c'est une fonction C^1. On a

$$\begin{aligned}(df \cdot V)(x) &= \int_0^1 dg_{\varphi^s(x)} \circ d\varphi^s_x \cdot V(x) ds = \int_0^1 dg_{\varphi^s(x)} \cdot V(\varphi^s(x)) ds = \int_0^1 \frac{d}{ds}(g \circ \varphi^s(x)) ds \\ &= (g \circ \varphi^1 - g)(x).\end{aligned}$$

Si V n'est que C^1, la fonction f n'est a priori que C^1, mais elle peut être régularisée, comme dans le cas des applications. Pour ceci, on considère d'abord une fonction continue strictement positive $\epsilon : B - A \longrightarrow]0, \infty)$ telle que $df \cdot V < -2\epsilon|V|$, $f > 2\epsilon$ et $f < 1 - 2\epsilon$ sur $B - A$. Alors, il existe une fonction f_1, de classe C^∞, telle que $|f - f_1| \leqslant \epsilon$ et $|df - df_1| \leqslant \epsilon$. La fonction f_1 vérifie encore $df_1 \cdot V < 0$ sur $B - A$, et elle se prolonge en une fonction continue sur X par les valeurs

0 sur A et 1 sur $X - B$. On régularise alors f_1 en posant $f_2 = \phi \circ f_1$, où $\phi : [0,1] \longrightarrow [0,1]$ est une fonction strictement croissante "assez plate" en 0 et en 1. ►

Comme dans le cas des applications, le linéarisé $L := dV_{x_0}$ donne des informations utiles sur la dynamique au voisinage du point fixe x_0. L'équation différentielle linéaire $v'(t) = L(V(t))$ est en effet une bonne approximation de l'équation $x'(t) = V(x(t))$ au voisinage du point x_0. Le flot de cette équation linéarisée est e^{tL}. Comme les valeurs propres de e^{tL} sont $e^{t\lambda}$, où λ est une valeur propre de L, on s'attend à un comportement asymptotiquement stable si $|e^{t\lambda}| < 1$ pour toute valeur propre λ de L, c'est à dire si toutes les valeurs propres de L ont des parties imaginaires strictement négatives.

Dans le cas où X est une variété, il n'est pas complètement évident que L est bien défini en tant qu'endomorphisme de $T_{x_0}X$. Si ϕ et ψ sont deux cartes de X en x_0, on peut considérer les représentants $V_\phi = \phi_* V$ et $V_\psi = \psi_* V$ de V dans ces cartes. On peut alors calculer les linéarisés dans les cartes, $L_\phi = d(V_\phi)_0$ et $L_\psi = d(V_\psi)_0$. Ce sont des endomorphismes de $\mathbb{R}^d$ et la question est de savoir si ils représentent le même endomorphisme de $T_{x_0}X$, c'est à dire si $L_\psi \circ df_0 = df_0 \circ L_\phi$, en notant $f = \psi \circ \phi^{-1}$. On a $V_\psi = f_* V_\phi$, c'est à dire que $V_\psi(f(x)) = df_x \cdot V_\phi(x)$. On différentie en $x = 0$, ce qui donne

$$L_\psi \circ df_0 = df_0 \cdot L_\phi + d^2 f_0 \cdot V_\phi(0).$$

On a donc bien l'égalité voulue si (et seulement si) $V_\phi(0) = 0$, c'est à dire si $V(x_0) = 0$. Il résulte des considération ci-dessus que le linéarisé L d'un champ de vecteur est bien défini en tant qu'endomorphisme de $T_{x_0}X$ lorsque x_0 est un point fixe de V.

On peut aussi définir ce linéarisé L sans passer par les cartes en utilisant le flot $\varphi^t(x)$, qui est défini pour tout t dans un voisinage de x_0. Comme $\varphi^t(x_0) = x_0$, on peut définir $G_t := d\varphi^t_{x_0}$. C'est un groupe d'isomorphismes de $T_{x_0}X$, c'est à dire que $G^0 = Id$ et $G^{t+s} = G^t \circ G^s$. Il existe donc un endomorphisme L de $T_{x_0}X$ tel que $G^t = e^{tL}$, c'est le linéarisé de V en 0. Au delà de ces questions de définitions, on retiendra l'expression

$$d(\varphi^t)_{x_0} = e^{tL}$$

qui relie le linéarisé du champ de vecteur à la différentielle de son flot.

Théorème 3.8. *Soit x_0 un point fixe du champ de vecteur V. Soit L le linéarisé de V en x_0, et soit χ le supremum des parties réelles des éléments du spectre de L.*

Si $\chi < 0$, alors x_0 est asymptotiquement stable. On dit que le point fixe est linéairement stable.

Si $\chi > 0$, alors x_0 n'est pas Lyapounov stable, donc pas asymptotiquement stable.

◄ Nous avons vu que le champ de vecteurs est Lyapounov ou asymptotiquement stable si et seulement si le flot φ^1 l'est. Comme $d(\varphi^1)_{x_0} = e^L$, le rayon spectral du linéarisé de φ^1 est égal à $R = e^\chi$. ►

Comme dans le cas des applications, il est utile d'introduire des normes adaptées à la dynamique.

Lemme 3.9. *Soit L une application linéaire continue de l'espace de Hilbert H. Supposons qu'il existe un intervalle $[a', b']$ qui contient les parties réelles des éléments du spectre de L, et soit $]a, b[$ un intervalle ouvert contenant $[a', b']$. Alors il existe un produit scalaire $(.,.)$ sur H, qui engendre une norme équivalente à la norme initiale, et tel que*

$$a(v,v) \leqslant (Lv, v) \leqslant b(v,v)$$

pour tout $v \in H$.

◄ Le rayon spectral de e^L est strictement inférieur à e^b donc par la formule du rayon spectral il existe $T > 0$ pour lequel $\|e^{TL}\| < e^{Tb}$ (norme initiale sur E). On pose

$$[v, w] = \int_0^T e^{-2tb} \langle e^{tL} v, e^{tL} w \rangle dt$$

et on calcule

$$\begin{aligned} 2[Lv, v] &= 2\int_0^T e^{-2tb} \langle e^{tL} Lv, e^{tL} v \rangle dt = \int_0^T e^{-2tb} \frac{d}{dt} \langle e^{tL} v, e^{tL} v \rangle dt \\ &= 2b \int_0^T e^{-2tb} \langle e^{tL} v, e^{tL} v \rangle dt + e^{-2Tb} \langle e^{TL} v, e^{TL} v \rangle - \langle v, v \rangle \\ &\leqslant 2b[v, v]. \end{aligned}$$

Ensuite, on choisit $S > 0$ tel que $|e^{-SL}| < e^{-Sa}$ (norme associée à $[.,.]$) et on pose

$$(v,w) = \int_0^S e^{2ta}[e^{-tL}v, e^{-tL}w]dt.$$

Le même calcul que ci-dessus montre que $(-Lv,v) \leqslant -a(v,v)$. De plus,

$$(Lv,v) = \int_0^S e^{2ta}[Le^{-tL}v, e^{-tL}v]dt \leqslant \int_0^S be^{2ta}[e^{-tL}v, e^{-tL}v]dt = b(v,v). \blacktriangleright$$

On peut montrer comme dans le cas d'une application que la dynamique au voisinage d'un point fixe asymptotiquement stable est topologiquement conjuguée à la dynamique du linéarisé. En fait, le résultat prend une forme plus forte dans le cas des flots : il n'y a, à conjugaison topologique près, qu'un seul point fixe linéairement stable en dimension d.

Théorème 3.10. *Soient V_1 et V_2 deux champs de vecteurs C^1 sur des variétés X_1 et X_2 de dimension d. Soient x_1 et x_2 des points fixes asymptotiquement stables de V_1 et V_2, de bassins U_1 et U_2. Il existe un homéomorphisme $\phi : U_1 \longrightarrow U_2$ qui conjugue les flots de V_1 et V_2. L'homéomorphisme ϕ est C^1 sur $U_1^* := U_1 - \{x_1\}$, et satisfait $d\phi_x \cdot V_1(x) = V_2(\phi(x))$ pour tout $x \in U_1^*$.*

Le résultat implique en particulier que le bassin est toujours homéomorphe à $\mathbb{R}^d$, qui est la bassin de 0 pour le champ de vecteurs $V_0(x) = -x$ sur $\mathbb{R}^d$.
◀ On va montrer que V_1 est conjugué à V_0 sur U_1.

On identifie localement (X_1, x_1) à $(\mathbb{R}^d, 0)$ par une carte en x_1, et on munit l'espace $\mathbb{R}^d$ d'un produit scalaire pour lequel $\langle V_1(x), x\rangle \leqslant b\langle x, x\rangle$ au voisinage de 0, avec $b < 0$, on note $|.|$ la norme correspondante. Pour $r > 0$ assez petit, la sphère $S = S(0,r)$ (que l'on peut aussi considérer comme une sous variété de X_1 par l'identification ci-dessus) est une section de Poincaré pour V_1 et pour V_0. On note aussi $B = B(0,r)$. Considérons l'application

$$F : \mathbb{R} \times S \ni (t,x) \mapsto \varphi^t(x) \in U_1^* = U_1 - \{x_1\}$$

où φ est le flot de V_1. Cette application induit une bijection de $\mathbb{R} \times S$ dans U_1^*, qui est un difféomorphisme local et donc un difféomorphisme.

Montrons la bijectivité. Il s'agit de vérifier que toute orbite $x(t)$ dans le bassin coupe S une fois et une seule. On constate d'abord que tout orbite qui contient des points dans B et des points hors de B coupe S, par connexité. Comme toute orbite tend vers x_1, toute orbite du bassin contient des points dans B. Donc les seules orbites qui pourraient ne pas couper S sont les orbites entièrement contenues dans B. On utilise finalement l'inégalité $\langle V_1(x), x\rangle \leqslant b\langle x, x\rangle$, satisfaite dans un voisinage de $\bar{B}$, pour montrer, à la fois que toute orbite différente de celle de x_1 sort de B (en grand temps négatif), et aussi que chaque orbite intersecte S au plus une fois.

De la même façon, l'application

$$G : \mathbb{R} \times S \ni (t,x) \mapsto e^{-t}x \in \mathbb{R}^d - \{0\}$$

est un difféomorphisme. L'application

$$F \circ G^{-1} : \mathbb{R}^d - \{0\} \ni x \mapsto \varphi^{-\ln|x|}(x/|x|) \in U_1^*$$

est donc un difféomorphisme. Comme 0 est asymptotiquement stable pour V_1, l'application ϕ qui est égale à $F \circ G^{-1}$ hors de 0 et à 0 en 0 est un homéomorphisme de $\mathbb{R}^d$ dans U_1. On constate que

$$\phi(e^{-t}x) = \varphi^{t-\ln|x|}(x/|x|) = \varphi^t(\phi(x)),$$

c'est à dire que ϕ conjugue les flots de V_1 et de V_0.

On trouve de la même façon une conjugaison topologique $\psi : \mathbb{R}^d \longrightarrow U_2$ entre V_0 et V_2, et donc une conjugaison topologique $\phi \circ \psi^{-1} : U_2 \longrightarrow U_1$ entre V_2 et V_1. ▶

L'analogue naturel de cet énoncé pour les applications n'est pas vrai comme le montre l'exercice ci-dessous.

Exercice 3.1. *Montrer que les dynamiques des applications $x \mapsto x/2$ et $x \mapsto -x/2$ sur $\mathbb{R}$ ne sont pas topologiquement conjuguées.*

Exercice 3.2. *Considérons le champ de vecteurs $V(x,y) = (y, -x - y^3)$.*

Le linéarisé permet-il de statuer sur la stabilité asymptotique de 0 ?

En considérant la fonction $f(x,y) = x^2 + y^2$, montrer que le point fixe 0 est asymptotiquement stable.

Exercice 3.3. *Soit* A *une matrice* $d \times d$*. Montrer l'équivalence entre :*

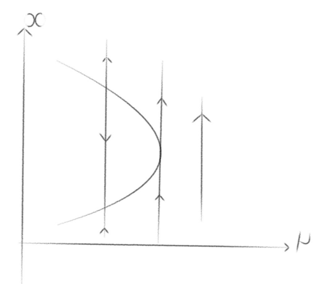
Toutes les solutions de l'équation linéaire $x'(t) = A \cdot x(t)$ *sont bornées.*

La matrice A *est diagonalisable dans* $\mathbb{C}$ *et ses valeurs propres sont imaginaires pures.*

Il existe un produit scalaire pour lequel A *est antisymétrique.*

4 Bifurcations

Nous avons vu qu'un point fixe non dégénéré survit à une petite perturbation de la dynamique. Nous étudions ici ce qui peut se passer dans le cas des points fixes dégénérés. Un exemple est la famille d'applications $\varphi(\mu, x) = \mu + x + x^2$, dont le comportement dynamique est résumé dans le diagramme ci-dessous.



Ce comportement est typique.

Théorème 4.1. *Soit $\varphi(\mu, x) : \mathbb{R} \times \mathbb{R} \longrightarrow \mathbb{R}$ est une famille C^2 d'applications de $\mathbb{R}$ telle que $\varphi(0,0) = 0$, $\partial_x \varphi(0,0) = 1$, $\partial_\mu \varphi(0,0) \neq 0$. Alors il existe un intervalle ouvert I contenant 0 et une fonction $\theta(x)$ telle que $\varphi(\theta(x), x) = x$ pour tout $x \in I$, $\theta(0) = 0$, et $\theta'(0) = 0$. De plus, pour $\mu \in I$ et $x \in I$, le point x est fixé par φ_μ si et seulement si $\theta(x) = \mu$.*

Si on suppose que $\partial^2_{xx}\varphi(0,0)$ est non nul et du même signe que $\partial_\mu \varphi(0,0)$, alors $\theta''(0) < 0$. Alors, pour $\mu < 0$ dans I, l'application φ_μ a deux points fixes dans I (l'un linéairement stable et l'autre linéairement instable), pour $\mu = 0$ elle a un point fixe, et pour $\mu > 0$ elle n'a pas de point fixe.

Si le signe de $\partial^2_{xx}\varphi(0,0)$ est opposé à celui de $\partial_\mu \varphi(0,0)$ alors $\theta''(0) > 0$ et on a les mêmes propriétés en échangeant $\mu > 0$ et $\mu < 0$.

◄ Le premier point est une application directe du théorème des fonctions implicites. En dérivant la relation $\varphi(\theta(x), x) = x$, on obtient

$$\partial_\mu \varphi(\theta(x), x) \cdot \theta'(x) + \partial_x \varphi(\theta(x), x) = 1.$$

En $x = 0$, ceci implique que $\theta'(0) = 0$. On dérive alors une seconde fois, en $x = 0$, et on obtient :

$$\partial_\mu \varphi(\theta(x), x) \cdot \theta''(x) + \partial^2_{xx}\varphi(\theta(x), x) = 0,$$

dont on déduit les propriétés voulues sur le signe de $\theta''(0)$. Pour déterminer la stabilité linéaire des points fixes obtenus, on doit déterminer le multiplicateur $\lambda(x) := \partial_x \varphi(\theta(x), x)$. On calcule que $\lambda'(0) = \partial^2_{xx}\varphi(0,0)$. Lorsque cette dérivée est non-nulle, on conclut que l'un des points fixes est linéairement stable, l'autre linéairement instable. ►

Dans de nombreuses applications, on considère une famille φ_μ dont toutes les applications préservent l'origine, c'est à dire que $\varphi(\mu, 0) = 0$.

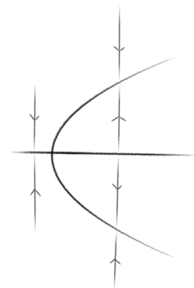
Théorème 4.2. *Soit $\varphi(\mu, x) : \mathbb{R} \times \mathbb{R} \longrightarrow \mathbb{R}$ est une famille C^3 d'applications de $\mathbb{R}$ telle que $\varphi(\mu, 0) = 0$ pour tout μ et $\partial_x \varphi(0,0) = 1$, $\partial^2_{\mu x}\varphi(0,0) \neq 0$.*

Le diagramme de bifurcation au voisinage de $(0,0)$ est localement la réunion de deux courbes C^2 qui se coupent transversalement en $(0,0)$ et dont l'une est l'axe horizontal.

Si $\partial^2_{xx}\varphi(0,0) \neq 0$, alors la seconde branche du diagramme de bifurcation est le graphe d'une application $\mu \longmapsto y(\mu)$. Il existe donc, pour $\mu \neq 0$ assez petit, exactement deux points fixes dans un voisinage de 0.

Si $\partial^2_{xx}\varphi(0,0) = 0$ alors la seconde branche est tangente à l'axe vertical en $(0,0)$. Dans ce cas, si $\partial^3_{xxx}\varphi(0,0)$ est non nul et du signe de $\partial^2_{\mu x}\varphi(0,0)$, alors la seconde branche contient deux orbites pour chaque $\mu < 0$ et aucune pour $\mu > 0$.

Suivant les signes des dérivées, on peut aussi déterminer la stabilité des points fixes obtenus. Par exemple, si $\partial^2_{\mu x}\varphi(0,0) > 0$ et $\partial^3_{xxx}\varphi(0,0) < 0$, alors le point fixe 0 est linéairement stable pour $\mu < 0$ et linéairement instable pour $\mu > 0$. Les points fixes de la seconde branche sont linéairement stables. On a donc, pour $\mu < 0$ un unique point fixe qui est linéairement stable. Pour $\mu > 0$, ce point fixe devient linéairement instable, mais deux points fixes linéairement stables nouveaux apparaissent. C'est la situation représentée ci-dessous.



◄ On a $\varphi(\mu,x) = x\psi(\mu,x)$, où

$$\psi(\mu,x) = \int_0^1 \partial_x\varphi(\mu,sx)ds$$

est une fonction C^2 qui vérifie $\psi(0,0) = 1$. L'équation $\varphi(\mu,x) = x$ est équivalente à l'équation $\psi(\mu,x) = 1$. On vérifie en dérivant l'égalité $\varphi = x\psi$ que $k\partial_x^k\psi(0,0) = \partial_x^k\varphi(0,0)$ pour tout $k \geqslant 0$, et que $\partial_\mu\psi(0,0) = \partial^2_{\mu x}\varphi(0,0)$. La preuve se termine alors comme la précédente. ►

Nous avons décrit des bifurcations pour des applications d'une variable réelle. Nous allons maintenant expliquer pourquoi les bifurcation les plus simples en dimension supérieure présentent des diagrammes de bifurcation similaires.

On considère pour ceci une famille d'applications φ_μ de $\mathbb{R}^d$, de classe C^r, et telle que $\varphi_0(0) = 0$. Soit $L = \partial_x\varphi(0,0)$ le linéarisé de φ_0 en 0. Soit E l'espace caractéristique de L associé à la valeur propre 1. Soit F la somme des autres espaces caractéristiques de L. C'est un supplémentaire de E dans $\mathbb{R}^d$ qui est invariant par L. L'énoncé ci-dessous permet de réduire l'étude des bifurcations de la famille φ_μ à l'étude des bifurcations d'une famille réduite ψ_μ d'applications de E de classe C^r. On note $x = (y,z)$ la décomposition des points de $\mathbb{R}^d$ associée à la décomposition $\mathbb{R}^d = E \oplus F$. On note φ_E, φ_F les composantes de φ.

Proposition 4.3. *Il existe une application locale $Z : \mathbb{R}\times E \longrightarrow F$, de classe C^r, qui vérifie $\partial_y Z(0,0) = 0$ et qui est telle que le diagramme de bifurcation de φ est l'image par $id \times Z$ du diagramme de bifurcation de la famille*

$$\psi(\mu,y) := \varphi_E(\mu,y,Z(\mu,y)).$$

Autrement dit, localement, le point $x = (y,z)$ est un point fixe de φ_μ si et seulement si $z = Z(\mu,y)$ et si y est un point fixe de ψ_μ.

Cette méthode est appelée réduction de Lyapounov-Schmidt. Elle permet de construire, localement, une sous variété tangente à $\mathbb{R}\times E$ dans $\mathbb{R}\times\mathbb{R}^d$ qui contient tous les points fixes. Attention, cette variété n'est pas forcément invariante par φ.

◄ L'équation $\varphi(u,x) = x$ est équivalente au système

$$\begin{cases} \varphi_E(\mu,y,z) = y \\ \varphi_F(\mu,y,z) = z \end{cases}.$$

Comme $\partial_z\varphi_F(0,0,0)$ n'a pas 1 pour valeur propre, la seconde équation se résout localement par le théorème des fonctions implicites. Il existe donc une fonction $Z(\mu,y)$ telle que la seconde équation est satisfaite si et seulement si $z = Z(\mu,y)$. De plus, comme $\partial_y\varphi_F(0,0,0) = 0$, on a $\partial_y Z(0,0) = 0$. Les solutions du système sont alors les points de la forme $(\mu,y,Z(\mu,y))$, où (μ,y) satisfait l'équation $\varphi_E(\mu,y,Z(\mu,y) = y$, c'est à dire où y est un point fixe de ψ_μ. ►

Les cas les plus simples sont ceux où E est de dimension 1.

Proposition 4.4. *Supposons que φ est une famille C^2 et que l'espace propre E est de dimension 1. Si $\partial_\mu\varphi_E(0,0)$ et $\partial^2_{yy}\varphi_E(0,0)$ sont non nuls et de même signe, alors le diagramme de bifurcation est localement une courbe C^2 de $\mathbb{R}\times\mathbb{R}^d$, tangente à $\{0\}\times E$, qui contient deux points pour chaque $\mu < 0$ et aucun point pour $\mu > 0$.*

◄ On applique le théorème 4.1 à l'application réduite ψ. On calcule pour ceci

$$\partial_\mu\psi(0,0) = \partial_\mu\varphi_E(0,0) + \partial_z\varphi_E(0,0)\circ\partial_y Z(0,0) = \partial_\mu\varphi_E(0,0)$$

et

$$\partial^2_{yy}\psi(0,0) = \partial^2_{yy}\varphi_E(0,0) + 2\partial^2_{yz}\varphi_E(0,0)\circ\partial_y Z(0,0) + \partial_z\varphi_E(0,0)\partial^2_{yy}Z(0,0) = \partial^2_{yy}\varphi_E(0,0)$$

puisque $\partial_z\varphi_E(0,0) = 0$. ►

On considère maintenant le cas des familles fixant l'origine, c'est à dire que $\varphi(\mu,0) = 0$.

Proposition 4.5. *Supposons que φ est une famille C^3 qui fixe l'origine, et que l'espace propre E est de dimension 1. Si $\partial^2_{\mu y}\varphi_E(0,0)$ est non-nul, alors le diagramme de bifurcation est localement la réunion de la courbe $\mathbb{R}\times\{0\}$ et d'un seconde branche C^2.*

Si $\partial^2_{yy}\varphi_E(0,0)\neq 0$, alors la seconde branche du diagramme de bifurcation est le graphe d'une application $\mu\longmapsto Y(\mu)$ de classe C^2. Il existe donc, pour $\mu\neq 0$ assez petit, exactement deux points fixes dans un voisinage de 0.

Si $\partial^2_{yy}\varphi_E(0,0)=0$, alors la seconde branche du diagramme de bifurcation est tangente à $\{0\}\times E$.

On peut caractériser plus précisément la forme de la seconde branche comme dans le théorème 4.2 en fonction de la quantité $\partial^3_{yyy}\psi(0,0)$. Mais l'expression de cette dérivée troisième en fonction de φ n'est pas simple.

◄ On applique le théorème 4.2 à l'application réduite ψ, après avoir vérifié que $\partial^2_{x\mu}\psi(0,0)=\partial^2_{x\mu}\varphi_E(0,0)$. ►

On rencontre par exemple le cas ci-dessus lorsqu'on considère un point fixe non dégénéré qui admet -1 comme valeur propre simple. Soit φ_μ une famille d'applications de $\mathbb{R}^d$ telle que $\varphi_0(0)=0$, telle que 1 n'est pas valeur propre du linéarisé, et telle que -1 est valeur propre simple. Comme le point fixe 0 est alors non-dégénéré pour φ_0, il existe localement une unique courbe régulière $q(\mu)$ de points fixes. Mais 0 est un point fixe dégénéré pour φ_0^2, donc il peut exister d'autres points de période 2 au voisinage de $(0,0)$. On note $L(\mu)$ le linéarisé de φ_μ en $q(\mu)$, E la droite propre de $L(0)$ associée à la valeur propre -1, et F le supplémentaire de E invariant par $L(0)$. On a alors

Proposition 4.6. *Soit $\varphi(\mu,x):\mathbb{R}\times\mathbb{R}\longrightarrow\mathbb{R}$ est une famille C^3 d'applications de $\mathbb{R}^d$ telle que $\varphi(0,0)=0$ et -1 est valeur propre de $L(0)=\partial_x\varphi(0,0)$, de multiplicité 1. On suppose que $\partial_{\mu|\mu=0}(\partial_y\varphi_E(\mu,q(\mu))\neq 0$. Alors il existe, localement, une branche de points périodiques de période 2, qui est tangente à $\{0\}\times E$.*

◄ Soit $q(\mu)$ la famille régulière locale de points fixes de φ_μ. Pour se ramener au cadre du résultat précédent, on considère la famille

$$\phi(\mu,x)=\varphi(\mu,\varphi(\mu,x+q(\mu))-q(\mu).$$

Toutes les applications de cette famille fixent 0, mais 0 est un point fixe dégénéré de ϕ_0. On a

$$\partial_x\phi(\mu,x)=\partial_x\varphi(\mu,\varphi(\mu,x+q(\mu)))\circ\partial_x\varphi(\mu,x+q(\mu)),$$

en particulier $\partial_x\phi(\mu,0)=\partial_x\varphi(\mu,q(\mu))\circ\partial_x\varphi(\mu,q(\mu))$. Le linéarisé $\partial_x\phi(0,0)$ a donc 1 pour valeur propre simple, d'espace propre E, et il préserve le supplémentaire F. On doit montrer que $\partial^2_{\mu y}\phi_E(0,0)$ est non nul. En notant π_E la projection sur E, le réel $\partial_y\phi_E(\mu,0)$ est le coefficient $\pi_E\circ L(\mu)^2_{|E}$. On a donc

$$\partial^2_{\mu y}\phi_E(0,0)=\pi_E\circ L(0)\circ L'(0)_{|E}+\pi_E\circ L'(0)\circ L(0)_{|E}.$$

Les hypothèses sur $L(0)$ impliquent que $\pi_E\circ L(0)=-\pi_E$, et que $L(0)_{|E}=-Id$. On obtient donc

$$\partial^2_{\mu y}\phi_E(0,0)=-2\pi_E\circ L'(0)_{|E}=-2\partial_{\mu|\mu=0}(\partial_y\varphi_E(\mu,q(\mu))\neq 0.$$

On a donc une seconde branche de points fixes de ϕ_μ, qui sont des points de période 2 pour φ_μ. Pour chaque valeur de μ, il y a un nombre pair de tel points, et on est donc dans le second cas de la proposition précédente, avec une branche tangente à $\{0\}\times E$. ►

Il est intéressant de préciser un peu ce qui précède en dimension 1.

Exercice 4.1. *Soit $\varphi(\mu,x)$ une famille C^3 d'applications de $\mathbb{R}$, telle que $\varphi(0,0)=0$ et $\partial_x\varphi(0,0)=-1$.*

On suppose, pour fixer les idées, que $\partial^2_{\mu x}(0,0)>0$ et que

$$2\partial^3_{xxx}\varphi(0,0)+3(\partial^2_{xx}f(0,0))^2>0.$$

Il existe alors un branche régulière $q(\mu)$ de points fixes. Pour $\mu<0$, le point fixe $q(\mu)$ est asymptotiquement stable et il est localement le seul point périodique.

Pour $\mu>0$, le point fixe $q(\mu)$ est linéairement instable, et il existe une orbite périodique de période 2, qui est linéairement stable.

On parle de bifurcation de doublement de période, car l'élément stable est un point de période 1 pour $\mu<0$ et un point de période 2 pour $\mu>0$.

Le résultat ci-dessus montre que la condition de non dégénérescence, qui interdit l'existence de bifurcations à période fixée, ne suffit pas pour autant à garantir l'absence de bifurcations d'orbites périodiques lorsqu'on considère simultanément toutes les périodes. Ce type de phénomène peut se produire dès qu'une valeur propre est une racine complexe de l'unité, ou simplement est de module 1. Le cas des racines non réelles est un peu plus délicat à étudier, car il ne se produit pas en dimension 1. On l'illustre en considérant des familles holomorphes.

Proposition 4.7. *Soit $\varphi_\mu(x) = \varphi(\mu, x) : \mathbb{C} \times \mathbb{C} \longrightarrow \mathbb{C}$ une famille holomorphe d'applications de $\mathbb{C}$ fixant l'origine. Supposons que $\lambda = \partial_z \varphi(0,0)$ est de module 1, et supposons que $\partial^2_{\mu z}\varphi(0,0) \neq 0$. Il existe alors, pour tout $\epsilon > 0$ une valeur μ du paramètre telle que $|\mu| \leqslant \epsilon$ et pour laquelle l'application φ_μ admet une orbite périodique contenue dans la boule $B(0, \epsilon)$.*

◄ On considère dans un premier temps le cas où λ est une racine de l'unité, $\lambda^k = 1$, $k \geqslant 1$. On cherche alors des orbites de périodes k. On écrit pour ceci $\varphi^k_\mu = z f(\mu, z)$, avec $f(0,0) = \lambda^k = 1$. On a $f(\mu, 0) = \partial_z(\varphi^k_\mu(0)) = (\partial_z \varphi(\mu, 0))^k$, et donc

$$\partial_\mu f(0,0) = k\lambda^{k-1}\partial^2_{\mu z}\varphi(0,0) \neq 0.$$

L'équation $f(\mu, z) = 0$ peut donc être résolue par le théorème des fonctions implicites : il existe une fonction holomorphe $u(z)$ telle que $f(\mu, z) = 1$ si et seulement si $\mu = u(z)$. Il existe donc des couples (μ, z) arbitrairement proches de $(0,0)$ pour lesquels $\varphi^k_\mu(z) = z$.

Supposons maintenant que λ n'est pas une racine de l'unité. L'application $\mu \longmapsto \partial_z \varphi_\mu(\mu, 0)$ est un difféomorphisme local. Il existe donc μ_1 arbitrairement petit pour lequel $\lambda_1 := \partial_z \varphi(\mu_1, 0)$ est une racine de l'unité et $\partial_{\mu z}\varphi(\mu_1, 0) \neq 0$. On peut appliquer ce qui précède pour trouver des (μ_2, z_2) arbitrairement proches de $(\mu_1, 0)$ pour lesquels $\varphi^k(\mu_2, z_2) = z_2$ (où k est tel que $\lambda_1^k = 1$). On notera, dans ce second cas, que la période k doit être choisie d'autant plus grande que l'on veut une orbite périodique plus proche de l'origine. ►

On comprend au vu des développements ci-dessus que toute valeur propre de module 1 crée la possibilité de phénomènes de bifurcations d'orbites périodiques (à période non fixée).

On dit qu'un point fixe est hyperbolique si le linéarisé n'a pas de valeur propre de module 1.

Théorème 4.8. *Soit $\varphi : \mathbb{R}^d \longrightarrow \mathbb{R}^d$ une application de classe C^1 dont 0 est un point fixe hyperbolique.*

Alors il existe un voisinage U de 0 tel que toute orbite $(x_n), n \in \mathbb{Z}$ contenue dans U est identiquement nulle. En particulier, aucune autre orbite périodique n'est entièrement contenue dans U.

Finalement, si $\varphi(\mu, x)$ est une famille C^1 d'applications telle que $\varphi_0 = \varphi$, alors pour tout μ petit, l'ouvert U contient une unique orbite entière de φ_μ, et cette orbite est un point fixe hyperbolique.

◄ Considérons directement une famille φ_μ. Comme 0 est non dégénéré, il existe une courbe $q(\mu)$ de points fixes de φ_μ.

L'hyperbolicité implique qu'il existe deux sous-espace E^s et E^u, invariants par $L(0)$, et tels que $L_{|E^s}$ et $L^{-1}_{|E^u}$ ont un rayon rayon spectral < 1. Il existe donc des constantes $a < 1$ et une norme Euclidienne sur E telle que $|L^{-1}_{|E^u}| < a$ et $|L_{|E^s}| < a$. Décomposons la matrice $L(\mu, x) := \partial_x \varphi(\mu, x)$ par bloc en

$$L(\mu, x) = \begin{bmatrix} L_{ss}(\mu, x) & L_{su}(\mu, x) \\ L_{us}(\mu, x) & L_{uu}(\mu, x). \end{bmatrix}$$

Fixons $\delta \in]0, 1-a[$, ce qui implique que $a^{-1} - \delta > 1$. Il existe $\epsilon > 0$ et un voisinage convexe U de 0 dans $\mathbb{R}^d$ tel que

$$|L_{ss}(\mu, x)| < a \quad , \quad |L_{uu}(\mu, x)^{-1}| < a \quad , \quad |L_{su}(\mu, x)| < \delta \quad , \quad |L_{us}(\mu, x)| < \delta \tag{1}$$

pour tout $(\mu, x) \in]-\epsilon, \epsilon[\times U$. On peut supposer de plus en diminuant ϵ que $q(\mu) \in U$ pour tout $\mu \in]-\epsilon, \epsilon[$. Ceci implique que

$$|\varphi^u_\mu(x) - q^u(\mu)| \geqslant a^{-1}|x_u - q^u(\mu)| - \delta|x_s - q^s(\mu)| \quad , \quad |\varphi^s_\mu(x) - q^s(\mu)| \leqslant a|x_s - q^s(\mu)| + \delta|x_u - q^u(\mu)|$$

Si $x = (x_s, x_u)$ est une condition initiale dans U telle que $|x_u - q_u(\mu)| \geqslant |x_s - q_u(\mu)|$ alors $|\varphi^u_\mu(x) - q_u(\mu)| \geqslant |\varphi^s_\mu(x) - q_s(\mu)|$ et

$$|\varphi^u_\mu(x) - q_u(\mu)| \geqslant (a^{-1} - \delta)|x_u - q_u(\mu)|.$$

Comme $a^{-1} - \delta > 1$, ceci implique, si $x \neq q(\mu)$, que l'orbite de x par φ_μ sort de U dans le futur. On montre de la même façon que toute orbite dont la condition initiale $x \neq q(\mu)$ dans U satisfaisant $|x_u - q_u(\mu)| \leqslant |x_s - q_u(\mu)|$ sort de U dans le passé. En conséquence, la seule orbite entièrement contenue dans U est le point fixe $q(\mu)$. ►

L'énoncé ci-dessus et sa preuve s'étendent sans modification au cas où $\mathbb{R}^d$ est remplacé par un espace de Banach, à condition de donner la définition suivante d'une application linéaire hyperbolique :

On dit que $L : B \longrightarrow B$ est hyperbolique si elle est continue et si il existe un décomposition $B = B^s \oplus B^u$ en deux sous-espaces fermés invariants par L tels que $|L^{-1}_{|B^u}| < a$ et $|L_{|B^s}| < a$, avec $a < 1$. Il est en fait équivalent de demander que le spectre de L soit disjoint du cercle unité.

Exercice 4.2. *Soit L un isomorphisme de B. Supposons qu'il existe une décomposition $B = E^s \oplus E^u$, une norme sur B, et $a > 1$ tels que :*

Pour tout $v = v_s + v_u$ vérifiant $|v_s| \leqslant |v_u|$, on a $|L_s v| \leqslant |L_u v|$ et $|Lv| \geqslant a|v|$;

Pour tout $v = v_s + v_u$ vérifiant $|v_u| \leqslant |v_s|$, on a $|L_u^{-1} v| \leqslant |L_s^{-1} v|$ et $|L^{-1} v| \geqslant a|v|$.

Montrer que L est hyperbolique.

On pourra montrer qu'il existe un application linéaire $A : E^s \longrightarrow E^u$, de norme $\leqslant 1$, dont le graphe est invariant par L.

5 Rotations quasi périodiques

On considère le tore $\mathbb{T}^d = \mathbb{R}^d/\mathbb{Z}^d$. C'est un groupe de Lie Abélien, c'est à dire à la fois une variété et un groupe commutatif, et les opérations de groupe sont différentiables. Les fonctions différentiables sur $\mathbb{T}^d$ s'identifient aux fonctions différentiables de $\mathbb{R}^d$ qui sont $\mathbb{Z}^d$-périodiques.

On le munit $\mathbb{T}^d$ de la distance $d(x, x') = \min_{y,y'} |y' - y|$ où le minimum est pris sur l'ensemble des représentants y et y' de x et x' dans $\mathbb{R}^d$. C'est un espace métrique compact. La distance d est invariante par translations, c'est à dire que $d(x + v, x' + v) = d(x, x')\ \forall v \in \mathbb{T}^d$.

On étudie ici la dynamique d'une translation $\varphi : x \mapsto x + \omega$ où ω est vu soit comme un élément de $\mathbb{R}^d$ soit comme un élément de $\mathbb{T}^d$ (l'application φ ne dépend que de l'image de ω dans $\mathbb{T}^d$).

Commençons par motiver cette étude en expliquant comment cette dynamique apparaît dans les systèmes linéaires en présence de valeurs propres de module 1. Soit $R(\omega)$ la matrice 2×2 de la rotation d'angle $2\pi\omega$. Soit L La matrice $2d \times 2d$ diagonalisable par blocs 2×2, et dont les blocs diagonaux sont $R(\omega_1), \ldots, R(\omega_d)$. Les valeurs propres complexes de L sont donc les nombres complexes $e^{2i\pi\omega_i}$. Notons (x_i, y_i) les coordonnées de $\mathbb{R}^{2d}$ correspondant à la décomposition par blocs. Les fonctions $n_i : x \mapsto (x_i^2 + y_i^2)$ sont invariante par la dynamique de L, c'est à dire que $n_i(Lx) = n_i(x)$. Pour $r_1, \ldots, r_d > 0$, le tore $T := \{x \in \mathbb{R}^{2d} : n_i(x) = r_i^2\ \forall \mathrm{i}\}$ est donc une sous-variété invariante. On parle de tore car cette sous-variété est difféomorphe à $\mathbb{T}^d$, un difféomorphisme est donné par l'application

$$J : \mathbb{T}^d \ni \theta = (\theta_1, \ldots, \theta_d) \mapsto (r_1 e^{2i\pi\theta_1}, \ldots, r_d e^{2i\pi\theta_d}) \in \mathbb{C}^d = \mathbb{R}^{2d}.$$

En restriction à T, la dynamique de L est conjuguée à celle de la translation φ. Plus précisément, on a

$$L\circ J = \varphi \circ J.$$

Nous allons maintenant décrire la dynamique en fonction du vecteur ω. La première remarque, indépendante de ω, est que les orbites sont des translatées les unes des autres. Elles ont donc toutes les mêmes propriétés.

Premier cas, ω rationnel. On dit que ω est rationnel si $\omega \in \mathbb{Q}^d$.

Proposition 5.1. *Si ω est rationnel, alors toutes les orbites sont périodique. Leur période minimale T est le plus petit entier pour lequel $T\omega \in \mathbb{Z}^d$.*

Second cas, ω non résonant. On dit que ω est non résonant si il n'existe aucun vecteur entier non-nul $k \in \mathbb{Z}^d$ tel que $k \cdot \omega \in \mathbb{Z}$.

Théorème 5.2. *Si ω est non résonant, alors toutes les orbites de φ sont denses dans $\mathbb{T}^d$. Réciproquement, si il existe une orbite dense, alors ω est non résonant.*

On utilise quelques lemmes algébriques :

Lemme 5.3. *Soit G un sous-groupe discret non trivial de $\mathbb{R}^d$. Il existe une forme linéaire ℓ telle que $\ell(G) = \mathbb{Z}$.*

◄ On considère une famille $\mathbb{R}^d$-libre maximale $h_1, \ldots h_r$ dans G, $r \geqslant 1$. Le groupe G est alors contenu dans l'espace vectoriel engendré par les h_i. On considère une forme linéaire l sur $\mathbb{R}^d$ telle que $l(h_1) = 1$ et $l(h_i) = 0$ pour $i > 1$. L'image $l(G)$ est un sou-groupe de $\mathbb{R}$. Soit elle est engendré par un réel a, et alors l/a vérifie $l(G) = \mathbb{Z}$, soit elle est dense.

Excluons ce second cas. Si l'image est dense, alors il existe une suite g_n de G tels que $x_1(n) := l_1(g_n)$ est une suite de réels non nuls qui tend vers 0. On écrit alors $g_n = x_1(n)h_1 + \cdots + x_r(n)h_t$. En notant $y_i(n)$ la partie fractionnaire de $x_i(n)$, on remarque que

$$g_n' := x_1(n)h_1 + y_2(n)h_2 + \cdots + y_r(n)h_r$$

appartient à G. La suite g_n' est contenue dans la boule fermée B de centre 0 de rayon $|h_1|+\ldots+|h_r|$, et $l(g_n') = l(g_n) = x_1(n)$. Comme B est compacte, $G \cap B$ est fini, donc $l(g_n') = x_1(n)$ prend ses valeurs parmi un nombre fini de possibilités, ce qui est en contradiction avec le fait que $x_1(n)$ est une suite de réels non nuls qui tend vers 0. ►

Lemme 5.4. *Soit G un sous-groupe fermé de $\mathbb{R}^d$. Si $G \neq \{0\}$ et $G \neq \mathbb{R}^d$, alors il existe une forme linéaire ℓ telle que $\ell(G) = \mathbb{Z}$.*

◄ La somme de deux sous-espaces vectoriels de $\mathbb{R}^d$ contenus dans G est contenue dans G. On note E la somme de tous les sous-espaces vectoriels de $\mathbb{R}^d$ contenus dans G, c'est un sous-espace vectoriel de $\mathbb{R}^d$ contenu dans G. Si $G \neq \mathbb{R}^d$, alors $E \neq \mathbb{R}^d$. On considère alors un supplémentaire F de E et l'intersection $G \cap F$, qui est un sous-groupe fermé de F. Ce sous-groupe ne contient aucun sous-espace vectoriel de F.

Montrons que G' est discret dans F. Sinon, il existe une suite $g_n \in G'$ telle que $g_n \longrightarrow 0$. On peut supposer en prenant une sous-suite que les directions $g_n/|g_n|$ convergent vers un vecteur unité $v \in F$. Montrons que la droite $\mathbb{R}v$ est alors contenue dans G, ce qui est une contradiction. Comme $|g_n| \longrightarrow 0$, pour tout $t \in \mathbb{R}$, il existe une suite a_n d'entiers tels que $a_n|g_n| \longrightarrow t$. Alors, on a $a_n g_n \longrightarrow tv$. Comme G est fermé, ceci implique que $tv \in G$.

Nous avons montré que G' est discret dans F. Par le lemme ci-dessous, il existe alors une forme linéaire non nulle ℓ' sur F telle que $\ell'(G') = \mathbb{Z}$. On prolonge ℓ' en une forme linéaire ℓ sur $\mathbb{R}^d$ nulle sur E. ►

◄ Terminons la preuve du théorème. On considère la fermeture O de l'orbite de 0. C'est un sous-groupe fermé de $\mathbb{T}^d$. On considère la préimage G de O dans $\mathbb{R}^d$. C'est un sous-groupe fermé qui contient ω et qui contient $\mathbb{Z}^d$. Si $O \neq \mathbb{T}^d$, alors $G \neq \mathbb{R}^d$. Il existe donc une forme linéaire ℓ telle que $\ell(G) = \mathbb{Z}$. Comme $\ell(\mathbb{Z}^d) \subset \mathbb{Z}$, la forme linéaire ℓ est à coefficients entiers. Comme $\omega \in G$, on a $\ell(\omega) \in \mathbb{Z}$, ce qui est une relation de résonance pour ω, et donc une contradiction si ω est non-résonant.

La réciproque est beaucoup plus facile. Si $k \in \mathbb{Z}^d$ est une relation de résonance, alors k engendre une application différentiable $f : \theta \longrightarrow k \cdot \theta \mod 1$ de $\mathbb{T}^n$ dans $\mathbb{T}$, qui est une submersion surjective. L'orbite du point x_0 est contenue dans la fibre $f^{-1}(f(\theta_0))$, qui est une sous variété fermée de $\mathbb{T}^d$ de codimension 1. ►

Nous avons fait tout le travail pour comprendre la structure des sous-groupes fermés de $\mathbb{R}^d$.

Théorème 5.5. *Soit G un sous-groupe fermé de $\mathbb{R}^d$. Alors il existe un sous-espace vectoriel E de $\mathbb{R}^d$ et des éléments $g_1, \dots, g_k$ de G linéairement indépendants modulo E, tels que G est engendré par E et les g_i.*

◄ On fait la preuve par récurrence sur d. Ou bien $G = 0$, ou bien $G = \mathbb{R}^d$, ou bien il existe une forme linéaire ℓ telle que $\ell(G) = \mathbb{Z}$. On applique l'hypothèse de récurrence au sous groupe $G \cap \ker \ell$ et on ajoute un point g tel que $\ell(g) = 1$. ►

Nous allons maintenant donner une autre preuve du théorème 5.2, qui nous conduira à une propriété plus forte. Il sera utile de considérer sur $\mathbb{T}^d$ des structures supplémentaires. Tout d'abord, nous munissons le tore d'une tribu, qui peut être définie comme la tribu borélienne, ou comme l'ensemble des parties B de $\mathbb{T}^d$ dont la préimage dans $\mathbb{R}^d$ est borélienne. On note qu'alors la bijection $\pi : [0,1[^d \longrightarrow \mathbb{T}^d$ est bi-mesurable entre les tribus boréliennes. On munit $\mathbb{T}^d$ d'une mesure de probabilité λ, que nous appellerons mesure de Lebesgue (ou mesure de Haar), et qui est l'unique mesure de probabilité Borélienne invariante par translation. C'est aussi l'image directe par la projection de la mesure de Lebesque sur $[0,1[^d$ (attention, ce n'est pas l'image directe par la projection de la mesure de Lebesgue sur $\mathbb{R}^d$).

On rappelle que, si $f : X \longrightarrow Y$ est une application mesurable entre espaces mesurés, l'image directe d'une mesure μ sur X est la mesure $f_*\mu$ sur Y telle que $f_*\mu(B) = \mu(f^{-1}(B))$. (en langage probabiliste, c'est la loi de la variable aléatoire f). En termes d'intégrales, ceci s'écrit

$$\int_Y g \, d(f_*\mu) = \int_X g \circ f \, d\mu$$

pour toute fonction mesurable positive.

L'invariance par translation de la mesure λ est la propriété $\varphi_*\lambda = \lambda$, qui est satisfaite par toutes les translations. On notera dans la suite $S^n f$ la somme $f + f \circ \varphi + \ldots + f \circ \varphi^{n-1}$.

Théorème 5.6 (Hermann Weyl)**.** *Soit φ une translation de vecteur non résonant sur le tore. Pour toute fonction continue f sur $\mathbb{T}^d$, on a, en topologie uniforme,*

$$\frac{1}{n} S^n f = \frac{1}{n}\left(f + f \circ \varphi + \cdots + f \circ \varphi^{n-1}\right) \longrightarrow \int_{\mathbb{T}^d} f \, d\lambda$$

Ce résultat implique la densité des orbites. En effet, si l'orbite de 0 n'était pas dense, il existerait une fonction f positive continue nulle sur l'orbite, mais non nulle. On a alors $f(0) + f \circ \varphi(0) + \cdots + f \circ \varphi^{n-1}(0) = 0$, et comme $\int f d\lambda > 0$, c'est une contradiction au vu du théorème.

◄ On commence par démontrer le résultat pour $f = e^{2i\pi k \cdot \theta}$, $k \in \mathbb{Z}^d$. Si $k \cdot \omega = 0$, alors $f \equiv 1$, donc le résultat est évident. Si $k \cdot \omega \neq 0$, on calcule :

$$f(\theta) + \cdots + f \circ \varphi^{n-1}(\theta) = e^{2i\pi\theta}\left(1 + e^{2i\pi k \cdot \omega} + \cdots + e^{(n-1)2i\pi k \cdot \omega}\right) = e^{2i\pi\theta} \frac{1 - e^{n2i\pi k \cdot \omega}}{1 - e^{2i\pi k \cdot \omega}}.$$

Cette quantité est bornée, donc $\left(f + f \circ \varphi + \cdots + f \circ \varphi^{n-1}\right)/n$ converge uniformément vers 0, qui est la moyenne de f.

Par linéarité, le résultat s'étend à tous les polynômes trigonométriques. En rappelant que les polynômes trigonométriques sont denses, pour la topologie uniforme, dans les fonctions continues, on peut alors étendre le résultat à toute fonction continue. En effet, soit f une fonction continue sur $\mathbb{T}^d$ et soit $\epsilon > 0$. Il existe alors g, polynôme trigonométrique tel que $|f - g| < \epsilon/10$. Il existe alors N tel que $|S^n g/n - \int g| < \epsilon/10$ pour tout $n \geqslant N$. Ceci implique que

$$|S^n f/n - \int f| \leqslant |S^n f/n - S^n g/n| + \left|S^n g/n - \int g\right| + \left|\int f - \int g\right| \leqslant 3\epsilon/10 < \epsilon. ►$$

Le corollaire suivant montre que cet énoncé donne une information plus quantitative que la densité des orbites :

Corollaire 5.7 (Hermann Weyl)**.** *Soit A une partie de $\mathbb{T}^d$ tel que $\lambda(\partial A) = 0$, et soit $N_A(n, \theta)$ le nombre de points parmi $x, \varphi(\theta), \dots, \varphi^{n-1}(\theta)$ qui sont dans A. Alors $N_A(n, \theta)/n \longrightarrow \lambda(A)$ pour tout $\theta \in \mathbb{T}^d$.*

◄ C'est exactement l'énoncé précédent pour la fonction $f = 1_A(x)$. Comme cette fonction n'est pas continue, nous devons justifier l'affirmation. Par régularité de la mesure λ, il existe, pour tout $\epsilon > 0$, un compact $K \subset A$ et un ouvert $U \supset A$ tels que

$$\lambda(U) - \epsilon \leqslant \lambda(\bar{A}) = \lambda(A) = \lambda(\mathring{A}) \leqslant \lambda(K) + \epsilon.$$

On considère alors des fonctions continues h et $g : \mathbb{T}^d \longrightarrow [0, 1]$ qui vérifient $h = 1$ sur $\bar{A}$ et 0 hors de U, et $g = 1$ sur K et 0 hors de $\mathring{A}$, de sorte que $g \leqslant 1_A \leqslant h$.

On a donc $S^n g \leqslant S^n_f \leqslant S^n_h$, donc

$$\lambda(A) - \epsilon \leqslant \int g = \lim S^n g(\theta)/n \leqslant \liminf S^n f(\theta)/n \leqslant \limsup S^n f(\theta)/n \leqslant \lim S^n h(\theta)/n \int h \leqslant \lambda(A) + \epsilon.$$

Comme ceci est vrai pour tout $\epsilon > 0$, on conclut que $\limsup S^n f(\theta)/n = \liminf S^n f(\theta)/n = \lambda(A)$ pour tout θ. ►

Cas général. Si $d = 1$, alors tout vecteur de rotation $\omega \in \mathbb{R}$ est soit rationnel soit non-résonant. Ce n'est toutefois pas le cas en dimension supérieure.

On note $Z(\omega)$ le sous-groupe de résonance de ω, c'est à dire le sous-groupe de $\mathbb{Z}^d$ constitué des vecteurs ξ tels que $\xi \cdot \omega \in \mathbb{Z}$. On note $E(\omega)$ le sous-espace vectoriel de $\mathbb{R}^d$ engendré par $Z(\omega)$, et $r(\omega)$ sa dimension, c'est l'ordre de résonance de ω.

Propriété 5.8. *ω est non résonant si et seulement si $r(\omega) = 0$.*

ω est rationnel si et seulement si $r(\omega) = d$.

◄ Le premier point est tautologique. Concernant le second, si $r(\omega) = d$ alors il existe $l_1, \dots, l_d \in \mathbb{Z}^d$, linéairement indépendants, tels que $l_i(\omega) \in \mathbb{Z}$. Il existe donc une matrice A inversible, à coefficients entiers, telle que $A\omega \in \mathbb{Z}^d$. La matrice A^{-1} est à coefficients rationnels, donc ω est est rationnel. ►

Théorème 5.9. *Considérons l'application $\varphi : \mathbb{T}^d \ni \theta \longmapsto \theta + \omega \in \mathbb{T}^d$ et soit r l'ordre de résonance de ω. Chaque orbite a pour adhérence une union finie de tores de dimension $d - r(\omega)$.*

On peut illustrer cet énoncé en prenant $\omega = (1/2, a)$, avec a irrationnel. L'orbite de zéro est alors dense dans la sous variété $\{0\} \times \mathbb{T} \cup \{1/2\} \times \mathbb{T}$.

Pour démontrer, et préciser, ce théorème, nous allons considérer des changements de variables $\theta \longmapsto A\theta$, avec $A \in Gl_d(\mathbb{Z})$. On rappelle qu'une matrice A à coefficients entiers engendre une application différentiable de $\mathbb{T}^d$. Cette application est un difféomorphisme si et seulement si A admet une inverse à coefficients entiers, c'est à dire si et seulement si $\det A = \pm 1$. On appelle automorphisme du tore ces difféomorphismes. L'automorphisme du tore associé à la matrice $A \in Gl_d(\mathbb{Z})$ est encore noté A, il conjugue la translation de vecteur ω à la translation de vecteur $A\omega$. Le théorème précédent découle donc du résultat algébrique suivant :

Théorème 5.10. *Soit $\omega \in \mathbb{T}^d$. Il existe $A \in Gl_d(\mathbb{Z})$ tel que $A\omega = (\omega_p, \omega_n)$, où $\omega_p \in \mathbb{T}^r$ est rationnel et $\omega_n \in \mathbb{T}^{d-r}$ est non résonant.*

Soit $F : \mathbb{T}^d \longrightarrow \mathbb{T}^r$ l'application donnée par les r premières composantes de TA, où T est une période de ω_p. L'application F est un morphisme de groupe différentiable et une submersion. Les adhérences des orbites de φ dans $\mathbb{T}^d$ sont les fibres de F (les préimages des points).

On a besoin de :

Proposition 5.11. *Pour tout entier k, le groupe $Gl_d(\mathbb{Z})$ agit transitivement sur les sous-espaces de dimension k de $\mathbb{Q}^d$.*

◄ Soit F un tel sous-espace et soit L une matrice rationnelle $d \times k$ dont F est l'image. En multipliant au besoin L par le produit de tous les dénominateurs de ses coefficients, on peut supposer que L est à coefficients entiers.

Nous allons montrer que l'on peut réduire L à une matrice $M = \begin{bmatrix} D & 0 \end{bmatrix}$ qui est la juxtaposition d'un bloc D diagonal $k \times k$ et du bloc nul par opérations élémentaires entières sur ses lignes et ses colonnes (c'est à dire par multiplication à gauche et à droite par des matrices de $Gl_d(\mathbb{Z})$ et $Gl_k(\mathbb{Z})$.

On réduit la matrice en appliquant successivement la suite de deux opérations suivantes : Mettre le plus petit (en valeur absolue) coefficient a non nul en haut à droite, puis remplacer les autres coefficients de la premières ligne et de la première colonne par les restes de leur division modulo a.

Chaque étape de cette construction fait strictement diminuer au moins un des coefficients non diagonaux, donc on se ramène en un nombre fini d'étapes à une matrice dont tous les coefficients de la permière ligne et de la première colonne sont nuls, sauf le premier. On recommence en prenant pour pivot le coeff $(2,2)$, etc ...

En fait, on peut, en plus, obtenir que les coefficients diagonaux $d_1, \dots, d_k$ de M sont des entiers tels que d_i divise d_{i+1} mais ce n'est pas nécessaire ici. Voire par exemple Artin, Algebra, section 12.4.

On a montré l'existence de deux matrices $A \in Gl_d(\mathbb{Z})$ et $B \in Gl_k(\mathbb{Z})$ telles que $ALB = M$. Ceci implique que $A(F) = \mathbb{Q}^k \times \{0_{d-k}\}$. ►

◄ On prouve maintenant le théorème. Soit $F(\omega)$ l'orthogonal de $Z(\omega)$ dans $\mathbb{Q}^d$, c'est à dire l'ensemble des vecteurs rationnels x tels que $l \cdot x = 0$ pour tout $l \in Z(\omega)$. Il existe un automorphisme $A \in Gl_d(\mathbb{Z})$ tel que $A(F(\omega)) = \{0_r\} \times \mathbb{Q}^{d-r}$.

On décompose $w := A\omega$ en deux composantes $w_r \in \mathbb{R}^r$ et $w_n \in \mathbb{R}^{d-r}$. Le groupe de résonances de w est l'ensemble des $l \circ A$, $l \in Z(\omega)$, c'est à dire que $Z(w) = A^t(Z(\omega)))$. Il est donc contenu dans $\mathbb{Z}^r \times \{0\}$. La première composante w_r est donc r résonante, donc rationnelle. La seconde composante w_n n'est pas résonante. ►

Temps continu. On termine en considérant le champ de vecteurs constant $V(\theta) = \omega$ sur $\mathbb{T}^d$, qui engendre le flot $\varphi^t(\theta) = \theta + \omega$. De tels flots apparaissent notamment dans les équations linéaires $\dot{x} = Lx$ en présence de valeurs propres imaginaires pures de L.

Les résultats sont très similaires à ceux du temps discret.

On définit le groupe de résonance $Y(\omega) = \{k \in \mathbb{Z}^d \,:\, k \cdot \omega = 0\}$, et l'ordre de résonance $q(\omega)$ comme la dimension de l'espace vectoriel engendré par $Y(\omega)$.

Théorème 5.12. *Les orbites sont denses dans des tores de dimension $d - q$.*

Il existe un automorphisme A du tore pour lequel $A\omega = (0, \omega_n) \in \mathbb{T}^q \times \mathbb{T}^{d-q}$ avec ω_n non résonant (c'est à dire $Y(\omega_n) = 0$).

La preuve est une variante simplifiée de celle du cas discret.

Mentionnons un dernier contexte dans lequel les tores quasi-périodiques apparaissent : la mécanique céleste.

Dans un système composé de k planètes et d'un soleil de très grande masse, on peut en première approximation supposer que chacune des planètes suit un mouvement périodique elliptique déterminé par son interaction avec le soleil. On peut paramétrer cette orbite par un flot de translations sur le cercle $\mathbb{T}^1$, au vu de l'exercice ci-dessous.

Exercice 5.1. *Soit $V(x)$ un champ de vecteurs sur $\mathbb{T}^1$ qui admet une orbite périodique non constante de période minimale T. Montrer que le champ V est globalement conjugué au champ de vecteurs constant $W(x) = 1/T$.*

On a donc, dans cette approximation du problème planétaire, des tores invariants de dimension n (le nombre de planètes) sur lesquels la dynamique est donnée par le flot d'un champ de vecteurs constant dont les coordonnées sont les inverses des périodes des planètes (qui dépendent des orbites choisies, c'est à dire du tore choisi). L'espace des phases, qui est ici l'ensemble des positions et vitesses possibles des planètes, est un espace de dimension $6n$, qui se décompose comme un union de tores de dimension n (avec des singularités et des dégénérescences correspondant aux collisions) sur chacun desquels la dynamique est un flot quasi périodique. Contrairement au cas linéaire, le vecteur fréquence ω dépend du tore, il y a des tores résonants et des tores non résonants.

6 Diverses propriétés d'irréductibilité en dynamique topologique

Les propriétés que nous avons démontrées pour les translations non résonantes ont un sens pour toute application continue d'un espace métrique compact (et parfois non compact) X.

L'application continue $\varphi : X \longrightarrow X$ est dite minimale si il n'existe aucun sous-ensemble fermé positivement invariant ($\varphi(X) \subset X$) non trivial. C'est équivalent à la densité de chaque orbite $O^+(x) := \{\varphi^n(x), n \geqslant 0\}$.

Un compact positivement invariant $Y \subset X$ est dit minimal si $\varphi_{|Y}$ est minimale. Il est alors invariant, c'est à dire que $\varphi(Y) = Y$.

Les points fixes et les orbites périodiques sont des compacts minimaux, ainsi que les tore quasi-périodiques non résonants.

Définition 6.1. *L'application $\varphi : X \longrightarrow X$ est dite uniquement ergodique si, pour chaque fonction f, il existe une constante $c(f)$ telle que $S^n f/n \longrightarrow c(f)$ ponctuellement, où $S^n f = f + f\circ\varphi + \cdots + f\circ\varphi^{n-1}$.*

Dans les exemples, on obtient même une convergence uniforme. On verra que c'est en fait équivalent : si φ est uniquement ergodique au sens ci-dessus, alors la convergence est uniforme.

Si φ est uniquement ergodique, alors il la fonction $f \longmapsto c(f)$ est linéaire et positive sur $C^0(X)$, c'est à dire que $c(f) \geqslant 0$ pour $f \geqslant 0$. Comme de plus $c(1) = 1$, on conclut par le théorème de représentation de Riesz qu'il existe une mesure de probabilité Borélienne m sur X telle que $c(f) = \int f dm$. La mesure de probabilité m est alors invariante par m. En effet

$$\int (f\circ\varphi)dm = \lim S^n(f\circ\varphi)(x)/n = \lim \left(S^n f(x) + f\circ\varphi^n(x) - f(x)\right)/n = \lim S^n f(x)/n = \int f dm$$

pour toute fonction continue f et tout point x, donc $\varphi_* m = m$.

Le théorème d'équirépartition de Weyl est alors valable relativement à la mesure m, c'est à dire que

$$N_A(n, x)/n \longrightarrow m(A)$$

pour tout x lorsque $n \longrightarrow \infty$, pourvu que $m(\partial A) = 0$. Plus généralement, la relation

$$S^n f(x)/n \longrightarrow \int f dm$$

est satisfaite, ponctuellement, pour toute fonction intégrable au sens de Riemann, c'est à dire pour toute fonction dont l'ensemble des points de discontinuité est de mesure nulle (pour la mesure m). Ceci découle de l'exercice suivant.

Exercice 6.1. *Pour toute fonction f, on note $I^-(f)$ le supremum des inégrales des fonctions continues inférieures à f, et $I^+(f)$ l'infimum des intégrales des fonctions continues supérieures à f.*

1. *En utilisant la séparabilité de $C(X, \mathbb{R})$, montrer qu'il existe deux suites g_n et h_n, croissantes et décroissantes, de fonctions continues, telles que $g_n \leqslant f \leqslant h_n$ et $\int h_n \longrightarrow I^-(f)$, $\int g_n \longrightarrow I^+(f)$.*
2. *Soient H et G les limites des suites h_n et g_n. Montrer que les fonctions H et $-G$ sont semi-continues inférieurement. Montrer que f est continue en x si et seulement si $H(x) = f(x) = G(x)$.*
3. *Montrer que f est Riemann intégrable si et seulement si $I^-(f) = I^+(f)$.*
4. *Montrer que $S^n f(x)/n \longrightarrow \int f$ pour tout x si f est Riemann intégrable.*

Propriété 6.2. *Si φ est uniquement ergodique, alors il existe une unique mesure de probabilité Borélienne invariante.*

On montrera plus tard que toute application continue d'un espace métrique compact admettant une unique probabilité invariante est uniquement ergodique (ce qui explique le nom).

◄ Soit m la probabilité pour laquelle $c(f) = \int f dm$. Comme $c(f\circ\varphi) = c(f)$, la mesure m est invariante. Soit μ une autre mesure invariante, alors $\int f d\mu = \int S^n f/n d\mu$ pour tout n. Par le théorème de convergence dominée, cette suite converge vers $\int(\int f dm) d\mu = \int f dm$, on a donc $\int f d\mu = \int f dm$ pour toute fonction continue f, et donc $\mu = m$. ►

Les translations φ sont des isométries. Bien qu'il existe des exemples d'homéomorphismes minimaux qui ne sont pas uniquement ergodiques, on a :

Proposition 6.3. *Soit $\varphi : X \longrightarrow X$ une isométrie d'un espace métrique compact. Si φ est minimale, elle est uniquement ergodique.*

◄ Soit f une fonction continue sur X. Elle est uniformément continue. Comme φ est une isométrie, les fonctions $f\circ\varphi^k$ sont uniformément équicontinues, et uniformément bornées. il en est donc de même des fonctions $S^n f/n$. Par le théorème d'Ascoli, la suite $S^n f/n$ admet une valeur d'adhérence g, qui est la limite de $S^n f/n$ le long d'une sous-suite n_k. Montrons que g est invariante. On a

$$g\circ\varphi = \lim(f\circ\varphi + \cdots + f\circ\varphi^{n_k})/n_k = \lim(S^{n_k}f/n_k) + (f\circ\varphi^{n_k} - f)/n_k = g.$$

La minimalité implique alors que g est constante (les ensembles $g \geqslant a, a \in \mathbb{R}$ sont invariants). Il reste à démontrer que $S^n f/n$ converge uniformément vers g. On considère la suite $M_n := \max S^n f$. Comme $S^{n+m}f = S^n f + (S^m f)\circ\varphi^n$, on a $M_{n+m} \leqslant M_n + M_m$. Un lemme très classique (lemme sous-additif, voire par exemple wikipedia) affirme alors que M_n/n converge vers $M = \inf M_n/n$. De la même façon, la suite $m_n = \min S^n f$ est sur additive, donc converge vers $m = \sup m_n/n$. Pour $\epsilon > 0$ et k assez grand, on a $g - \epsilon \leqslant S^{n_k} f/n_k \leqslant g + \epsilon$, et donc

$$g - \epsilon \leqslant m_{n_k}/n_k \leqslant m \leqslant M \leqslant M_{n_k}/n_k \leqslant g + \epsilon.$$

On a donc $M = m = g$, c'est à dire que $S^n f/n$ converge uniformément vers g. ►

Réciproquement, on a :

Proposition 6.4. *Soit φ une application continue d'un espace métrique compact X. Supposons que φ est uniquement ergodique de mesure invariante m, et soit Y le support de m. Alors Y est invariant et $\varphi_{|Y}$ est minimal.*

Le support d'une mesure est le complémentaire du plus grand ouvert de mesure nulle.
◄ Le complémentaire $X - Y$ est un ouvert de mesure nulle, donc c'est aussi le cas de sa préimage $\varphi^{-1}(X - Y)$. Ceci implique que $\varphi^{-1}(X - Y)$ est disjoint de Y, c'est à dire que Y est positivement invariant. Soit Z un fermé positivement invariant contenu dans Y, et soit $f : X \longrightarrow [0, 1]$ une fonction continue nulle sur Z et strictement positive sur $X - Z$. La première propriété implique que $S^n f(x) = 0$ pour tout $x \in Z$, et donc que $\int f dm = 0$ si Z est non vide. La seconde propriété implique que $\int f dm > 0$ si $Z \neq Y$. Z est donc soit vide soit égal à Y. ►

Proposition 6.5. *Soit $\varphi : X \longrightarrow X$ une application continue du métrique compact X. Il existe un compact invariant minimal.*

◄ On peut utiliser le Lemme de Zorn. Nous allons donner une preuve plus constructive issue des notes de Milnor. On considère une base d'ouverts dénombrable U_i. Pour chaque i, on note V_i le passé de U_i, c'est à dire $\cap_{j\geqslant 0}\varphi^{-j}(U_i)$. On construit par récurrence une suite décroissante de compacts invariants non-vides Y_n tels que $Y_0 = X$. On pose alors $Y_{n-1} \cap (X - V_n)$ si cet ensemble est non vide (c'est un compact invariant), et $Y_n = Y_{n-1} = Y_{n-1} \cap V_n$ dans l'autre cas, c'est alors clairement un compact non vide invariant. On considère $Y = \cap Y_n$, qui est un compact non-vide invariant.

Soit $H \subset \mathbb{N}$ l'ensemble des indices i tels que $U_i \cap Y$ est non-vide. Pour $i \in H$, on a $Y \subset Y_i \subset V_i$, ce qui montre que toute orbite de Y passe par U_i. Comme les ouverts $U_i \cap Y, i \in H$ constituent une base d'ouverts de Y, ceci implique la minimalité de Y. ►

Il existe une caractérisation intéressante des orbites dont l'adhérence est minimale. On dit que le point x est presque périodique si pour tout $\epsilon > 0$ il existe N tel que tout intervalle de temps de longueur N contient un entier T pour lequel $d(\varphi^T(x), x) < \epsilon$.

Proposition 6.6. *L'adhérence de l'orbite de x est minimale si et seulement si l'orbite de x est presque périodique.*

◄ Soit x un point presque périodique. Soit y et z deux points de l'adhérence de $O^+(x)$. Fixons $\epsilon > 0$. On veut montrer que l'orbite de y intersecte $B(y, \epsilon)$. Il existe k tel que $d(\varphi^k(x), z) < \epsilon/2$. Il existe donc ϵ_1 tel que $d(\varphi^k(x'), z) < \epsilon$ lorsque $d(x, x') < \epsilon_1$. Soit N tel que tout intervalle de temps de longueur N contient un temps T pour lequel $d(\varphi^T(x), x) < \epsilon_1/2$. Comme X est compact, les applications φ^i sont uniformément continues, il existe donc $\delta > 0$ tel que $\varphi^i(q, q') < \epsilon_1$ si $d(q, q') < \delta$ et si $i \in \{0, 1, \dots, N\}$. Comme y est dans l'adhérence de l'orbite de x, il existe un entier t tel que $d(\varphi^t(x), y) < \delta$. Il existe ensuite un entier $i \in \{0, 1, \dots, N\}$ tel que $d(\varphi^{t+i}(x), x) < \epsilon_1/2$. Alors,

$$d(\varphi^i(y), x) < d(\varphi^i(y), \varphi^i\circ\varphi^t(x)) + d(\varphi^{i+t}(x), x) < \epsilon_1$$

et donc $d(\varphi^{i+k}(y), z) < \epsilon$.

Soit x un point qui n'est pas presque périodique, soit Y l'adhérence de son image. Il existe $\epsilon > 0$ et une suite a_n de temps tel que, pour chaque n, aucun des points $\varphi^{a_n}(x), \dots, \varphi^{a_n+n}(x)$ n'est dans $B(x, \epsilon)$. On peut supposer en extrayant une sous-suite que $\varphi^{a_n}(x)$ a une limite $y \in Y$. Pour tout $k \geqslant 0$, le point $\varphi^{a_n+k}(x)$ est hors de la boule $B(x, \epsilon)$. A la limite, on déduit que $d(\varphi^k(y), x) \geqslant \epsilon$. Ceci étant vrai pour tout $k \geqslant 0$, x n'est pas dans l'adhérence de l'orbite de y, et donc Y n'est pas minimal. ►

La minimalité est une façon de dire que les orbites remplissent l'espace X. Il est utile de considérer des notions plus faibles dans la même direction. On se place ici dans le cadre plus général

L'application $\varphi : X \longrightarrow X$ est dite topologiquement transitive si, pour tous ouverts U et V, il existe $n \in \mathbb{N}$ tel que $\varphi^{-n}(V) \cap U$ est non-vide. Il est clair qu'une application minimale est transitive, la réciproque n'est pas vraie.

Voici deux exemples de dynamiques transitives : L'application $\varphi : \theta \mapsto 2\theta$ sur $\mathbb{T}$. Pour montrer la transitivité, on constate que l'image d'un intervalle I de longueur l est un intervalle de longueur $\min(2l, 1)$. Pour n assez grand, $\varphi^n(I)$ est donc de longueur 1. On verra ci-dessous qu'il existe alors une orbite dense. Il n'est pas évident de donner explicitement une telle orbite.

Dans le second exemple, $X = \{0, 1\}^{\mathbb{N}}$, l'ensemble des suites de zéros et de uns, et l'application est le décalage σ qui à la suite $a_0, a_1, \ldots$ associe la suite $a_1, a_2, \ldots$. L'espace X est muni de la topologie produit, c'est un espace compact métrisable. Il y a de nombreuses distances possibles, comme $d(a, b) = \sum 2^{-i}|a_i - b_i|$ ou $D(a, b) = 2^{-k}$, où k est le plus grand entier tel que les suites a et b coïncident jusqu'au rang k. Cette seconde distance est ultramétrique, c'est à dire que $D(a, b) \leqslant \max(d(a, a'), d(a', b)$ pour tout a'. On remarque que $D(\sigma(a), \sigma(b)) = \min(2D(a, b), 1)$: on a une dilatation comme dans l'exemple précédent. On montre alors la transitivité comme pour l'exemple précédent : si U est un ouvert de X, alors il existe n tel que $\sigma(U) = X$. Concrètement, U contient un cylindre, c'est à dire l'ensemble des suites de X dont les k premiers termes sont fixés (c'est aussi une boule de rayon 2^{-k} pour la distance D), et il est évident que l'image d'un tel cylindre par σ^{k+1} est X entier.

En fait, ces deux exemples sont fortement reliés. En effet on peu considérer l'application

$$p : X \ni (a_i) \mapsto \sum_i 2^{-i} a_i \mod 1 \in \mathbb{T}.$$

Cette application est continue, et c'est une semi-conjugaison entre φ et σ, c'est à dire que $\varphi \circ p = p \circ \varphi$. La transitivité de φ découle ainsi de celle de σ. De plus, l'image d'une condition initiale dont l'orbite est dense est une condition initiale dont l'orbite et dense (ceci permet de décrire certaines conditions initiales pour φ dont l'orbite est dense.) Même si p n'est pas injective, elle est très proche de l'être. En effet tous les points ont une unique préimage, sauf un nombre dénombrable de points, qui ont chacun deux préimages.

Exercice 6.2. *Soit $\varphi : X \longrightarrow X$ une application continue topologiquement transitive (où X est un espace métrique, ou même topologique quelconque). Si X admet un point isolé, montrer que X est fini.*

Propriété 6.7. *Si φ est une isométrie d'un espace métrique X qui est transitive, alors elle est minimale.*

Ceci implique qu'on peut partitionner X en compacts invariants minimaux disjoints (qui sont les adhérences des orbites).

◄ Soit $x \in X$ et $B(y, \epsilon)$ une boule ouverte. On doit montrer que l'orbite de X intersecte $B(y, \epsilon)$. La transitivité implique l'existence d'un point z et d'un temps n tels que $d(x, z) < \epsilon/2$ et $d(\varphi^n(z), y) < \epsilon/2$. Comme φ^n est une isométrie, ceci implique que $d(\varphi^n(x), y) \leqslant d(\varphi^n(x), \varphi^n(z)) + d(\varphi^n(z), x) < \epsilon$. ►

Propriété 6.8. *Si $\varphi : X \longrightarrow X$ est transitive, alors pour tous ouverts U et V, il existe des entiers n arbitrairement grands pour lesquels $\varphi^{-n}(V) \cap U$ est non vide.*

◄ Fixons $N \geqslant 0$, et posons $W = \varphi^{-N}(V)$. La transitivité de φ implique l'existence d'un entier $k \geqslant 0$ tel que $\varphi^{-k}(W) \cap U$ est non vide, c'est à dire tel que $\varphi^{-N-k}(V) \cap U$ est non vide. ►

Proposition 6.9. *Supposons que X est un espace métrique séparable et ayant la propriété de Baire (par exemple un espace complet ou un espace localement compact). L'application continue $\varphi : X \longrightarrow X$ est transitive si et seulement si il existe un point x tel que $\omega(x) = X$ (en particulier, l'orbite de x est dense).*

La dynamique $\theta \mapsto 2\theta$ sur $\mathbb{Q}/\mathbb{Z}$ est transitive, mais n'a pas d'orbite dense. Tous ses points sont pré-périodiques.

◄ On considère une base dénombrable d'ouverts, U_i. On définit $V(k, i) := \cup_{n \geqslant k} \varphi^{-n}(U_i)$. C'est l'ensemble des points x pour lesquels il existe $n \geqslant k$ tel que $\varphi^n(x) \in U_i$. La propriété précédent implique que $V(k, i)$ est un ouvert dense. La propriété de Baire implique alors que l'intersection $G := \cap_{i,k} V(k, i)$ est dense. Soit x un point de G, et U un ouvert. Il existe alors i tel que $\bar{U}_i \subset U$. Il existe une suite strictement croissante n_k telle que $\varphi^{n_k}(x)$ est contenu dans U_i pour tout x. Les valeurs d'adhérence de cette suite sont des points de $\omega(x)$ contenus dans $\bar{U}_i$, donc dans U. L'ensemble $\omega(x)$ intersecte tous les ouverts, donc il est dense. ►

Exercice 6.3. *Soit φ un homéomorphisme de l'espace métrique compact X sans point isolé. Supposons qu'il existe une orbite $\{\varphi^n(x), n \in \mathbb{Z}\}$ qui est dense. Montrer qu'alors il existe une orbite positive $\{\varphi^n(y), n \in \mathbb{N}\}$ qui est dense. Donner un contre exemple si X a des points isolés (on pourra s'inspirer de la dynamique représentée par le premier dessin du chapitre 2).*

On peut affaiblir encore la minimalité et la transitivité de diverses façons. Un compact positivement invariant K est particulièrement important pour la dynamique si il est asymptotiquement stable. On définit la stabilité asymptotique d'un compact invariant comme celle d'un point fixe :

Le compact positivement invariant K est dit Lyapounov stable si, pour tout voisinage V de K, il existe un voisinage U de K tel que $\varphi^n(U) \subset V$ pour tout $n \in \mathbb{N}$.

Il est dit asymptotiquement stable si il est Lyapounov stable et si de plus il existe un voisinage V de K tel que toute orbite issue de V converge vers K, c'est à dire que $d(\varphi^n(x), K) \longrightarrow 0$ pour tout $x \in V$.

Le bassin de K est alors l'ensemble des points x dont l'orbite converge vers K. C'est un voisinage ouvert de K.

Si X est localement compact, on montre comme dans le cas des points fixes les énoncés suivants :

Théorème 6.10. *Supposons que X est localement compact et que K est un compact positivement invariant ($\varphi(K) \subset K$). Supposons de plus qu'il existe une fonction de Lyapounov stricte en K, c'est à dire une fonction continue $f : U \longrightarrow [0, \infty)$, où U est un voisinage de K, telle que les fonctions f et $f - f \circ \varphi$ sont strictement positives sur un voisinage $V - K$ pour un voisinage V de K.*

Alors K est asymptotiquement stable.

Théorème 6.11. *On suppose X localement compact. Soit K un compact positivement invariant asymptotiquement stable, de bassin B. Alors il existe une fonction de Lyapounov continue $f : X \longrightarrow [0, 1]$ qui a les propriétés suivantes : $f = 1$ sur $X - B$, $f \in]0, 1[$ sur $B - K$, $f_{|K} = 0$, $f \circ \varphi - f < 0$ sur $B - K$.*

Exercice 6.4. *Si il existe un compact invariant asymptotiquement stable non trivial, alors φ n'est pas transitive.*

Une notion d'irréductibilité naturelle est la non-existence d'un compact invariant non-trivial asymptotiquement stable.

On peut aussi considérer les fonctions de Lyapounov, c'est à dire les fonctions continues telles que $f \circ \varphi \leqslant f$. Si φ est transitive, alors toute fonction de Lyapounov est constante. Une notion naturelle d'irréductibilité est la non-existence de fonctions de Lyapounov non constantes. Il est aussi utile de considérer un type particulier de fonctions de Lyapounov. On dit que f est une bonne fonction de Lyapounov si c'est une fonction de Lyapounov dont l'ensemble des valeures strictes est dense dans $\mathbb{R}$. On peut illustrer la différence entre ces notions lorsque $\varphi = id$. Dans ce cas, toute fonction continue est une fonction de Lyapounov, mais les bonnes fonctions de Lyapounov sont les fonctions constantes.

Finalement, on définit la transitivité par chaînes. Une suite x_i est une ϵ-chaîne si $d(\varphi(x_i), x_{i+1}) < \epsilon$ pour tout n. On dit que φ est transitif par chaînes si, pour toute $\epsilon > 0$, et tous x, y dans X, il existe une ϵ-chaîne finie reliant x à y. La transitivité implique la transitivité par chaînes, mais ce n'est pas réciproque (l'application identité est un contre exemple).

Théorème 6.12. *Les trois propriétés suivantes sont équivalentes pour une application continue $\varphi : X \longrightarrow X$ de l'espace métrique compact X. Elles sont satisfaites si φ est transitive.*

- *φ est transitif par chaînes.*
- *Toute bonne fonction de Lyapounov est constante.*
- *Il n'existe pas de compact invariant asymptotiquement stable non trivial (c'est à dire ni vide ni égal à X).*

◄ On montre les deux équivalences.

Supposons que φ n'est pas transitive par chaînes. Il existe alors un point x, et un réel $\epsilon > 0$ tel que l'ouvert $U \neq X$, où U est l'ouvert des points accessibles depuis x par une ϵ-chaîne. Si $y \in \varphi(U)$, alors la boule ouverte $B(y, \epsilon)$ est contenue dans U. On a donc $\overline{\varphi(U)} \subset U$. Nous avons montré qu'il existe alors une fonction de Lyapounov non constante à valeurs dans $[0, 1]$, qui n'a que deux valeurs neutres, 0 et 1. C'est donc une bonne fonction de Lyapounov non constante.

Supposons qu'il existe une bonne fonction de Lyapounov f non constante, posons $a = \min f$ et $b = \max f$. L'ensemble des valeurs strictes de f est un ouvert dense de $[a, b]$. Soit $]c, d[$ un intervalle maximal dans cet ouvert. Le compact $K = f^{-1}([a, c])$ est alors asymptotiquement stable. Si U est un voisinage de K, on considère $m = \min_{x \in \partial U} f$, on a $m > c$ car ∂U est un compact disjoint de K. On considère alors une valeur stricte $e \in]c, m[$. L'ouvert $V = f^{-1}([a, e[)$ est un voisinage positivement invariant de K contenu dans U. On a montré que K est Lyapounov stable. Finalement, si x vérifie $f(x) < d$, alors $\omega(x)$ est contenu dans un niveau neutre de f inférieur à d, donc inférieur à c. On a donc $\omega(x) \subset K$, ce qui montre la stabilité asymptotique.

Toujours en supposant l'existence de la fonction f, on montre que φ n'est pas transitive par chaînes. On considère en effet comme ci-dessus une valeur stricte $e \in]a, b[$. On note $e' = \max_{f^{-1}(a,e])} f \circ \varphi$. On a $e' \leqslant e$, et, si $e' = e$, alors il existe un point x tel que $f(x) = e$ et $f \circ \varphi(x) = e$, ce qui contredit le caractère strict de la valeur e. On a donc $e' < e$. Soit ϵ

la distance entre les compacts disjoints $f^{-1}([e, b])$ et $f^{-1}([a, e'])$. Aucune ϵ-chaîne ne peut sortir de $f^{-1}([a, e])$, ce qui contredit la transitivité par chaînes

Si il existe un compact invariant asymptotiquement stable A non trivial, de bassin B, on montre comme dans les cas des points fixes qu'il existe une fonction de Lyapounov $f : X \longrightarrow [0, 1]$ qui est nulle sur A, égale à 1 hors de B, et qui vérifie $f \circ \varphi - f < 0$ sur $B - A$. C'est une bonne fonction de Lyapounov. ►

On a en fait montré l'énoncé suivant, qui sera réutilisé :

Lemme 6.13. *Soit X un espace métrique compact, φ une application continue de X, et f une fonction de Lyapounov continue. Si e est une valeur stricte de f, alors il existe $\epsilon > 0$ tel que toute ϵ-chaîne commençant dans $\{f < e\}$ reste dans $\{f < e\}$.*

7 Récurrence, récurrence par chaines

On considère ici une application continue φ d'un espace métrique X, pas nécessairement compact.

Le point x est dit récurrent si $x \in \omega(x)$.

Propriété 7.1. *L'ensemble R des points récurrents vérifie $\varphi(R) = R$.*

◄ On a $\varphi(\omega(x)) = \omega(x)$, et $\omega(x) = \omega(\varphi(x))$. Si x est récurrent, alors $x \in \omega(x) = \omega(\varphi(x))$, donc $\varphi(x) \in \omega(\varphi(x))$. Réciproquement soit $x \in \omega(x)$ un point récurrent. Comme $\varphi(\omega(x)) = \omega(x)$, il existe $y \in \omega(x)$ tel que $\varphi(y) = x$. Alors, $y \in \omega(x) = \omega(y)$, donc y est récurrent. On a montré que $x \in \varphi(R)$. ►

L'ensemble des points récurrents n'est pas forcément fermé. Un exemple est donné par le doublement $\theta \longmapsto 2\theta$ sur $\mathbb{T}$. Le point $1/2$ n'est pas récurrent (il est pré-periodique, mais pas périodique), alors que les points récurrents sont denses puisque l'application est transitive.

On peut considérer aussi comme exemple le pendule pesant, qui est donné par les équations différentielles $\dot{\theta} =$
PENDULE

Propriété 7.2. *Si φ est une application continue d'un espace métrique compact X, alors il existe un point récurrent.*

◄ On considère un compact invariant minimal Y. Tout $y \in Y$ vérifie $\omega(y) = Y$. En effet, $\omega(y)$ est un compact invariant non vide don égal à Y. ►

On dit que U est un ouvert errant si $\varphi^k(U)$ est disjoint de U pour tout $k \in \mathbb{N}$. C'est équivalent au fait que les préimages $\varphi^{-k}(U)$, $k \in \mathbb{N}$, sont toutes disjointes.

Propriété 7.3. *Si φ est un homéomorphisme, alors l'ouvert U est errant pour x si et seulement si les ouverts $\varphi^k(U)$, $k \in \mathbb{Z}$ sont disjoints. L'ouvert U est donc errant pour φ si et seulement si il est errant pour φ^{-1}.*

◄ Si il existe $k < l$ tels que $\varphi^k(U) \cap \varphi^l(U)$ est non vide, alors $U \cap \varphi^{l-k}(U)$ est non-vide, donc U est errant pour φ. ►

On dit que x est un point errant si il est contenu dans un ouvert errant, sinon on dit que x est non errant. Soit Ω l'ensemble des points non-errants.

Propriété 7.4. *L'ensemble non-errant Ω est fermé et vérifie $\varphi(\Omega) \subset \Omega$.*

◄ L'ensemble non-errant est le complémentaire de la réunion des ouverts errants. C'est donc un fermé. Si $\varphi(x)$ est errant, alors il est contenu dans un ouvert errant U. Alors, l'ouvert $\varphi^{-1}(U)$ est errant, et il contient x donc x est errant. On a donc $\varphi(\Omega) \subset \Omega$. ►

Propriété 7.5. *Pour tout $x \in X$, l'ensemble limite $\omega(x)$ est contenu dans l'ensemble non errant. En particulier, tout point récurrent est non errant.*

◄ Soit $y \in \omega(x)$ et soit U un voisinage de y. Il existe alors n et $m \geqslant n + 1$ tels que $\varphi^n(x) \in U$ et $\varphi^m(x) \in U$, et donc $\varphi^{m-n}(U) \cap U$ est non vide. ►

Propriété 7.6. *Si x est non-errant, alors pour tout voisinage U de x il existe n arbitrairement grand tel que $\varphi^{-n}(U) \cap U$ est non vide.*

◄ Si x est périodique, la conclusion est satisfaite.

Si x n'est pas périodique, les points $x, \varphi(x), \ldots, \varphi^N(x)$ sont disjoints pour tout N. Il existe donc un voisinage ouvert U de x tel que U, $\varphi(U), \ldots, \varphi^N(U)$ sont disjoints. Si x est non errant, il existe j tel que $\varphi^j(U) \cap U$ est non vide, et nécessairement $j \geqslant N$. ►

Considérons maintenant le cas d'un flot φ^t sur l'espace métrique X. Pour tout ouvert U, il existe toujours $t \neq 0$ tel que $\varphi^t(U) \cap U$ est non-vide. On définit les points non errants par la propriétés suivante :

Le point x est non errant si, pour tout voisinage U de x, il existe des temps t arbitrairement grands pour lesquels $\varphi^{-t}(U) \cap U$ est non-vide.

Proposition 7.7. *Soit V un champ de vecteurs complet sur la variété X. Le point $x \in X$ est errant si et seulement si il existe une boite de flot globale autour de l'orbite de x, c'est à dire un ouvert invariant U contenant X et un difféomorphisme $\psi : U \longrightarrow W \times \mathbb{R}$ tel que $\psi_* V = (0, 1)$, où W est un ouvert de $\mathbb{R}^{d-1}$.*

◄ On considère un plongement F d'un voisinage ouvert W de 0 dans $\mathbb{R}^{d-1}$ telle que $F(0) = x$ et $dF_0(\mathbb{R}^{d-1})$ est un supplémentaire de $V(x)$. On considère alors l'application $\phi(y,t) = \varphi(t, F(y))$. Nous avons déjà démontré que ϕ est un difféomorphisme local en $(0,0)$. Comme $\phi(y, t+s) = \varphi^t(\phi(y,s))$, on a $d\phi_{(0,t)} = d\varphi^t_0 \circ d\phi_{(0,0)}$, donc ϕ est un difféomorphisme local en chaque point $(0,t)$.

Montrons maintenant que ϕ est injectif sur $W \times \mathbb{R}$ si W est assez petit. Sinon, il existe deux suites y_n et z_n tendant vers 0 deux suites de temps $t_n \geqslant s_n$ telles que $\phi(y_n, t_n) = \phi(z_n, s_n)$, c'est à dire $\varphi^{t_n - s_n}(y_n) = z_n$.

Puisque x est errant, la suite $t_n - s_n$ est bornée. On peut donc supposer qu'elle converge vers une limite τ, et alors $\varphi^\tau(x) = x$. Comme le point x n'est pas périodique, on a $\tau = 0$. Comme l'application ϕ est injective au voisinage de $(0,0)$, c'est une contradiction. ►

Exercice 7.1. *Soit V un champ de vecteurs C^1 sur une variété X, x un point de X, et $T > 0$ un temps tel que l'orbite $t \longmapsto \varphi^t(x)$ est injective sur $[0,T]$. Montrer que ce segment d'orbite est contenu dans un tube de flot, c'est à dire qu'il existe un ouvert U de X et un difféomorphisme $\psi : U \longrightarrow W \times]a,b[$ tel que $\psi(x) = (0,0)$, $\psi_* V \equiv (0,1)$ et $[0,T] \subset]a,b[$.*

En notant $\Omega(f)$ l'ensemble non errant, il faut prendre garde au fait que $\Omega(f_{|\Omega})$ n'est pas forcément égal à Ω.

On décrit un exemple sur $\mathbb{T} \times \mathbb{R}$. On se donne une fonction dérivable 1-périodique à valeurs positives $f : \mathbb{R} \longrightarrow [0,1]$ telle que $|f'| < 1/2$, telle que $f(0) = 0$, $f(1/2) = 0$, et $f(\theta) > 0$ si θ n'est pas dans $\mathbb{Z}/2$. L'application $\psi(\theta) = \theta + f(\theta)$ sur $\mathbb{T}$ a pour ensemble non-errant les points 0 et $1/2$. Considérons alors l'application

$$\varphi(\theta, z) = (\psi(\theta) + z^2, z/2).$$

On dit finalement que le point x est récurrent par chaînes si, pour tout $\epsilon > 0$, il existe une ϵ-chaîne non triviale reliant x à x.

On introduit à ce propos les relations suivantes : on dit que $x \rhd y$ si, pour tout $\epsilon > 0$, il existe une ϵ-chaine non triviale (c'est à dire comportant au moins 2 points) reliant x à y. On note aussi $\unrhd$ la relation $x \unrhd y$ si et seulement si $x = y$ ou $x \rhd y$. Le point x est donc récurrent par chaines si et seulement si $x \rhd x$. Si x est récurrent par chaines, alors les relations $x \rhd y$ et $x \unrhd y$ sont équivalentes. En effet, $x \rhd x \unrhd y \Rightarrow x \rhd y$.

Exercice 7.2. *Montrer que la relation $\rhd$ est fermée et transitive.*

Propriété 7.8. *$x \rhd y \iff \varphi(x) \unrhd y$.*

◄ Supposons que $x \rhd y$ et $y \neq \varphi(x)$. Fixons $\epsilon > 0$ et $\delta << \epsilon$. Il existe un δ-chaîne $x = z_1, z_2, \ldots, z_k = y$, et forcément $k \geqslant 3$. Alors, La suite $\varphi(x), z_3, \ldots, z_k$ est une ϵ-chaine. En effet

$$d(\varphi \circ \varphi(x), z_3) \leqslant d(\varphi \circ \varphi(x), \varphi(z_2)) + d(\varphi(z_2), z_3) < \epsilon$$

si δ est assez petit, car $d(\varphi(x), z_2) < \delta$ et φ et continue en $\varphi(x)$.

La réciproque est évidente. ►

En particulier, le point x est récurrent par chaines si et seulement si $\varphi(x) \unrhd x$.

Théorème 7.9. *$x \unrhd y$ si et seulement si $f(y) \leqslant f(x)$ pour toute bonne fonction de Lyapounov f. Plus précisément, si $x \not\unrhd y$ alors il existe une fonction de Lyapounov $f : X \longrightarrow [0,1]$ telle que $f = 0$ au voisinage de x, $f = 1$ au voisinage de y, et dont les seules valeurs neutres sont 0 et 1.*

Ce résultat implique la fermeture des relations $\rhd$ et $\unrhd$. Il implique aussi que $\varphi(x) \rhd \varphi(y)$ (ceci peut aussi être démontré directement). si $x \rhd y$.

◄ Si il existe une bonne fonction de Lyapounov telle que $f(y) > f(x)$, alors il existe une valeur stricte e de f dans l'intervalle $]f(x), f(y)[$. Alors des arguments déjà vus montrent que, pour ϵ assez petit, aucune ϵ-chaîne ne peut sortir de l'ouvert $\{f < e\}$, donc on ne peut pas avoir $x \unrhd y$.

Soit $x \not\unrhd y$, alors il existe $\epsilon > 0$ tel que y n'est pas contenu dans l'ouvert U des points accessibles depuis x par des ϵ-chaînes. De plus, il existe une petite boule ouverte B centrée en x qui ne contient pas y dans son adhérence, et telle que $\overline{\varphi(B)} \subset U$. En posant $V = U \cup B$, on a $\overline{\varphi(V)} \subset U \subset V$. On considère alors une fonction $f : X \longrightarrow [0,1]$ égale à 1 hors V, égale à 0 sur $\overline{\varphi(V)}$, et à valeurs dans $]0,1[$ sur l'ouvert restant. C'est une bonne fonction de Lyapunov pour laquelle $1 = f(y) > f(x)$.

On peut alors considérer une fonction continue $g : \mathbb{R} \longrightarrow \mathbb{R}$ telle que $g = 0$ sur $(-\infty, a]$, g est strictement croissante sur $[a,1]$ et $g = 1$ sur $[1,\infty)$, avec $a \in]f(x), 1[$. La fonction $g \circ f$ est alors une bonne fonction de Lyapounov qui vérifie les conditions de l'énoncé. ►

Théorème 7.10. *L'ensemble RC des points récurrents par chaines est un compact invariant. C'est l'ensemble des points x qui sont neutres pour toutes les bonnes fonctions de Lyapounov. Il existe une bonne fonction de Lyapounov qui est stricte en dehors de RC.*

◄ Le point x est récurrent par chaines si et seulement si $\varphi(x) \trianglerighteq x$ c'est à dire si et seulement si x est neutre pour toute bonne fonction de Lyapounov f. Ceci implique que RC est compact.

Considérons maintenant l'espace de Banach séparable $C(X, \mathbb{R})$ des fonctions continues. Soit $\mathcal{L}$ l'ensemble des bonnes fonctions de Lyapounov à valeurs dans $[0, 1]$ et dont les seules valeurs neutres sont 0 et 1. Nous avons démontré ci-dessus qu'il existe $f \in \mathcal{L}$ telle que $f(x) \in]0, 1[$ (c'est donc une valeur stricte) si x n'est pas récurrent par chaines. Soit $f_i, i \geqslant 1$ une suite dense dans $\mathcal{L}$. Posons $h := \sum_i 3^{-i} f_i$. La série converge uniformément, donc f est une fonction continue. C'est une fonction de Lyapounov. Si x n'est pas récurrent par chaines, alors il existe $f \in \mathcal{L}$ tel que $f \circ \varphi(x) < f(x)$, donc il existe i tel que $f_i \circ \varphi(x) < f_i(x)$, et donc $h \circ \varphi(x) < h(x)$. La fonction h est donc stricte hors de RC.

Il reste à montrer que h est une bonne fonction de Lyapounov. Si x est neutre pour h, il est neutre pour f_i pour tout i, et donc $f_i(x) \in \{0, 1\}$. La valeur $h(x)$ est donc contenue dans le cantor des points $x \in [0, 1]$ dont le développement en base 3 ne contient pas de 2. Le complémentaire de ce Cantor est dense. ►

La relation $\triangleright$ est fermée, transitive, et, en restriction à RC, reflexive (c'est à dire que $x \triangleright x$ $\forall x \in RC$). La relation symétrisée $x \bigtriangleup y :\Leftrightarrow x \triangleright y$ et $x \triangleright y$ est donc une relation d'équivalence fermée sur RC.

Les classes d'équivalence de la relation $\bigtriangleup$ sont appelées les composantes transitives par chaines de RC. Elles constituent une partition de RC en compacts invariants. L'invariance découle du fait que $x \triangleright \varphi(x)$ et, pour tout $x \in RC$, $\varphi(x) \in RC$ et $\varphi(x) \trianglerighteq x$ donc $\varphi(x) \triangleright x$.

Théorème 7.11. *La restriction de φ à RC est récurrente par chaines. La restriction de φ à chaque composante transitive de RC est transitive par chaines.*

On dit que l'ensemble récurrent par chaines est intérieurement récurrent par chaines, et que les composantes transitives par chaines sont intérieurement transitives par chaines.

◄ Le second énoncé implique le premier.

Soit $K \subset RC$ une composante transitive et x_0 un point de K. Notons F^+ l'ensemble des points y tels que $x_0 \trianglerighteq y$ et F^- l'ensemble des points y tels que $y \trianglerighteq x_0$. On remarque que $F^+ \cap F^- = K$. Pour tout $x \in X - F^-$, il existe une fonction de Lyapunov $f_x : X \longrightarrow [0, 1]$ qui vaut 1 au voisinage de x_0 et 0 au voisinage de x, et donc les seules valeurs neutres sont 0 et 1. Comme plus haut, on considère une suite f_i qui est dense dans l'ensemble des fonctions de $\mathcal{L}$ qui sont nulles en x_0, puis $f^+ := -1 + \sum_{i \geqslant 1} 3^{-i} f_i$. C'est une bonne fonction de Lyapounov à valeurs dans $[-1, 0]$, nulle en x_0 (donc sur F^-) et strictement négative hors de F^-.

On construit de la même façon une bonne fonction de Lyapounov $f^+ : X \longrightarrow [0, 1]$ qui est strictement positive hors de F^+ et nulle en x_0, donc sur F^+.

Pour tout voisinage U de K, il existe $\varepsilon > 0$ tel que toute ε-chaine reliant un point de K à un point de K reste dans U. En effet, il existe $a > 0$ tel que $\{f^+ < a\} \cap \{f^- > -a\} \subset U$. Soit $b \in [0, a]$ une valeur stricte de f^+. Pour ε assez petit, aucune ε-chaine ne peut sortir du domaine $\{f^+ < b\}$, donc les ε-chaines issues de K restent dans $\{f^+ < a\}$. On montre de manière similaire qu'une ε-chaine arrivant dans K est contenue dans $\{f^- > -a\}$. Les ε-chaines partant dans K et arrivant dans K sont donc contenues dans $\{f^+ < a\} \cap \{f^- > -a\}$, donc dans U.

Fixons maintenant $\epsilon > 0$ et x, y dans K. On veut montrer qu'il existe une ϵ-chaine de x à y qui est contenue dans K. On prend $\varepsilon << \delta << \epsilon$, et on pose $U = \cup_{z \in K} B(z, \delta)$, c'est un voisinage de K. Il existe une ε-chaine $x = z_1, z_2, \dots, z_k = y$ dans X. Si ε est assez petit par rapport à δ, alors cette ε-chaine est contenue dans U. Pour chaque i, il existe donc un point $q_i \in K$ tel que $d(z_i, q_i) < \delta$ (on prend $q_i = z_i$ dans le cas où $z_i \in K$, notamment pour $i = 1$ et $i = k$). On a alors

$$d(q_{i+1}, \varphi(q_i)) \leqslant d(q_{i+1}, z_{i+1}) + d(z_{i+1}, \varphi(z_i)) + d(\varphi(z_i), \varphi(q_i)) < \delta + \varepsilon + \epsilon/2 < \epsilon$$

à condition que δ soit choisi assez petit par rapport à ϵ (au vu de la continuité uniforme de φ). ►

Soit X une variété Riemannienne, c'est à dire que chacun des espaces tangents $T_x X$ est muni d'un produit scalaire. Si f est une fonction lisse sur X, alors le gradient inverse $V(x)$ de f est le champ de vecteurs sur X tel que $\langle V(x), \cdot \rangle = -df_x$ pour tout x. La fonction f est, de manière évidente, une fonction de Lyapounov pour le champ V.

Exercice 7.3. *Supposons que la fonction f est minorée et propre. Montrer que le flot de V est complet.*

En s'inspirant de cette situation, on dit que f est une fonction de Lyapounov régulière du champ V si c'est une fonction lisse telle que $df_x \cdot V(x) < 0$ sauf aux points où $df_x = 0$. Dans le cas des applications, on dit que f est une fonction de Lyapounov régulière si c'est une fonction de Lyapounov lisse telle que $f \circ \varphi(x) < f(x)$ pour tout x qui n'est pas un point critique de f. Autrement dit, on demande que les points neutres de f soient des points critiques de f. En conséquence, les valeurs neutres de f sont des valeurs critiques. La remarque suivante contribue à justifier la notion de bonne fonction de Lyapounov :

Proposition 7.12. *Toute fonction de Lyapounov régulière est une bonne fonction de Lyapounov.*

◄ Au vu du théorème de Sard, l'ensemble des valeurs régulières est de mesure totale, donc il est dense. ►

8 Mesures invariantes, mesures ergodiques

On s'intéresse pour commencer à l'existence de mesures de probabilité invariantes pour une application, c'est à dire de mesures telles que $\varphi_* m = m$.

Propriété 8.1. *Pour que la mesure m soit invariante, il suffit qu'il existe une famille $\mathcal{Y}$ d'ensembles Boréliens qui est stable par intersections finies et qui engendre la tribu Borélienne pour laquelle $\varphi_* m(Y) = m(Y)$ pour tout $Y \in \mathcal{Y}$.*

Il suffit aussi qu'il existe une partie $\mathcal{F} \subset C(X)$ engendrant un sou-espace dense, telle que $\int f \circ \varphi dm = \int f dm$ pour tout $f \in \mathcal{F}$.

◄ La première partie découle du résultat classique de théorie de la mesure selon lequel deux mesures finies qui coïncident sur une famille d'ensembles stable par intersections finies coïncident sur la tribu engendrée par cette partie (c'est le théorème de classe monotone). La seconde affirmation est laissée en exercice. ►

Exercice 8.1. *Montrer que l'application $\varphi : \mathbb{T} \ni x \mapsto 2x \in \mathbb{T}$ préserve la mesure de Lebesgue.*

Une première fois en montrant que, si $I \subset \mathbb{T}$ est un intervalle de longueur $l < 1$, alors la préimage de $\varphi^{-1}(I)$ est la réunion de deux intervalles disjoints de longueur $l/2$.

Une seconde fois en montrant que $\int e_k \circ \varphi d\lambda = \int e_k d\lambda$ pour tout k, où $e_k(x) = \exp(2i\pi kx)$.

La mesure de dirac δ_x est invariante si et seulement si x est un point fixe, puisque $\varphi_* \delta_x = \delta_{\varphi(x)}$. De la même façon, si x est un point fixe de période T, alors la mesure $(\delta_x + \delta_{\varphi(x)} + \cdots + \delta_{\varphi^{T-1}(x)})/T$ est invariante.

Le résultat suivant est donc évident si il existe un point fixe ou une orbite périodique.

Théorème 8.2. *Soit $\varphi : X \longrightarrow X$ une application continue, où X est un espace métrique compact.*

L'application φ admet une mesure de probabilité invariante.

Avant de démontrer ce théorème, nous allons étudier l'espace $\mathcal{P}(X)$ des mesures de probabilité Boréliennes sur X. On rappelle que cet espace s'identifie au convexe fermé des formes linéaires continues l sur $C(X)$ qui sont positives ($l(f) \geqslant 0$ si $f \leqslant 0$) et vérifient $l(1) = 1$. Un exemple de mesure de probabilité est le Dirac δ_x au point x, qui correspond à la fonctionnelle d'évaluation $f \mapsto f(x)$.

On munit $\mathcal{P}(X)$ de la topologie faible-$*$, c'est à dire de la topolgie engendrée par les applications linéaires $\mu \mapsto \int f d\mu, f \in C(X)$.

Cette topologie est métrisable et compacte. C'est un fait général (la boule unité du dual d'un espace de Banach séparable est un espace métrisable compact), mais nous allons en donner une présentation *ad hoc*.

On définit pour ceci la distance

$$D(\mu, \eta) = \sup_{f \in Lip_1(X)} \left(\int f d\mu - \int f d\eta \right)$$

où Lip_a est l'ensemble des fonctions a-Lipschitz.

Montrons d'abord que c'est bien une distance. L'inégalité triangulaire est claire. De plus, si $D(\mu, \eta) = 0$, alors $\int f d\mu = \int f d\eta$ pour toute fonction Lipschitz, et donc pour toute fonction continue, par densité.

Exercice 8.2. *Montrer que l'application $X \ni x \mapsto \delta_x \in \mathcal{P}(X)$ est un plongement isométrique de (X, d) dans $(\mathcal{P}(X), D)$.*

Montrons maintenant que $\mu_n \longrightarrow \mu$ pour la distance d si et seulement si $\int f d\mu_n \longrightarrow \int f d\mu$ pour tout $f \in C(X)$.

Supposons pour commencer que $D(\mu_n, \mu) \longrightarrow 0$. Les éléments de $\mathcal{P}(X)$ sont tous 1-Lipschitz en tant que fonctions sur $C(X)$. La convergence sur la partie dense des fonctions Lipschitz implique donc la convergence sur tout $C(X)$.

Pour la réciproque, on fixe un point x_0 et on note L l'ensemble des fonctions 1-Lipschitz qui sont nulles en x_0. Par Ascoli, c'est un compact, et de plus on voit que

$$D(\mu, \eta) = \sup_{f \in L} \left(\int f d\mu - \int f d\eta \right).$$

Comme les éléments de $\mathcal{P}(X)$ sont tous 1-Lipschitz en tant que fonctionnelles sur L, et que L est compact, la convergence ponctuelle implique la convergence uniforme, c'est à dire que $D(\mu_n, \mu) \longrightarrow 0$.

Proposition 8.3. $(\mathcal{P}(X), D)$ *est un espace métrique compact.*

◄ Par Ascoli, l'espace des fonctionnelles 1-Lipschitz sur le métrique compact L qui valent 0 en 0 est compact. $\mathcal{P}(X)$ est un fermé de cet espace. ►

On utilisera la propriété suivante :

Propriété 8.4. *Si $U \subset X$ est compact, l'application $m \mapsto m(U)$ est semi-continue inférieurement.*

◄ La fonction 1_U qui vaut 1 sur U et 0 hors de U est le supremum des fonctions continues $f : X \longrightarrow [0,1]$ qui sont nulles hors de U. On a donc $m(U) = \int 1_U dm = \sup \int f dm$ où le supremum est pris sur l'ensemble des fonctions continues à valeurs dans $[0,1]$ nulles sur U. La fonction $m \mapsto m(U)$ est donc le supremum d'une famille de fonctions continues, donc elle est semi-continue inférieurement. ►

La démonstration du théorème est maintenant facile :
◄ Pour tout $\mu \in \mathcal{P}(X)$, on pose $S_*^n \mu = \mu + \varphi_* \mu + \cdots + \varphi_*^{n-1} \mu$, de sorte que $S_*^n \mu / n \in \mathcal{P}(X)$. Par compacité, il existe une sous-suite n_k telle que $S_*^{n_k} \mu / n_k$ converge vers une limite m dans $\mathcal{P}(X)$. Montrons que m est invariante. On a

$$\varphi_*(S_*^{n_k} \mu / n_k) = (S_*^{n_k} \mu + \varphi_*^{n_k} \mu - \mu)/n_k$$

qui a la limite donne $\varphi_* m = m$. ►

Revenons maintenant au cas des applications uniquement ergodiques.

Théorème 8.5. *Soit $\varphi : X \longrightarrow X$ une applications continue de l'espace métrique compact X. Alors les propriétés suivantes sont équivalentes :*
— *φ est uniquement ergodique.*
— *Il existe une unique mesure de probabilité invariante*
— *Pour tout $\mu \in \mathcal{P}(X)$, la suite de mesures $S_*^n \mu$ converge vers une limite m indépendante de μ, et cette convergence est uniforme en μ.*
— *Pour toute fonction f, la suite $S^n f / n$ converge uniformément vers une constante.*

◄ Nous avons déjà démontré que le premier point implique le second.

Supposons qu'il existe une unique mesure de probabilités invariante m. Alors pour tout $\mu \in \mathcal{P}(X)$, m est la seule valeur d'adhérence de la suite $S_*^n \mu$, ce qui implique que cette suite converge vers m. Supposons que la convergence n'est pas uniforme par rapport à μ. Il existe alors une suite μ_n de mesures de probabilité telle que $D(S_*^n \mu_n / n, m) \geqslant \epsilon > 0$. Posons $\eta_n := S_*^n \mu_n / n$. Soit η_k une sous-suite le long de laquelle η_{n_k} converge vers un limite η. On va montrer que η est invariante, donc que $\eta = m$, ce qui est une contradiction. Comme précédemment, l'invariance de η s'obtient par passage à la limite dans l'égalité $\varphi_* \eta_n = \eta_n + (\varphi_*^n \mu_n - \mu_n)/n$.

En supposant le troisième point, on a en particulier convergence, uniforme en x, des mesures $S_*^n \delta_x / n$ vers m. Pour toute fonction Lipschitz f, on a donc

$$S^n f(x)/n = \int f d(S_*^n \delta_x / n) \longrightarrow \int f dm$$

uniformément en x. Par un argument de densité déjà utilisé, on en déduit que cette convergence uniforme a lieu pour tout $f \in C(X)$. ►

Exercice 8.3. *Soit $Y \subset X$ un compact invariant asymptotiquement stable de bassin B et soit m une mesure de probabilité invariante. Montrer que $m(B) = m(Y)$.*

Si x_1 et x_2 sont deux points fixes de φ, alors toute combinaison convexe de δ_{x_1} et de δ_{x_2} est une mesure de probabilité invariante pour φ. Pourtant, les mesures δ_{x_1} et δ_{x_2} ont un sens dynamique plus clair que la mesure $(\delta_{x_1} + \delta_{x_2})/2$.

Cette remarque conduit à la notion de mesure ergodique.

Propriété 8.6. *Les propriétés suivantes sont équivalentes pour une mesure de probabilité m invariante par une application mesurable $\varphi : X \longrightarrow X$.*
— *Tout ensemble mesurable A tel que $\varphi^{-1}(A) = A$ est de mesure 1 ou 0.*
— *Toute fonction mesurable f telle que $f \circ \varphi = f$ est constante presque partout.*
— *Toute fonction mesurable bornée f telle que $f \circ \varphi = f$ p.p. est constante presque partout.*
On dit alors que la mesure m est ergodique.

On dira par fois que φ est ergodique our la mesure m, où juste que φ est ergodique lorsque la mesure m est sous-entendue. Cette terminologie ne doit pas faire oublier que l'ergodicité est une propriété d'une masure invariante particulière, alors que l'unique ergodicité est une propriété d'une système dynamique topologique.
◄ Suposons le premier point et considérons une fonction invariante f. Alors pour tout $a \in \mathbb{R}$, l'ensemble $\{f > a\}$ est invariant, il est donc de mesure 0 ou 1. Comme $\cap_n \{f > n\}$ est vide, on a $m(\{f > a\}) = 0$ pour a assez grand. De

même $m(\{f > a\}) = 1$ pour a assez petit. Il existe donc une unique valeur de transition b telle que $m(\{f > b\}) = 0$ et $m(\{f > a\}) = 1$ pour $a < b$, c'est à dire que $f = a$ p.p.

Soit maintenant f une fonctions telle que $f\circ\varphi = f$ p.p. Posons $g = \limsup f\circ\varphi^n$, de sorte que $g\circ\varphi = g$. Si l'on a supposé le second point, alors la fonction g est constante presque partout. Il reste à montrer que la fonction g est presque partout égale à la fonction f. Soit Z l'ensemble de mesure pleine sur lequel $f\circ\varphi = f$. Alors $f\circ\varphi^2 = f\circ\varphi$ sur $Z_1 := \varphi^{-1}(Z)$. On montre ainsi que

$$f = f\circ\varphi = f\circ\varphi^2 = \cdots$$

sur $A := \cap_{i\geqslant 0}\varphi^{-i}(Z)$. Comme φ préserve la mesure m, chacun des ensembles $\varphi^{-i}(Z)$ est de mesure pleine, et donc aussi l'ensemble A. Clairement, $f = g$ sur A.

Finalement, on déduit le premier point du troisième en utilisant la fonction indicatrice de A. ►

L'unique mesure invariante d'un système uniquement ergodique est ergodique. En effet, pour tout ensemble mesurable invariant A de mesure strictement positive, la mesure $m_A(Y) := m(Y \cap A)/m(A)$ est invariante. Comme $m_A(A) = 1$, on conclut que $m_A \neq m$ si $m(A) \neq 1$, ce qui contredit l'unique ergodicité.

Dans le cas où φ est une application continue d'un espace métrique compact X, on peut donner une caractérisation différente de l'ergodicité. On remarque pour commencer que l'ensemble $\mathcal{PI}$ des mesures de probabilité invariantes est une partie convexe et compacte de $\mathcal{P}(X)$. Dans un convexe C, on dit que le point x est extrémal si il n'est pas une combinaison convexe non-triviale d'autres points de C.

Proposition 8.7. *La mesure invariante m est ergodique si et seulement si c'est un point extrémal du convexe $\mathcal{PI}$.*

◄ Si m n'est pas ergodique, alors il existe une partition de X en deux ensembles mesurables invariants A et B de mesure srictement positive. On pose alors $m_A(Y) = m(A\cap Y)/m(A)$ et $m_B(Y) = m(B\cap Y)/m(B)$, ce sont des mesures invariantes et $m = m(A)m_A + m(B)m_B$, donc m n'est pas extrémale.

On montrera la réciproque plus tard. ►

Nous allons maintenant montrer que toute application continue d'un espace métrique compact admet une mesure invariante ergodique. C'est un résultat semblable à celui qui donne l'existence de compacts invariants minimaux.

Théorème 8.8. *Il existe une mesure invariante ergodique.*

De plus, toute mesure invariante peut être approchée par des combinaisons convexes finies de mesures ergodiques.

Dans un espace vectoriel, on dit qu'une partie C est convexe si elle est stable par combinaisons convexes finies. L'enveloppe convexe d'une partie est le plus petit ensemble convexe contenant cette partie, c'est l'ensemble des combinaisons convexes finies d'éléments de cette partie. L'énoncé ci-dessus est un cas particulier du théorème de Krein-Milman, qui affirme qu'une partie convexe et compacte d'un espace vectoriel topologique est la fermeture de l'enveloppe convexe des ses points extrémaux. La preuve, dans le cas général, que nous donnons ci-dessous n'est pas très différente de la preuve générale. Nous tirons toutefois partie d'un argument de séparabilité spécifique.

◄ Considérons une suite $f_i, i \geqslant 1$ dense dense dans $C(X)$. Soit $\mathcal{P}_1$ l'ensemble des mesures $m \in \mathcal{P}$ qui minimisent la fonctions $m \mapsto \int f_1 dm$ sur $\mathcal{P}$. On définit ensuite par récurrence une suite décroissante de convexes compacts tels que $\mathcal{P}_{i+1}$ est l'ensemble des minima de la fonction $m \mapsto f_i dm$ sur $\mathcal{P}_i$. L'intersection $\mathcal{P}_\infty$ des ces compacts est non vide. Montrons que toute mesure $m \in \mathcal{P}_\infty$ est ergodique.

Supposons que m s'écrit $m = t\mu_1 + (1-t)\mu_0$ où μ_0 et μ_1 sont invariantes et où $t \in]0,1[$. On montre alors par récurrence que la propriété suivante est vraie pour tout i : $\int f_i d\mu_0 = \int f_i d\mu_1 = \int f_i dm$ donc μ_0, μ_1 appartiennent à $\mathcal{P}_i$. L'égalité $\int f_i d\mu_0 = \int f_i d\mu_1$ pour tout i implique que $\mu_0 = \mu_1$ par densité des f_i.

Nous avons montré l'existence d'une mesure extrémale, donc ergodique. Le fait que le convexe compact non vide $\mathcal{P}_\infty$ ne contient que des points extrémaux montre qu'il ne contient qu'un seul point. Ceci implique que le diamètre des compacts convexes $\mathcal{P}_i$ tend vers 0.

Considérons maintenant une mesure invariante quelconque μ, et pour tout k, considérons le convexe compact K_k constitué des mesures η telles que $\int f_i d\mu = \int f_i d\eta$ pour tout $i \leqslant k$. Les K_k forment une intersection décroissante de parties compactes non vides de $\mathcal{P}$, dont l'intersection est le point μ. Leur diamètre tend donc vers 0. La seconde partie du théorème est démontrée si chaque K_k contient une combinaison convexe finie de mesures invariantes extrémales, ce que nous allons maintenant montrer.

On considère l'application

$$L : \mathcal{P} \ni \eta \longrightarrow \left(\int f_1 d\eta, \dots, \int f_k d\eta\right) \in \mathbb{R}^k.$$

Son image C est un convexe compact de $\mathbb{R}^k$. Par le théorème de Minkowski, chaque point de C est une combinaison convexe de points extrémaux de C (en fait, de $k+1$ points extrémaux). En particulier,

$$L(\mu) = a_0 e_0 + \cdots + a_k e_k$$

où les e_i sont des points extrémaux de C, avec $a_i \geqslant 0$, $\sum a_i = 1$. Une légère modification de l'argument précédent montre de plus que chaque point extrémal de C est l'image d'un point extrémal de $\mathcal{P}$, donc il existe des mesures extrémales $\mu_0, \dots, \mu_k$ telles que $L(\mu_i) = e_i$. On a alors $L(a_0\mu_0 + \cdots + a_k\mu_k) = L(\mu)$ c'est à dire que $a_0\mu_0 + \cdots + a_k\mu_k \in K_k$. ►

Le résultat suivant fait un parallèle avec l'unique ergodicité.

Théorème 8.9. *La mesure m est ergodique si et seulement si $S^n f/n \longrightarrow \int f dm$ dans $L^2(X,m)$ pour tout $f \in L^2(X,m)$.*

C'est le théorème ergodique en moyenne de Von Neumann. L'énoncé est aussi vrai dans L^1, mais la démonstration dans L^2 est particulièrement simple. On commence par le

Lemme 8.10. *L'espace des fonctions f telles que $S^n f/n \longrightarrow \int f dm$ dans $L^2(X,m)$ est un fermé de $L^2(X,m)$.*

◄ Ceci découle du fait que les applications $f \mapsto S^n f/n$ et $f \mapsto \int f dm$ sont 1-Lipshitz de $L^2(X,m)$ dans lui-même. La seconde affirmation découle du calcul

$$\left|\int f dm - \int g dm\right| \leqslant \int |f-g| dm \leqslant \sqrt{\int (f-g)^2 dm} = \|f-g\|. ►$$

◄ Supposons que le critère de convergence a lieu. Alors toute fonction $f \in L^2$ qui vérifie $f \circ \varphi = f$ est égale dans L^2 (c'est à dire p.p.) à sa moyenne $\int f dm$. En particulier, toute fonction mesurable bornée vérifiant $f \circ \varphi = f$ p.p. est presque partout constante.

Réciproquement, Posons $H = L^2(X,m)$, c'est un espace de Hilbert. Soit $T : H \longrightarrow H$ l'application $f \mapsto f \circ \varphi$. L'invariance de m implique que T est unitaire, c'est à dire que

$$\langle Tf, Tg\rangle = \int (f\circ\varphi)(g\circ\varphi) dm = \int fg dm = \langle f, g\rangle.$$

L'ergodicité de m est équivalente au fait que le noyau de $T - I$ est réduit aux fonctions constantes, que l'on identifie à $\mathbb{R}$.

L'orthogonal de $\mathbb{R}$ est l'espace E des fonctions de moyenne nulle, il contient l'image de $T - I$. Montrons que cette image est dense dans E. Il suffit pour ceci de démontrer que toute fonction orthogonale à $\mathrm{im}(T-I)$ est constante. Si f est une telle fonction, alors pour tout $g \in H$ on a $\langle g, Tf\rangle = \langle g, f\rangle$ et en particulier $\langle f, Tf\rangle = \|f\|^2 (= \|Tf\|^2)$. Calculons maintenant

$$\|Tf - f\|^2 = \|Tf\|^2 + \|f\|^2 - 2\langle f, Tf\rangle = 0$$

donc $Tf = f$, donc f est constante. On a montré que le sous-espace $\mathbb{R} \oplus \mathrm{im}(T-I)$ est dense dans H.

On vérifie facilement que $S^n f \longrightarrow 0$ si $f \in \mathrm{im}(T-I)$. On termine la démonstration du théorème par un argument de densité semblable à celui qui a déjà été utilisé concernant l'unique ergodicité. Ici, il s'agit de vérifier que l'ensemble des fonctions f pour lesquelles $S^n f \longrightarrow \int f dm$ est fermé dans H. ►

En fait, les arguments de la preuve précédente montrent

Propriété 8.11. *Même en l'absence d'ergodicité, on a $H = \ker(T-I) \oplus \overline{\mathrm{im}(T-I)}$ donc $S^n f \longrightarrow \pi f$ dans L^2 pour toute fonction $f \in H$, où π est la projection orthogonale sur l'espace des fonctions invariantes.*

Revenons maintenant au fait que les mesures ergodiques sont extrémales.

◄ Soit $m = t\mu_1 + (1-t)\mu_0$ une mesure invariante, où μ_0 et μ_1 sont des mesures invariantes et $t \in]0,1[$. Si m est ergodique, les mesures μ_0 et μ_1 le sont aussi. Pour toute fonction mesurable bornée f, on a donc $S^n f/n \longrightarrow \int f dm$ dans $L^2(X,m)$. Comme $\mu_1 \leqslant m/t$ et $\mu_0 \leqslant m/(1-t)$, on déduit que $S^n f/n \longrightarrow \int f dm$ dans $L^2(X,\mu_0)$ et dans $L^2(X,\mu_1)$.

Comme par ailleurs $S^n f/n \longrightarrow \int f d\mu_0$ dans $L^2(X,\mu_0)$ et $S^n f/n \longrightarrow \int f d\mu_1$ dans $L^2(X,\mu_1)$ on a, pour toute fonction mesurable bornée $\int f d\mu_0 = \int f dm = \int f d\mu_1$, c'est à dire que $\mu_0 = m = \mu_1$. ►

Propriété 8.12. *L'application $x \mapsto 2x$ sur $\mathbb{T}$ est ergodique. Plus précisément, la mesure de lebesgue λ est ergodique pour cette application.*

Cette application n'est pas uniquement ergodique, car λ est une mesure invariante, et le dirac δ_0 en est une autre.

◄ Posons $f_k = e^{2i\pi kx}$. Les fonctions f_k constituent une base orthonormée de $H = L^2(\mathbb{T},\lambda)$. On a $Tf_k = f_{2k}$. Comme les vecteurs f_k sont orthogonaux, on obtient, pour $k \neq 0$ que

$$\|S^n f_k\| = \|f_k + Tf_k + \cdots + T^{n-1} f_k\| = \sqrt{n}$$

et donc $S^n f_k/n \longrightarrow 0$. Par ailleurs, $S^n f_0/n = f_0 \longrightarrow f_0 = 1$. On déduit que la convergence L^2 a lieu sur l'espace dense engendré par les f_k, et donc que l'application est ergodique. ►

On dit que la mesure *m* est mélangeante si

$$\int f \circ \varphi^n g dm \longrightarrow \int f dm \int g dm$$

pour toutes fonctions f et g dans $L^2(X, m)$.

9 Théorèmes Ergodiques

Le contexte de ce chapitre est celui d'un espace de probabilité (X, m) et d'une application mesurable $\varphi : X \longrightarrow X$ qui préserve la mesure m. On considère la tribu $\mathcal{I}$ des ensembles invariants, c'est à dire des ensembles A tels que $\varphi^{-1}(A) = A$.

Le premier résultat est le théorème de récurrence de Poincaré.

Théorème 9.1. *Soit A un ensemble mesurable de mesure strictement positive. Alors presque tous les points $x \in A$ reviennent dans A une infinité de fois.*

◄ Soit R l'ensemble des points de A qui ne reviennent pas dans A. Les préimages $\varphi^{-i}(R)$ sont deux à deux disjointes. En effet, si il existe $i < j$ tels que $\varphi^i(x) \in R$ et $\varphi^j(x) \in R$ Alors $\varphi^{j-i}(\varphi^i(x)) \in R \subset A$, ce qui contredit la définition de R.

Comme $\sum_{i \geqslant 0} m(\varphi^{-i}(R)) = \sum_{i \geqslant 0} m(R) < \infty$, on a $m(R) = 0$.

Soit $B = A - \cup_{i \geqslant 0} A \cap \varphi^{-i}(R)$. On a $m(B) = m(A)$, et les points de B reviennent dans A une infinité de fois. ►

On peut quantifier mieux les temps de retours. Pour tout ensemble mesurable A et tout x, on note $T_A(x)$ le premier temps $n \geqslant 1$ pour lequel $\varphi^n(x) \in A$ (on pose $T_A(x) = \infty$ si le point x ne revient pas dans A).

Théorème 9.2 (Kac)**.** *On a l'inégalité $\int_A T_A dm \leqslant 1$, avec égalité dans le cas où m est ergodique.*

◄ Soit $A_i, i \geqslant 1$ l'ensemble des points de A sur lesquels $T_A = i$. L'ensemble A est donc la réunion disjointe des ensembles A_i et d'un ensemble de mesure nulle.

On a $\varphi^{-1}(A) \cap A = A_1$, et on note $B_1 := \varphi^{-1}(A) - A_1$. On a donc $m(B_1) = m(A) - m(A_1) = \sum_{i \geqslant 2} m(A_i)$.

Soit B_i l'ensemble des points de $X - A$ sur en lesquels $T_A = i$. C'est donc le complémentaire de $\varphi^{-(i-1)}(A)$ dans $\varphi^{-i}(A)$. La réunion $B = A \cup B_1 \cup B_2 \cup \cdots$ est invariante, elle est donc de mesure 1 dans le cas ergodique.

On remarque que $\varphi^{-2}(A)$ est la réunion disjointe de A_2, de $\varphi^{-1}(A_1)$, et de B_2, donc que $m(B_2) = \sum_{i \geqslant 3} m(A_i)$. De la même façon, on a $m(B_k) = \sum_{i \geqslant k+1} m(A_i)$, donc

$$\begin{aligned} 1 \geqslant m(B) = m(A) + m(B_1) + m(B_2) + \cdots &= \sum_{i \geqslant 1} m(A_i) + \sum_{i \geqslant 2} m(A_i) + \cdots \\ = m(A_1) + 2m(A_2) + 3m(A_3) + \cdots &= \int_A T_A dm. \end{aligned}$$

L'inégalité est de plus une égalité dans le cas ergodique. ►

Propriété 9.3. *Une fonction mesurable $f : X \longrightarrow \mathbb{R}$ est invariante si et seulement si elle est mesurable par rapport à la tribu invariante $\mathcal{I}$.*

◄ Il est clair qu'une fonction Borel mesurable et invariante est $\mathcal{I}$-mesurable.

Réciproquement, soit f une fonction $\mathcal{I}$-mesurable. Alors, pour tout $a \in \mathbb{R}$, l'ensemble $\{f = a\}$ est invariant, ce qui implique que f est invariante. ►

Le cas où m est ergodique correspond au cas ou les fonctions $\mathcal{I}$-mesurables sont presque constantes (c'est à dire égales à une constante en dehors d'une ensemble de mesure nulle).

Pour toute fonction $f \in L^1(X)$ on note f_I l'espérance conditionnelle de f relativement à la tribu $\mathcal{I}$. C'est l'unique fonction invariante pour laquelle $\int_A f dm = \int_A f_I dm$ pour tout ensemble invariant A. Bien entendu, cette condition ne détermine f_I que modulo égalité p.p., c'est en ce sens qu'il faut considérer l'unicité. On supposera cependant toujours que f_I est invariante partout, et pas seulement presque partout. Dans le cas ergodique, $f_I = \int f dm$.

Pour rappel on peut déduire l'existence d'une telle fonction du théorème de Radon-Nikodym appliqué aux mesures $fm << m$ sur la tribu $\mathcal{I}$. Ce théorème nous donne en effet une densité I-mesurable pour la mesure $(fm)_{|\mathcal{I}}$ par rapport à la mesure $m_{|\mathcal{I}}$, qui est précisément la fonction f_I cherchée.

Le résultat majeur de la théorie ergodique est le Théorème Ergodique de Birkhoff que voici.

Théorème 9.4. *Soit (X, m) un espace de probabilité et φ une application mesurable de X qui préserve m. Pour toute fonction $f \in L^1(X, m)$, on a*

$$S^n f / n \longrightarrow f_I$$

presque partout.

◄ Posons $M^n f = \max(f, f + f \circ \varphi, \ldots, S^n f)$ et $M f = \sup_{n \geqslant 0} M^n f$. La fonction Mf prend ses valeurs dans $\mathbb{R} \cup \{+\infty\}$, on a

$$M^{n+1} f = \max(f + M^n f \circ \varphi, f) = f + M^n f \circ \varphi + \max(0, -M^n f \circ \varphi),$$

et donc

$$M f = f + M f \circ \varphi + \max(0, -M f \circ \varphi).$$

Lemme 9.5. *L'ensemble $A := \{Mf = +\infty\}$ est invariant, c'est à dire que $\varphi^{-1}(A) = A$.*

◄ On le déduite de l'identité ci-dessus car $\max(0, -Mf\circ\varphi)$ est positif et fini en tout x. ►

Lemme 9.6. *On a $\int_A f \geqslant 0$.*

◄ En posant $g_n := \max(0, -M^n f\circ\varphi)$, on a, pour tout n

$$\int f dm = \int_A M^{n+1} f dm - \int_A M^n f\circ\varphi dm - \int_A g_n dm$$
$$= \int_A M^{n+1} f - M^n f dm - \int_A g_n dm \geqslant \int_A g_n dm.$$

Dans ce calcul, on a utilisé d'abord l'invariance de la mesure m, puis le fait évident que $M^{n+1}f \geqslant M^n f$. Comme g_n décroit vers 0, le théorème de convergence monotone implique que $\int g_n dm$ décroit vers 0, on obtient l'inégalité souhaitée à la limite. ►

Pour terminer la preuve du théorème, on applique ce qui précède à la fonction $h = f - f_I - \epsilon$. En notant encore A l'ensemble $\{Mh = \infty\}$, on déduit que $\int_A h dm \geqslant 0$. Comme A est invariant, on a d'autre part $\int_A h = -\epsilon m(A)$, et donc $m(A) = 0$. On a donc $Mh < \infty$ presque partout, ce qui implique que $\limsup S^n h/n \leqslant 0$ presque partout. On a donc $\limsup S^n f/n \leqslant \epsilon + f_I$ presque partout, et comme ceci est vrai pour tout $\epsilon > 0$, on obtient

$$\limsup S^n f/n \leqslant f_I.$$

L'inégalité opposée s'obtient de la même façon (ou en appliquant ce qui précède à la fonction $-f$). Le théorème ergodique est prouvé. ►

Supposons que φ admet une inverse mesurable. Cette inverse préserve alors la mesure m, et le théorème ergodique s'applique dont aussi en temps négatif :

$$S^{-n} f/n \longrightarrow f_I$$

presque partout. Il est intéressant de remarquer en particulier que

$$\lim_{n\longrightarrow+\infty} S^n f(x)/n = \lim_{n\longrightarrow+\infty} S^{-n} f(x)/n$$

pour presque tout point x, ce qui n'est pas une évidence.

Corollaire 9.7. *Deux mesures invariantes ergodiques différentes d'une même application mesurable sont mutuellement singulières. En conséquence, toute mesure invariante ergodique est extrémale.*

◄ Soient μ_0 et μ_1 deux mesures invariantes ergodiques. Pour toute fonction f mesurable bornée, on a $S^n f/n \longrightarrow \inf f d\mu_0$ sur un ensemble mesurable A_0 qui vérifie $\mu_0(A_0) = 1$. On a aussi $S^n f/n \longrightarrow \inf f d\mu_1$ sue un ensemble mesurable A_1 qui vérifie $\mu_1(A_1) = 1$. Si les mesures μ_0 et μ_1 sont différentes, alors il existe une fonction f pour laquelle $\int f d\mu_0 \neq \int f d\mu_1$, les ensembles A_0 et A_1 correspondant sont disjoints.

Soit $m = t\mu_1 + (1-t)\mu_0$ une mesure invariante, où μ_0 et μ_1 sont des mesures invariantes et $t \in]0, 1[$. Si m est ergodique, les mesures μ_0 et μ_1 le sont aussi. Mais les mesures m et μ_0 ne peuvent pas être mutuellement singulières, ce qui est une contradiction. ►

On déduit de la convergence presque partout des résultats de convergence en moyenne. On commence par un lemme utile sur l'espérance conditionnelle :

Lemme 9.8. *Soit $p \in [1, \infty)$. Pour tout $f \in L^p$, on a $|f_I|^p \leqslant (|f|^p)_I$. En conséquence, $f_I \in L^p$ et $\|f_I\|_p \leqslant \|f\|_p$.*

C'est un cas particulier de l'inégalité de Jensen, qui affirme que $(g\circ f)_I \leqslant g\circ(f_I)$ pour toute fonction convexe g.
◄ La fonction $t \longmapsto |t|^p$ est convexe, et s'écrit comme un supremum de de fonctions linéaires,

$$|t|^p = p \max_{s\in\mathbb{R}} st - a(s) \quad , \quad |t|^p = p \sup_{s\in\mathbb{Q}} st - a(s)$$

où $a(s) = |s|^{q-1}/q$ (mais la valeur n'est pas importante). On a donc $|f|^p = p\max_s sf - a(s)$. Pour tout tout x et tout s, on en déduit que $|f|^p_I(x) \geqslant psf_I(x) - pa(s)$ presque partout

$$|f|^p_I(x) \geqslant p \sup_s sf_I(x) - a(s) = |f_I(x)|^p. \text{ ►}$$

Dans le cas $p = 2$, on comprend bien le résultat ci-dessus en remarquant que f_I est le projeté orthogonal de f sur le sous-espace fermé $L^2(X, I, m) \subset L^2(X, m)$ des fonctions invariantes de carré intégrable. En effet, en notant g ce projeté, on a que $\langle g, h\rangle = \langle f, h\rangle$ pour toute $h \in L^2(X, I, m)$. En prenant pour h la fonction caractéristique d'un ensemble invariant A, on déduit que $\int_A g = \int_A f$, ce qui montre que $g = f_I$.

Théorème 9.9. *Soit $p \in [1, \infty)$ et $p \in L^p(X, m)$, on a $S^n f/n \longrightarrow f_I$ dans L^p.*

◄ Si f est mesurable bornée, alors la convergence presque partout de la suite equibornée $S^n f$ vers f_I implique la convergence dans tous les L^p. Le résultat découle donc de l'observation que l'ensemble G des fonctions f satisfaisant la conclusion de l'énoncé est fermé dans L^p. Soit en effet f une fonction de $\bar{G}$. Alors pour tout $\epsilon > 0$ il existe une fonction $g \in G$ telle que $\|f - g\|_p < \epsilon/3$. On a alors

$$\begin{aligned}\|S^n f/n - f_I\|_p &\leqslant \|S^n f/n - S^n g/g\|_p + \|S^n g/n - g_I\|_p + \|g_I - f_I\|_p \\ &\leqslant 2\|f - g\|_p + \|S^n g/n - g_I\|_p \leqslant \epsilon\end{aligned}$$

pour n assez grand. ►

Propriété 9.10. *L'application $x \mapsto 2x$ sur $\mathbb{T}$ est ergodique. Plus précisément, la mesure de lebesgue λ est ergodique pour cette application.*

Cette application n'est pas uniquement ergodique, car λ est une mesure invariante, et le dirac δ_0 en est une autre.
◄ Posons $f_k = e^{2i\pi kx}$. Les fonctions f_k constituent une base orthonormée de $H = L^2(\mathbb{T}, \lambda)$. On a $f_k \circ \varphi = f_{2k}$. Comme les vecteurs f_k sont orthogonaux, on obtient, pour $k \neq 0$ que

$$\|S^n f_k\| = \|f_k + f_k \circ \varphi + \cdots + f_k \circ \varphi^{n-1}\| = \sqrt{n}$$

et donc $S^n f_k/n \longrightarrow 0$. Par ailleurs, $S^n f_0/n = f_0 \longrightarrow f_0 = 1$. On déduit que la convergence L^2 a lieu sur l'espace dense engendré par les f_k, et donc que l'application est ergodique. ►

On a utilisé le critère d'ergodicité suivant :

Propriété 9.11. *Si il existe une partie dense $F \subset L^2(X, m)$ telle que la suite $S^n f/n \longrightarrow \int f dm$ pour tout $f \in F$, alors la mesure m est ergodique.*

En préparation de la définition suivante, mentionnons aussi :

Propriété 9.12. *La mesure m est ergodique si et seulement si la suite $S^n f/n$ converge faiblement vers $\int f dm$ dans L^2 pour toute $f \in L^2$.*

◄ L'ergodicité implique la convergence forte dans de $S^n f/n$ comme nous l'avons montré, donc sa convergence faible.

Réciproquement, soit f une fonction invariante. Le critère implique que $\int f(f \circ \varphi^n) dm = \int f^2 dm \longrightarrow \left(\int f\right)^2$, et donc que f est presque partout constante. ►

En fait, dans l'exemple ci-dessus, on a la propriété plus forte suivante :

Définition 9.13. *On dit que la mesure m est mélangeante si $f \circ \varphi^n$ converge vers $\int f dm$ faiblement dans L^2 pour tout $f \in L^2$, c'est à dire si*

$$\int g(f \circ \varphi^n) dm \longrightarrow \int g dm \int f dm$$

pour tous f et g dans L^2.

Il suffit de vérifier que cette convergence a lieu sur une partie engendrant un sous-espace dense de L^2, et notamment sur les fonctions caractéristiques des ensembles. Le mélange est donc équivalent à la propriété

$$m(\varphi^{-n}(A) \cap B) \longrightarrow m(A)m(B),$$

ce qui s'interprète comme une propriété d'indépendance asymptotique des ensembles B et $\varphi^{-n}(A)$. Il est souvent très utile dans les exemples de remarquer que la convergence

Exercice 9.1. *Montrer que la mesure m est ergodique si et seulement si la suite $m(\varphi^{-n}(A) \cap B)$ converge en moyenne de Césaro vers $m(A)m(B)$ pour tous ensembles mesurables A et B.*

Il suffit que cette convergence ait lieu lorsque A et B appartiennent à une algèbre d'ensembles engendrant la tribu.

Revenons maintenant au cas de l'application $x \mapsto 2x$. Comme $\int f_k f_l dm$ est nul sauf si $k = l$, la suite $\int (f_l)(f_k \circ \varphi^n)$ est nulle à partir d'un certain rang, et donc elle converge vers 0. On a ainsi montré la convergence désirée sur une partie dense, et donc la propriété de mélange. De manière générale :

Propriété 9.14. *Toute mesure mélangeante est ergodique.*

◄ Soit A un ensemble invariant, le mélange implique que $m(A) = m(A)^2$ et donc $m(A) = 0$ ou 1. ►

Le théorème ergodique de Birkhoff implique la loi forte des grands nombres.

Rappelons en le contexte. On se donne une suite de variables aléatoires intégrables identiquement distribuées et indépendantes $f_n, n \geqslant 0$, et le théorème affirme que $(f_1 + \cdots + f_n)/n \longrightarrow \int f$ presque partout.

En notant P la loi des fonctions f_i, on peut modéliser cette situation en considérant l'espace de probabilité $X = \mathbb{R}^{\mathbb{N}}$, muni de la mesure produit, les fonctions f_i étant les coordonnées des points. Si $\sigma : X \longrightarrow X$ est le décalage, on a alors $f_i = f \circ \varphi^i$, en posant $f = f_0$.

La loi forte des grands nombres découle alors du théorème ergodique de Birkhoff pour peu que l'ergodicité de la mesure produit m (relativement à l'application σ) soit démontrée.

Propriété 9.15. *Soit (Y, P) un espace de probabilité et (X, m) l'espace des suites indépendantes d'éléments de Y, c'est à dire que $X = Y^{\mathbb{N}}$ et m est la mesure produit. Alors la mesure m est ergodique, et même mélangeante pour le décalage σ.*

Soit $\mathcal{A}$ la tribu asymptotique $\mathcal{A} := \cap_{n \geqslant 0} \tau(x_n, x_{n+1}, \ldots)$, où $\tau(x_n, x_{n+1}, \ldots)$ est la tribu engendrée par les fonctions $x_i, i \geqslant n$. La loi du 0 ou 1 de Kolmogorov affirme que la tribu asymptotique est triviale, c'est à dire que ses éléments ont tous mesure 0 ou 1 pour la mesure produit m.

La démonstration est la suivante : La tribu asymptotique est indépendante des tribus $\tau(x_1, \ldots, x_n)$ pour tout n, c'est à dire que $m(A \cap B) = m(A)m(B)$ pour tout A asymptotique et $B \in \tau(x_1, \ldots, x_n)$. Pour A fixé, les mesures positives $B \mapsto m(A \cap B)$ et $B \mapsto m(A)m(B)$ sont égales sur l'algèbre $\cup \tau(x_1, \ldots, x_n)$, donc par le lemme de classe monotone elles sont égales sur toute la tribu produit. Comme A est mesurable pour cette tribu, on conclut que $m(A) = m(A)^2$, donc $m(A) \in \{0, 1\}$.

L'ergodicité de la mesure produit résulte alors de la remarque suivante :

Propriété 9.16. $\mathcal{I} \subset \mathcal{A}$.

◄ Soit A un ensemble mesurable invariant. Il appartient à la tribu $\tau(x_1, x_2, \ldots)$. L'ensemble $A = \varphi^{-1}(A)$ appartient donc à la tribu $\tau(x_1 \circ \varphi, x_2 \circ \varphi, \ldots) = \tau(x_2, x_3, \ldots)$. Par récurrence, A appartient à chacune des tribus $\tau(x_n, x_{n+1}, \ldots)$, et donc il appartient à la tribu asymptotique. ►

La tribu asymptotique est plus grande que la tribu invariante. Pour s'en convaincre, on fixe un point y de X et on considère l'événement A constitué des suites x_n telles que $x_{2n} = y$ pour n assez grand. La loi du 0 ou 1 est donc plus forte que la simple ergodicité.

La loi du 0 ou 1 conduit plus généralement à la notion de mesure exacte :

Définition 9.17. *Dans le cas général d'une application mesurable $\varphi : X \longrightarrow X$ préservant la mesure m, on dit que m est exacte si la tribu asymptotique $\mathcal{A} := \cap_{n \in \mathbb{N}} \varphi^{-n}(\tau)$ est triviale, c'est à dire que ses ensembles ont mesure 0 ou 1, où τ est la tribu de X.*

Le caractère mélangeant de la mesure produit résulte donc du résultat suivant :

Propriété 9.18. *Toute mesure exacte est mélangeante, donc ergodique.*

◄ Remarquons pour commencer que, comme plus haut, la tribu invariante est contenue dans la tribu asymptotique. Toute mesure exacte est donc clairement ergodique.

Pour démontrer le mélange, on considère l'espace de Hilbert $H = L^2(X, m)$, et les sous-espaces fermés $H_n = L^2(X, \varphi^{-n}(\tau), m)$. Ces sous-espaces forment une suite décroissante, dont l'intersection est $L^2(X, \mathcal{A}, m)$. L'hypothèse d'exactitude nous dit que cette intersection est l'ensemble des fonctions constantes, identifié à $\mathbb{R}$.

Posons $F_n = H_n \cap H_{n+1}^{\perp}$, c'est donc l'orthogonal de H_{n+1} dans H_n. Posons $G_n = \mathbb{R} \oplus F_1 \oplus F_2 \oplus \cdots \oplus F_n$ et $G = \cup_n G_n$. Le sous-espace G est dense dans H.

L'application $f \mapsto f \circ \varphi^n$ envoie H sur H_n, de sorte que $\int g(f \circ \varphi^n) dm = \int f \int g$ à partir d'un certain rang pour $f, g \in G_n$. Par l'argument classique de densité, le mélange en découle. ►

Nous avons notamment montré ci-dessus la propriété de mélange dans les deux cas suivants : Application $x \mapsto 2x$ et mesure de Lebesgue, et décalage sur $\{0, 1\}^{\mathbb{N}}$ et mesure produit avec $P(0) = (1) = 1/2$. Les démonstration proposées sont très différentes, mais les résultats son équivalents. En effet, l'application

$$\psi : \{0, 1\}^{\mathbb{N}} \ni (a_i) \mapsto \sum a_i 2^{-i-1} \in \mathbb{T}$$

conjugue le décalage et la multiplication. De plus, l'image $\psi_* m$ de la mesure produit m sur $\{0,1\}^{\mathbb{N}}$ est la mesure de Lebesgue λ sur $\mathbb{T}$.

Exercice 9.2. *Soient σ et φ deux applications mesurables des espaces de probabilités (X, m) et (Y, λ). Soit $\psi : X \longrightarrow Y$ une application mesurable telle que $\psi_* m = \lambda$, et telle que $\varphi \circ \psi = \psi \circ \varphi$. Si m est ergodique ou mélangeante, montrer que λ a la même propriété.*

Soit $Z \subset Y$ une partie invariante de mesure totale. Montrer que $\varphi_{|Z}$ est ergodique ou mélangeante pour la mesure $m_{|Z}$ si et seulement si φ l'est.

Pour démontrer la réciproque, il faut utiliser le fait que ψ est presque une bijection. Si $Z \subset \mathbb{T}$ est l'ensemble des points irrationnels. C'est un ensemble invariant de mesure totale. La restriction de φ à Z est ergodique et mélangeante.

alors chaque point de Z a une seule préimage dans $\{0,1\}^{\mathbb{N}}$, et l'application ϕ qui a un point associe sa préimage est mesurable. Comme $\lambda(Z) = 1$,

10 Décomposition Ergodique

Voir Ricardo Mañé *Ergodic Theory and Differentiable Dynamics*, chapitre II.6.

On se donne un espace métrique compact X muni de sa tribu Borélienne ainsi qu'une application mesurable $\varphi : X \longrightarrow X$. On ne se donne pas a priori une mesure de probabilité invariante, mais les résultats qui suivent sont vides si il n'existe pas de probabilité invariante. On peut par exemple penser au cas où φ est continue. On va s'intéresser en particulier au comportement asymptotique de la suite des mesure empiriques $S_*^n \delta_x / n$. Pour fixer les idées, on commence par une conséquence directe du théorème ergodique :

Proposition 10.1. *Soit m est une mesure invariante. La mesure m est ergodique si et seulement si les mesures empiriques $S_*^n \delta_x / n$ convergent vers m pour m-presque tout x.*

◄ On considère une partie dénombrable dense $F \subset C(X)$. Pour tout $f \in F$, le théorème ergodique de Birkhoff implique que $\int f d(S_*^n \delta_x) = S^n f(x)/n$ converge vers $\int f dm$ en dehors d'une ensemble Z_f tel que $m(Z_f) = 0$. L'ensemble $Z = \cup_{f\in F} Z_f$ est de mesure nulle, et $\int f d(S_*^n \delta_x) \longrightarrow \int f dm$ pour tout $f \in F$ si $x \in Z$. Pour ces points x, on en conclut que $S_*^n \delta_x / n \longrightarrow m$.

Réciproquement, supposons que $S_*^n \delta_x / n \longrightarrow m$ pour m-presque tout x. Alors pour toute fonction continue, on a $S^n f(x)/n \longrightarrow \int f dm$ dans L^2. Comme $C(X)$ est dense dans L^2, ceci implique l'ergodicité. ►

Théorème 10.2. *Considérons l'ensemble A des points x admettant la propriété suivante :*

La suite $S_*^n \delta_x / n$ converge vers une probabilité invariante m_x, qui est ergodique et dont le support contient x.

Alors l'ensemble A est de mesure totale pour toute mesure invariante.

Si φ est un homéomorphisme, on obtient la même conclusion en définissant A comme l'ensemble des points x ayant la propriété suivante : Les suites $S_*^n \delta_x / n$ et $S_*^{-n} \delta_x / n$ convergent vers une même probabilité invariante ergodique m_x, dont le support contient x.

Exercice 10.1. *Montrer que les points de A sont récurrents.*

On considère la suite d'applications mesurables $S_*^n \delta_x / n$ à valeurs dans l'espace métrique compact $\mathcal{P}(X)$. Des arguments généraux de théorie de la mesure impliquent que l'ensemble R sur lequel cette suite converge est Borélien, et que la limite $x \longmapsto m_x := \lim S_*^n f(x)/n$ est mesurable sur R (en tant que fonction de R dans $\mathcal{P}(X)$).

Lemme 10.3. *L'ensemble Borélien R sur lequel $m_x := \lim S_*^n \delta_x / n$ existe est un ensemble mesurable invariant qui est de mesure totale pour toute probabilité invariante m. De plus, on a $m_{\varphi(x)} = m_x$ pour tout $x \in R$.*

◄ Comme ci-dessus, on considère une partie dénombrable dense $F \subset C(X)$. Le théorème ergodique de Birkhoff implique qu'il existe un ensemble Z de mesure totale sur lequel $\int f d(S_*^n \delta_x) = S^n f(x)/n$ converge pour tout $f \in F$. Par un argument de densité, on montre alors que la suite $S^n f(x)/n$ est de Cauchy pour tout $x \in Z$ et $f \in C(X)$, et donc convergente. L'application linéaire $f \longmapsto \lim S^n f(x)/n$ est alors représentée par une mesure m_x, qui est la limite de la suite $S_*^n \delta_x$. On a donc $Z \subset R$. ►

Lemme 10.4. *Pour toute fonction mesurable bornée f et toute mesure invariante m, on a $\int f dm_x = f_I(x)$ m-presque partout.*

On posera $\tilde{f}(x) := \int f dm_x$ pour toute fonction mesurable bornée f. La fonction $\tilde{f}$ est bien définie et mesurable sur l'ensemble invariant R, c'est un représentant de l'espérance conditionnelle f_I pour toute mesure invariante m. On étendra de plus $\tilde{f}$ en une fonction invariante définie sur tout X, par exemple en posant $\tilde{f} = 0$ hors de R. La notation $\tilde{f}$ est plus précise que la notation f_I car c'est une fonction qui est bien définie sur un ensemble R indépendant de f, et qui ne dépend pas du choix d'une mesure invariante m.

◄ Il suffit de démontrer que la fonction $x \longmapsto \int f dm_x$ vérifie la propriété qui caractérise f_I, c'est à dire qu'elle est mesurable et que $\int_A f dm = \int_A \left(\int f dm_x \right) dm(x)$ pour tout borélien invariant A. Le Borélien invariant A étant fixé, notons E l'espace des fonctions mesurables bornées f qui vérifient ces propriétés.

On a $C(X) \subset E$. En effet si $f \in C(X)$, la fonction $x \longmapsto \int f dm_x$ est mesurable comme composition de l'application mesurable $x \longmapsto m_x$ et de la fonction continue $m \longmapsto \int f dm$. De plus $S^n f(x)/n \longrightarrow \int f dm_x$ sur l'ensemble R, et donc m-presque partout. Par le théorème de convergence dominée, on obtient

$$\int_A f dm = \int_A (S^n f/n) dm = \int_A \left(\int f dm_x \right) dm(x).$$

Le théorème de convergence monotone implique que E est stable par convergence monotone.

Comme les fonctions indicatrices des ouverts et des fermés sont des limites monotones de fonctions continues, elles sont dans E. On considère alors l'ensemble C des parties Y de X telles que $1_Y \in E$. La stabilité par convergence monotone de E implique que C est stable par réunion croissante, elle est aussi stable par complémentation : c'est une classe monotone. Le lemme de classes monotone implique que toute classe monotone contenant les fermés contient les Boréliens. On obtient donc que C contient tous les Boréliens. L'espace E contient donc toutes les fonctions mesurables à image finie, et donc toutes les fonctions mesurables bornées puisqu'une telle fonction est limite croissante de fonctions mesurables à image finie. ►

Au vu du théorème de Birkhoff, on déduit :

Corollaire 10.5. *Soit f une fonction mesurable bornée invariante, l'ensemble des points x tels que $f(x) = \int f dm_x$ est de mesure totale pour toute probabilité invariante.*

Dans le cas où φ est continue, la mesure m_x est invariante pour chaque $x \in R$. Dans le cas général, on a :

Lemme 10.6. *L'ensemble $R_1 \subset R$ des points x tels que la mesure m_x est invariante est un Borélien invariant de mesure totale pour toute probabilité invariante.*

◄ Pour toute fonction $f \in C(X)$, la fonction $f\circ\varphi$ est mesurable et bornée, donc les deux égalités $\int (f\circ\varphi) dm_x = \lim S^n f(\varphi(x))/n$ et $\int f dm_x = \lim S^n f(x)/n$ sont satisfaites sur un ensemble mesurable S_f qui est de mesure totale pour toute mesure invariante. Comme $\lim S^n f(\varphi(x))/n = \lim S^n f(x)/n$, on déduit que $\int f\circ\varphi dm_x = \int f dm_x$ sur S_f. Soit F une partie dense dénombrable de $C(X)$. On a $\int f\circ\varphi dm_x = \int f dm_x$ pour tout $f \in F$ et tout x dans $S_F := \cap_{f\in F} S_f$, qui est un ensemble mesurable de mesure totale. Pour $x \in S_F$, on en déduit que l'égalité a lieu pour toute fonction continue f, et donc que $\varphi_* m_x = m_x$. ►

Lemme 10.7. *Soit $F \subset C(X)$ une partie dense. Soit m une mesure de probabilité invariante qui a la propriété que, pour toute fonction $f \in F$, la fonction f_I est constante m-presque partout. Alors la mesure m est ergodique.*

Lemme 10.8. *L'ensemble $R_2 \subset R_1$ des points x tels que la mesure m_x est invariante et ergodique est de mesure totale pour toute mesure invariante.*

En particulier, l'existence d'une mesure invariante implique l'existence d'une mesure invariante ergodique, ce qui généralise l'énoncé prouvant l'existence de mesures ergodiques dans le cas continu.
◄ Soit f une fonction mesurable bornée invariante. Soit $R_2(f)$ l'ensemble des points x tels que

$$\int f^2 dm_x = f^2(x) \quad \text{et} \quad \int f dm_x = f(x).$$

Comme f et f^2 sont des fonctions mesurables bornées, l'ensemble $R_2(f)$ est un Borélien invariant de mesure totale pour toute probabilité invariante (au vu du corollaire ci-dessus). Pour $x \in R_2(f)$, on a $(\int f dm_x)^2 = \int f^2 dm_x$, donc la fonction f est m_x-presque constante et égale à sa moyenne $\int f dm_x = f(x)$.

On considère maintenant une partie dense $F \subset C(X)$. L'énoncé découle de l'affirmation $R_2 = \cap_{f\in F} R_2(\tilde{f})$. En effet, si x est dans cette intersection et $f \in F$, alors $\tilde{f}(y) = \tilde{f}(x)$ m_x-presque partout, c'est à dire que $S^n f(y)/n \longrightarrow \tilde{f}(x)$ pour m_x presque tout y. Comme F est dénombrable, il existe donc un ensemble Y tel que $m_x(Y) = 1$ et $S^n f(y)/n \longrightarrow \tilde{f}(x)$ pour tout $f \in F$ et $y \in Y$. Par un argument de densité classique, on déduit que $S^n_* \delta_y/n \longrightarrow m_x$ pour tout $y \in Y$. On a vu que ceci implique l'ergodicité de m_x. ►

Proposition 10.9. *Soit A un Borélien qui est de mesure nulle pour toute probabilité invariante ergodique, alors A est de mesure nulle pour toute probabilité invariante μ.*

◄ On a $\mu(A) = \int_X m_x(A) d\mu(x)$. Comme m_x est ergodique pour μ-presque tout x, on a $m_x(A) = 0$ pour μ presque tout x, donc $\mu(A) = 0$. ►

◄ Finissons la démonstration du théorème 10.2. On a déjà montré qu'il existe un Borélien invariant R_2, qui est de mesure totale pour toute probabilité invariante, et tel que $S^n\delta_x/ \longrightarrow m_x$ pour tout $x \in R_2$, et m_x est une probabilité invariante ergodique pour tout $x \in R_2$. De plus, l'application $x \longmapsto m_x$ est mesurable sur R_2. On étudie maintenant le support des mesures $m_x, x \in R_2$.

On considère une suite de fonctions C^1 décroissantes $g_k : [0, \infty) \longrightarrow [0, 1]$ telles que $g_k = 1$ sur $[0, 1/k]$ et $g_k(2/k) = 0$. Le point x appartient au support de la mesure de probabilité m si et seulement si

$$\int g_k \circ d(x, y) dm(y) > 0$$

pour tout k. Chacune de ces inégalités définit un ouvert U_k de $X \times \mathcal{P}(X)$. En effet si $h(x, y)$ est une fonction Lipschitz sur $X \times X$, alors la fonction

$$X \times \mathcal{P}(X) \ni (x, m) \mapsto \int_X h(x, y) dm(y)$$

est continue. Ceci découle de l'inégalité $\left|\int h(x, y)dm(y) - \int h(x, y')dm'(y)\right| \leqslant k|y' - y| + \int h(x, y)d(m' - m)(y)$. L'intersection $B := \cap_k U_k$, qui est l'ensemble des couples (x, m) tels que le point x appartient au support de m, est donc une partie mesurable de $X \times \mathcal{P}(X)$. Comme l'application $x \mapsto (x, m_x)$ est mesurable, L'ensemble A des points $x \in R_2$ tels que x est dans le support de m_x est un Borélien.

Nous avons démontré que A est mesurable. Si m est une probabilité invariante ergodique, alors m presque tout point x a la propriété que $m_x = m$ et que x est dans le support de m, donc $m(A) = 1$. Comme ceci est vrai pour toute probabilité invariante ergodique, c'est aussi vrai pour toute probabilité invariante, par la proposition.

Si m est un mesure ergodique, alors m-presque tout point x a la propriété que $S^n_* \delta_x \longrightarrow m$ et que x est dans le support de m. On conclut que $m(A) = 1$. Comme A est de mesure totale pour toute probabilité invariante ergodique, il est de mesure totale pour toute probabilité invariante.

Dans le cas inversible, Les mesures $m^+_x := m_x$ et $m^-_x := \lim S^{-n}_* \delta_x / n$ sont bien définies et dépendent mesurablement de x sur le Borélien $A(\varphi) \cap A(\varphi^{-1})$, qui est de mesure totale pour toute mesure invariante. L'ensemble

$$A_1 := \{x \in A(\varphi) \cap A(\varphi^{-1}) : m^+_x = m^-_x\}$$

est borélien. Si m est une probabilité invariante ergodique, on déduit de théorème ergodique que $m^+_x = m^-_x = m$ pour m-presque tout x. Autrement dit, $m(A_1) = 1$ pour toute probabilité invariante ergodique, et donc $m(A_1) = 1$ pour toute probabilité invariante. ►

Pour interpréter les résultats ci-dessus, considérons l'application $\varphi(x) = x + x(1 - x)/2$, sur $X = [0, 1]$. La dynamique est très simple, il y a deux points fixes 0 et 1. Toutes les autres orbites sont strictement croissantes et convergent vers 1. Il y a seulement deux mesures de probabilité ergodiques, qui sont δ_0 et δ_1. Les mesures de probabilités invariantes sont donc les mesures $m = (1 - t)\delta_0 + t\delta_1$, $t \in [0, 1]$. Comme $\varphi^n(x) \longrightarrow 1$ pour tout $x \in]0, 1]$, on a a $S^n_* \delta_x \longrightarrow \delta_1$ pour ces valeurs de x, et bien sur $S^n_* \delta_0 \longrightarrow \delta_0$. On déduit que $R_2 = X$. Les mesures m_x sont définies et ergodiques pour tout x.

Cependant, on a $m^-_x = m^+_x$ seulement si $x \in \{0, 1\}$ (qui est invariant et de mesure totale). Les points x qui appartiennent au support de m_x sont aussi les points de $\{0, 1\}$.

Considérons le shift bilatère φ sur l'espace produit $X = Y^{\mathbb{Z}}$ où (Y, P) est un espace de probabilité. Contrairement au shift sur $X^{\mathbb{N}}$, l'application φ est inversible. Nous allons montrer l'ergodicité du shift

On considère les deux projections π^+ et π^- de $Y^{\mathbb{Z}}$ sur $Y^{\mathbb{N}}$ où π^+ est la projection sur les facteurs correspondant aux indices positifs, et $\pi^-(x_n) = \pi^+(x_{-n})$. La projection π^+ conjugue φ avec le shift unilatère σ, et la projection π^- conjugue l'inverse φ^{-1} avec le shift unilatère. On considère la mesure produit μ sur X et la mesure produit μ^+ sur $Y^{\mathbb{N}}$. On a $\pi^{\pm}_* \mu = \mu^+$.

L'ergodicité du décalage unilatère donne l'existence d'un borélien $A^+ \subset Y^{\mathbb{N}}$ tel que $\mu^+(A^+) = 1$ et tel que $(\delta_x + \delta_{\sigma(x)} + \cdots + \delta_{\sigma^{n-1}(x)})/n \longrightarrow \mu^+$. L'ensemble $B^+ = (\pi^+)^{-1}(A^+)$ vérifie donc $\mu(B^+) = 1$.

11 Dynamique symbolique

Soit Y un ensemble fini à k éléments, on prendra en général $Y = \{1, \dots, k\}$. On considère l'espace métrisable compact $X = Y^{\mathbb{N}}$ des mots infinis et le décalage à k symboles $\sigma : X \longrightarrow X$. On munit X de la distance $d(x, y) = 2^{-i}$ où i est le premier indice pour lequel $x_i \neq y_i$. C'est une distance ultramétrique, c'est à dire que $d(x, z) \leqslant \max(d(x, y), d(y, z))$. Le diamètre de X pour cette distance est 1. De plus, si $d(x, x') < 1$, alors $d(\sigma(x), \sigma(x')) \geqslant 2d(x, x')$. Plus précisément, le compact X est la réunion disjointe de k parties compactes et ouvertes X_i (les suites commençant par i) et $\sigma : X_i \longrightarrow X$ est un homéomorphisme qui dilate les distances d'un facteur 2.

On appelle cylindre les ensembles C de la forme $\pi_l^{-1}(A)$, où $\pi_l : X \longrightarrow Y^{l+1}$ est la projection sur les $l + 1$ premiers facteurs et où A est une partie quelconque de Y^{l+1}. Les cylindres sont ouverts et fermés. Ils constituent une base dénombrable d'ouverts, et une algèbre d'ensembles qui engendre la tribu produit (par définition de la tribu produit). Cette tribu produit est aussi la tribu Borélienne associée à la topologie produit. En effet, comme les cylindres sont fermés, la tribu engendrée par les cylindres est plus grande que la tribu borélienne. Réciproquement, comme tout ouvert est une réunions dénombrable de cylindres, tout ouvert est mesurable pour la tribu produit, donc la tribu produit est plus petite que la tribu borélienne.

On dit que C est un cylindre élémentaire si il est de la forme $\pi_l^{-1}(x)$ pour un point x de Y^{l+1}

L'application σ a de nombreux compacts invariants. Il y a k points fixes, qui sont les suites constantes. Il y a des orbites périodiques de toutes périodes.

Exercice 11.1. *Montrer que le décalage σ a k^T orbites de période T (mais T n'est pas forcément la période minimale).*

On dit que $\Sigma \subset X$ est de type fini si il existe un nombre fini de mots finis tels que Σ est l'ensemble des mots ne contenant pas ces mots finis. Autrement dit, Σ est de type fini si et seulement si il existe des cylindres élémentaires $C_1, \dots, C_l$ tels que

$$\Sigma = X - \bigcup_{i \geqslant 0, j \in \{1, \dots, l\}} \sigma^{-i}(C_j).$$

Toute partie de type fini est un compact invariant par σ. On vérifie facilement que Σ est de type fini si et seulement si il existe un cylindre C tel que $\Sigma = \cup_{i \geqslant 0} \sigma^{-i}(C)$.

Étudions en détails le cas d'un ensemble de type fini engendré par des mots interdits de longueur 2. Il existe alors une matrice carrée A de taille $k \times k$, à coefficients positifs, tel que $A(i, j) = 0$ si et seulement si le mot ij est interdit. On peut poser $A(i, j) = 1$ pour les mots ij autorisés, mais ce n'est pas nécessaire. Autrement dit, on considère l'ensemble $\Sigma_A \subset \{1, \dots, k\}^{\mathbb{N}}$, constituée des suites y_n telles que $A(y_n, y_{n+1}) > 0$ pour tout n. Cette partie est invariante par le décalage σ, la restriction σ_A de σ à Σ_A est le décalage de type fini associé à A.

La matrice A est dite transitive, ou irréductible, si pour tous états i et j, il existe $n \geqslant 0$ tel que $A^n(i, j) > 0$. Cela signifie qu'il existe un mot prenant la valeur i puis la valeur j. Il est équivalent de demander qu'il existe $n \geqslant 0$ tel que $S^n A := I + A + \cdots + A^{n-1}$ est à coefficients strictement positifs.

Proposition 11.1. *Le décalage de type fini associé à la matrice A est transitif si et seulement si la matrice A est transitive.*

◄ Si la matrice A n'est pas transitive, il existe i et j tels que $A^n(i, j) = 0$ pour tout n. Soit X_i est le cylindre $\pi_0^{-1}(i)$, où π_0 est la projection sur le premier facteur. Il n'existe aucune suite prenant la valeur i puis la valeur j, donc il n'existe aucune orbite partant de C_i et passant par C_j.

Supposons maintenant que A est transitive, et soient C et C' deux cylindres, déterminés par des suites finies admissibles x, x' avec $x = x_0 \cdots x_n$ et $x' = x'_0 \cdots x'_m$. Par transitivité de A, il existe une suite finie y telle que $y_0 = x_k$ et $y_l = x'_1$. Notons $\sigma(y)$ le mot $y_1 y_2 \cdots y_l$. Soit alors z un élément de Σ_A commençant par $x\sigma(y)\sigma(x')$ (juxtaposition des mots finis). La suite z commence par x, donc elle est dans le cylindre C. De plus, la suite $\sigma^{n+l-1} z$ commence par x', c'est donc un élément de C'. Dans le cas où $x_n = x'_0$, on prend $z = x\sigma(x')$. ►

On rappelle qu'une application continue φ est topologiquement mélangeante si, pour tous ouverts U et V, il existe N tel que $\varphi^{-n}(U) \cap V$ est non vide pour tout $n \geqslant N$. Si A est la matrice de permutation circulaire des coordonnées, alors le décalage σ_A est transitif, mais pas topologiquement mélangeant. Pour une telle matrice, tous les états ont une période égale à k.

Proposition 11.2. *Le décalage de type fini associé à la matrice A est topologiquement mélangeant si et seulement si la matrice A est apériodique.*

◄ Comme dans la preuve précédente, si il existe i et j tel que $A^n(i, j) = 0$ pour une infinité de valeurs de n, alors le système n'est pas mélangeant puisque pour $\sigma^{-n}(C_j) \cap C_i$ est vide pour ces valeurs de n.

Supposons maintenant que A^n est à coefficients positifs pour $n \geqslant N$. On montre comme ci-dessus que, si C est un cylindre déterminé par une suite de longueur l, et si $n \geqslant N$, alors $\varphi^{-(l+n)}(C) = \Sigma_A$ pour $n > N$. Ceci implique le mélange. ►

Nous allons maintenant nous intéresser aux décalages du point de vue de la théorie ergodique. Comme les cylindres sont compacts, toute suite décroissante de cylindres a une intersection non-vide. Cette propriété implique que toute mesure de probabilité additive sur les cylindres est en fait sigma additive, et se prolonge donc en une mesure de probabilité sur X. Comme tout cylindre est une réunion finie disjointe de cylindres élémentaires, une mesure de probabilité sur X est en fait déterminée par sa valeur sur les cylindres élémentaires, c'est à dire par les valeurs

$$m(y_0, y_1, \dots, y_l) := m(\pi_l^{-1}(y_0, \dots, y_l))$$

qui s'interprète comme la probabilité qu'une suite commence par $y_0, \dots, y_l$. Réciproquement, la famille de réels positifs $m(y)$, $y \in \cup_{l \geqslant 0} Y^l$ détermine une mesure de probabilité si et seulement si elle est compatible, c'est à dire si et seulement si

$$m(y_0, \dots, y_l) = \sum_{i=1}^{k} m(y_0, \dots, y_l, i)$$

pour tout $l \geqslant 0$ et tout $y = (y_0, \dots, y_l) \in Y^{l+1}$.

Si $p = (p_1, \dots p_k)$ est une probabilité sur Y, alors la mesure produit $m = p^{\mathbb{N}}$ est invariante pour le décalage. Rappelons que $m = p^{\mathbb{N}}$ correspond à

$$m(y_0, \dots, y_l) = p(y_0)p(y_1) \cdots p(y_l).$$

Elle est ergodique, comme nous l'avons déjà montré.

Nous allons étudier une famille plus générale de mesures invariantes, les mesures de Markov, qui correspondent aux chaînes de Markov. On dit que Q est une matrice stochastique si c'est une matrice carrée $k \times k$ à coefficients positifs telle que la somme des coefficients de chaque ligne vaut 1, c'est à dire que $Q\mathbf{1} = \mathbf{1}$, où $\mathbf{1}$ est le vecteur de $\mathbb{R}^k$ dont les coefficients sont égaux à 1. La donnée d'une matrice stochastique Q et d'une mesure de probabilité $p = (p_1, \dots, p_k)$ sur $\{1, \dots, k\}$ détermine une mesure sur X, déterminée par

$$m(y_0, \dots, y_l) = p(y_0)Q(y_1, y_2) \cdots Q(y_{l-1}, y_l).$$

La condition de compatibilité découle directement du fait que Q est une matrice stochastique. On rappelle l'interprétation probabiliste de la mesure m, qui est la loi d'une suite de variables aléatoires à valeurs dans Y, telle que la loi de y_0 est donnée par p et, à chaque étape la loi de y_{n+1} si y_n vaut i est $j \mapsto Q(i, j)$.

Propriété 11.3. *Le rayon spectral d'une matrice stochastique est égal à* 1.

◄ Si Q est une matrice stochastique, alors Q^n est une matrice stochastique pour tout n. Ceci implique en particulier que la suite Q^n est bornée, et donc que le rayon spectral de Q est inférieur ou égal à 1. C'est donc 1, puisque 1 est valeur propre de Q. ►

On remarque que l'application $p \longmapsto pQ$ (p est une matrice ligne) préserve l'espace $\mathcal{P}(Y)$ des probabilités sur Y. De plus,

Propriété 11.4. *Si m_p est la mesure de Markov associée à la matrice Q et à la probabilité p, alors $\sigma_* m_p = m_{pQ}$.*
En particulier, la mesure m_p est invariante par σ si et seulement si $pQ = p$.

On dit que p est un probabilité stationnaire si $pQ = p$.

◄ Soit $y \in Y^{l+1}$ un mot fini. La préimage par σ du cylindre élémentaire associé à y est le cylindre associé à

$$Y \times \{y\} = \bigcup_{i \in Y} (i, y_0, \dots, y_l) \subset Y^{l+2}.$$

On a donc

$$\begin{aligned}(\sigma_* m_p)(y_0, \dots, y_l) &= \sum_{i \in Y} p(i)Q(i, y_0) \cdots Q(y_{l-1}, y_l) = (pQ)(y_0)Q(i, y_0) \cdots Q(y_{l-1}, y_l) \\ &= m_{pQ}(y_0, \dots, y_l).\end{aligned}$$

Les mesures $\sigma_* m_p$ et m_{pQ} sont égales sur les cylindres élémentaires, donc sur tous les ensembles. ►

Théorème 11.5. *Pour toute matrice stochastique Q, il existe $p \in \mathcal{P}(Y)$ tel que $pQ = p$, et donc tel que la mesure de Markov m_p est invariante par σ.*

◄ La méthode est classique. On considère n'importe quel $q \in \mathcal{P}(Y)$, et la suite $q_n := (q + qQ + \cdots + qQ^{n-1})/n$. Par compacité de $\mathcal{P}(Y)$, cette suite a des valeurs d'adhérence, et ces valeurs d'adhérence p vérifient $pQ = p$. ►

Étudions maintenant les propriétés ergodiques des mesures de Markov. On commence par un lemme naturel :

Lemme 11.6. *Soit Q une matrice stochastique. Si Q admet une unique probabilité invariante p, alors la mesure de Markov m_p correspondante est ergodique pour l'application σ. De plus, la matrice S^nQ/n converge vers P, la matrice stochastiques dont toutes les lignes sont égales à p.*

◄ Pour tout $q \in \mathcal{P}(Y)$, on a $(q + qQ + \cdots + qQ^{n-1})/n \longrightarrow p$. En effet, chaque valeur d'adhérence de cette suite est une probabilité invariante, donc est égale à p. Ceci implique que la matrice $S^nQ := (I + Q + Q^2 + \cdots + Q^{n-1})/n$ converge vers la matrice P dont toutes les lignes sont égales à p.

Soient C et C' les deux cylindres élémentaires déterminés par les mots finis $y = y_0 \cdots y_l$ et $z = z_0 \cdots z_m$. Pour n assez grand, on a

$$m_p(\sigma^{-n}(C) \cap C') = p(z_0)Q(z_0, z_1) \cdots Q(z_{l-1}, z_l)Q^{n-m-1}(z_m, y_0)Q(y_0, y_1) \cdots Q(y_{l-1}, y_l).$$

Si deux suites a_n et b_n sont égales à partir d'un certain rang, alors la convergence vers a de la suite a_n en moyenne de cesaro est équivalente à la convergence de b_n vers a en moyenne de Cesaro. Ici, on a

$$m_p(C')(S^nQ/n)(z_m, y_0)Q(y_0, y_1) \cdots Q(y_{l-1}, y_l) \longrightarrow m_p(C')m_p(C)$$

donc

$$\frac{1}{n}\sum_{i=0}^{n-1} m_p(\sigma^{-n}(C) \cap C') \longrightarrow m_p(C')m_p(C).$$

Cette convergence a alors lieu pour tous ensembles mesurables C et C'. Lorsque $C = C'$ est invariant, on obtient que $m_p(C) \in \{0, 1\}$. ►

La dernière affirmation de la preuve ci-dessus n'est pas évidente, nous la détaillons maintenant.

Propriété 11.7. *Soit $\mathcal{C}$ une algèbre d'ensemble qui engendre la tribu. Soit m une mesure de probabilité invariante d'une application mesurable φ.*

Si $m(\varphi^{-n}(C) \cap C') \longrightarrow m(C)m(C')$ pour tous C et C' dans $\mathcal{C}$, alors la mesure m est mélangeante.

Si $m(\varphi^{-n}(C) \cap C')$ tend vers $m(C)m(C')$ en moyenne de cesaro pour tous C et C' dans $\mathcal{C}$, alors la mesure m est ergodique.

◄ Posons $m_n(C, C') = m(\varphi^{-n}(C) \cap C')$ et $M_n(C, C') = \frac{1}{n}\sum_{i=1}^{n} m(\varphi^{-i}(C) \cap C')$.

Les preuves sont identiques, nous allons par exemple traiter le cas de la suite M_n. On considère dans un premier temps l'ensemble des parties C de X telles que $M_n(C, C') \longrightarrow m(C)m(C')$ pour tout $C' \in \mathcal{C}$. On va montrer que cette ensemble est une classe monotone, et donc que c'est l'ensemble de tous les mesurables. Il est clairement stable par complémentation. Soit C_k une suite croissante d'ensembles vérifiant la propriété ci-dessus et $C = \sup C_k$. On vérifie facilement que $\liminf M_n(C, C') \geqslant m(C_k)m(C')$ pour tout $C' \in \mathcal{C}$. En appliquant cette inégalité à $X - C'$ on déduit que $\lim M_n(C, C') = m(C)m(C')$, et donc que C vérifie la propriété voulue.

On a montré que $M_n(C, C') \longrightarrow m(C)m(C')$ pour tout ensemble mesurable C et tout $C' \in \mathcal{C}$. On considère maintenant l'ensemble des parties C' de X telles que cette convergence a lieu pour tout borélien C. On montre comme ci-dessus que c'est une classe monotone, et donc que c'est l'ensemble des boréliens. ►

Théorème 11.8. *Soit Q une matrice stochastique. Les propriétés suivantes sont équivalentes :*

1. *La matrice Q est transitive.*
2. *La matrice Q admet une unique probabilité invariante p, dont les coefficients sont tous strictement positifs.*
3. *$S^nQ/n \longrightarrow P$, une matrice stochastique à coefficients strictement positifs dont toutes les lignes sont égales.*
4. *Il existe une probabilité invariante p à coefficients strictement positifs et telle que la mesure de Markov m_p est ergodique.*

◄ 1 ⇒ 2 Supposons la matrice Q transitive. On va montrer que tout vecteur $f \in \mathbb{R}^Y$ tel que $Qf = f$ est un multiple de $\mathbf{1}$. Soit f un tel vecteur, et soit $r \in Y$ un des états en lesquels $f(r)$ est maximal. Comme $\sum_i Q(r,i)f(r) = f(r) = \sum_i Q(r,i)f(i)$ on obtient que $\sum_i Q(r,i)(f(r) - f(i)) = 0$. Comme c'est une somme dont les termes sont positifs, on conclue que chaque terme est nul, et que $f(i) = f(r)$ si $Q(r,i) > 0$. On montre ensuite de la même façon que $f(j) = f(r)$ si il existe i tel que $Q(r,i) > 0$ et $Q(i,j) > 0$, c'est à dire si $Q^2(r,j) > 0$, puis par récurrence que $f(j) = f(i)$ si il existe n tel que $Q^n(r,i) = 0$, c'est à dire pour tout $i \in Y$ au vu de la transitivité. L'espace propre de valeur propre 1 pour Q est donc engendré par $\mathbf{1}$. L'espace propre de la transposée est donc de dimension 1.

Montrons maintenant que $p(i) > 0$ pour tout i. Si il existe i tel que $p(i) = 0$, alors on a $0 = \sum_j p(j)Q(j,i)$, ce dont on déduit que $p(j) = 0$ si $Q(j,i) > 0$. Par le même type de récurrence que ci-dessus, on déduit que $p(j) = 0$ pour tout j en utilisant la transitivité, ce qui est contradictoire puisque p est une probabilité.

2 ⇒ 3 est donné par le Lemme.

3 ⇒ 1 Si $S^nQ/n \longrightarrow P$ à coefficients strictement positifs, alors S^nQ est à coefficients strictement positifs pour n grand.

2 ⇒ 4 Est donné par le Lemme.

4 ⇒ 3 Supposons que m_p est ergodique. Dans ce cas on a la convergence en moyenne de Césaro de $m_p(\sigma^{-1}(C) \cap C')$ vers $m_p(C)m_p(C')$ pour tous C et C' mesurables. En appliquant cette convergence lorsque $C' = \pi_0^{-1}(i)$ et $C = \pi_0^{-1}(j)$, on obtient que $p(i = Q^{n-1}(i,j) \longrightarrow p(i)p(j)$ pour tous i, j. Comme $p(i) > 0$ ceci prouve 2. ►

Donnons nous maintenant une matrice A transitive pas nécessairement stochastique.

Théorème 11.9. *Si σ_A est transitif, il existe une mesure de Markov m ergodique dont le support est Σ_A.*

◄ Si la matrice A est stochastique, c'est une conséquence immédiate de l'énoncé précédent. On considère l'unique probabilité invariante p et la mesure de Markov m_p correspondante. Cette mesure est ergodique, et son support est Σ_A. En effet, un cylindre élémentaire Q est de mesure nulle si et seulement si il est engendré par un mot non admissible.

Dans le cas où la matrice A n'est pas stochastique, on utilise une construction due à Parry. Le théorème de Perron Frobenius implique qu'il existe $f \in \mathbb{R}^Y$, à coefficients positifs, et $a > 0$ tels que $Af = af$. Le même argument que ci-dessus permet de montrer, puisque A est transitive, que les coefficients de f sont tous non nuls.

On pose alors $Q(i,j) = A(i,j)f(j)/af(i)$. C'est une matrice stochastique telle que $Q(i,j) > 0$ si et seulement si $A(i,j) > 0$. On peut donc appliquer la construction précédente et trouver une probabilité invariante p de Q, pour laquelle la mesure de Markov m_p est ergodique et a pour support $\Sigma_Q = \Sigma_A$.

On remarque que p est aussi le vecteur propre à gauche de A pour la matrice A, c'est à dire que $pA = ap$. ►

Théorème 11.10. *Soit A une matrice carrée à coefficients positifs donc chaque colonne contient un coefficient non nul. Il existe alors $a > 0$ et un vecteur propre f à coefficients positifs tels que $Af = af$.*

Si la matrice A est transitive, alors f est à coefficients strictement positifs.

On dit que la matrice Q est apériodique si il existe n tel que Q^n est à coefficients strictement positifs. L'exemple typique d'une matrice stochastique transitive mais pas apériodique est la matrice d'une permutation circulaire des états.

Théorème 11.11. *Les propriétés suivantes sont équivalentes pour une matrice stochastique Q.*

La matrice Q est apériodique.

$Q^n \longrightarrow P$, une matrice stochastique dont toutes les lignes sont égales.

La valeur propre 1 est de multiplicité algébrique égale à 1, et toutes les autres valeurs propres complexes sont de module strictement inférieur à 1.

Il existe une mesure invariante p à coefficients strictement positifs telle que la mesure de Markov m_p est mélangeante.

◄ Supposons que $Q^n \longrightarrow P$. Le calcul fait ci-dessus montre alors que $m_p(\sigma^{-n}(C) \cap C') \longrightarrow m_p(C)m_p(C')$ pour tous cylindres élémentaires C et C', et donc pour tous ensembles mesurables C et C'. Le système est donc mélangeant.

►

On peut penser à A comme à une relation sur l'ensemble $\{1, \dots, k\}$, et donc à une dynamique multivaluée. (avec iRj si et seulement si $A(i,j) > 0$). L'état $i \in Y$ est dit errant si $A^n(i,i) = 0$ pour tout $n \geqslant 1$. Sinon, il est dit récurrent (il n'y a pas lieu de distinguer point non-errant et point récurrent dans ce cas simple ou l'espace des états est discret, ces points sont aussi les points récurrents par chaines).

On peut noter $\triangleright$ la relation $i \triangleright j$ si et seulement si il existe $n \geqslant 1$ tel que $A^n(i,j) > 0$. C'est en fait la relation de connection par chaines. La relation $\triangleleft\triangleright$ est une relation d'équivalence sur l'ensemble des points récurrents. On dit que A est transitive si elle l'est en tant que dynamique multivaluée, c'est à dire si pour tous i et j dans $1, \dots, k$, il existe N tel que $A^N(i,j) > 0$. Si A n'est pas transitive, alors

Propriété 11.12. *Si la matrice A n'est pas transitive, alors il existe une partition de l'alphabet Y en deux parties non vides I et J qui ont la propriété que $A^n(i,j) = 0$ pour tout n, $i \in I$, et $j \in J$.*

◄ On peut en donner une démonstration élémentaire, mais nous allons en donner une qui utilise les considérations précédentes sur les fonctions de Lyapounov des relations dynamiques. En effet ici la relation associée à A est fermée, et engendre la relation transitive fermée $iLj \Leftrightarrow \exists n : A^n(i,j) > 0$. Si la matrice A n'est pas transitive, alors la relation L n'est pas totale, donc il existe une fonction de Lyapounov non triviale, c'est à dire ici un vecteur $f \in \mathbb{R}^k$ qui a la propriété que $f(j) \leqslant f(i)$ si $A(i,j) > 0$. En considérant une valeur appropriée de la fonction f, on partitionne $\{1, \dots, k\}$ en deux parties I et J telles que $f(i) < f(j)$ si $u \in I, j \in J$. Ceci implique que $A(i,j) = 0$. ►

Dans la situation de la propriété, l'ensemble $J^{\mathbb{N}}$ vérifie $\sigma^{-1}(\Sigma_A \cap J^{\mathbb{N}}) = \Sigma_A \cap J^{\mathbb{N}}$.

Proposition 11.13. *L'état i est récurrent pour A si et seulement si tout mot commençant par i est non-errant pour σ_A.*

◄ Soit y un mot dont le premier état y_0 est récurrent. Pour montrer que y est un point non errant, il faut montrer que pour tout l, il existe un mot z dont les $1 + l$ premiers termes sont ceux de y, et dont une des décalées $\sigma^j(z)$ a la même propriété. Comme le mot $y_0, \dots, y_l$ est admissible, on a $y_0 \rhd y_l$, donc $y_l \rhd y_0$, il existe alors un mot admissible $y_l, x_1, \dots, x_r, y_0$, si bien que le mot fini $z = y_0, \dots, y_l, x_1, \dots, x_r, y_0, \dots y_l$ a la propriété voulue.

Réciproquement, Supposons que tout mot y commençant par i est récurrent pour σ_A. Soit $x \in Y^{l+1}$ un mot fini commençant par i. Il existe alors $n \geqslant l + 2$ et un point z dans le cylindre $C = \pi_l^{-1}(x)$ tel que $\sigma^n(z) \in C$. Ceci montre que $y_l \rhd i$. ►

On dit que τ est une période de l'état i si $A^\tau(i,i) > 0$. L'ensemble $T(i)$ des périodes de i est un semi-groupe de $\mathbb{N}$, c'est à dire qu'il est stable par addition. L'état i est dit errant si $T(i) = \{0\}$.

Lemme 11.14. *Soit T un semi-groupe de $\mathbb{N}$ qui contient un élément non nul. Alors il existe un nombre entier $r > 0$, le PGCD des éléments de T, tel que tous les éléménts de T sont des multiples de r, et tous les multiples de r sont des éléments de T sauf un nombre fini d'entre eux.*

◄ La suite r_k des PGCD des k premiers éléments de P est une suite décroissantes à valeurs dans $\mathbb{N}$, donc elle se stabilise à une valeur r. Il existe alors une relation $r = n_1p_1 + \cdots + n_kp_k$, où $n_i \in \mathbb{Z}$ et p_i sont les éléments de P classés par ordre croissant. Posons maintenant $m = |n_1|p_1 + \cdots + |n_k|p_k$. Soit a un grand multiple de r. Par division Euclidienne, on écrit $a = bm + qr$, où $0 \leqslant q < m/r$, et donc

$$a = (b|n_1| + qn_1)p_1 + \cdots + (b|n_k| + qn_k)p_k.$$

Si $a > m^2/r$, donc $b > m/r$, alors les coefficients sont tous des entiers positifs donc $a \in P$. ►

Pour chaque i, on note $\tau(i)$ le PGCD des périodes de l'état i. On convient que $\tau(i) = 0$ si i est un état errant.

Propriété 11.15. *Si A est transitive, tous les états ont la même période.*

◄ Soient i et j deux états, et soient n et m tels que $A^n(i,j) > 0$ et $A^m(j,i) > 0$. Alors si $l\tau(i)$ est une période de i, $l\tau(i) + n + m$ est une période de j et donc est un multiple de $\tau(j)$. On a donc, $l\tau(i) + n + m = 0 \mod \tau(j)$ pour tout l assez grand, ce qui implique que $\tau(i) = 0$ modulo $\tau(j)$, c'est à dire que $\tau(j)$ divise $\tau(i)$. De manière symétrique, $\tau(i)$ divise $\tau(j)$, donc $\tau(i) = \tau(j)$. ►

Lemme 11.16. *Les périodes $\tau(i), i \in Y$ sont toutes égales à 1 si et seulement si il existe N tel que $A^N(i,j) > 0$ pour tous i et j. On dit alors que A est apériodique.*

◄ Si tous les coefficients de A^N sont strictement positifs, alors il en est de même de A^n pour tout $n \geqslant N$. Ceci implique en particulier que $p(i) = 1$ pour tout i.

Réciproquement, si $p(i) = 1$, alors il existe k_i tel que $A^n(i,i) > 0$ pour tout $n \geqslant k_i$. Il existe donc k tel que tous les coefficients diagonaux ed A^n sont strictement positifs lorsque $n \geqslant k$. Si de plus A est transitive, alors pour tout (i,j), il existe $m(i,j)$ tel que $A^{m(i,j)}(i,j) > 0$. Comme $A^{m(i,j)+n}(i,j) \geqslant A^n(i,i)A^{m(i,j)}(i,j)$, on déduit que ces coefficients sont strictement positifs pour tout $n \geqslant k$, donc que A^N est à coefficients strictement positifs pour N assez grand. ►

Le théorème suivant montre en particulier que toute matrice stochastique admet une loi invariante.

Théorème 11.17. *Soit B une matrice à coefficients positifs sans états errants.*

Alors il existe une valeur propre réelle strictement positive b et un vecteur propre p à coefficients positifs tel que $Bp = bp$. Toutes les autres valeurs propres complexes de B ont un module inférieur à b.

Si B est transitive, alors p est à coefficients strictement positifs, et c'est le seul vecteur propre de B à coefficients positifs. De plus la multiplicité algébrique de b comme valeur propre de B est égale à 1

Si B est apériodique, alors toutes les autres valeurs propres ont un module strictement inférieur à b.

Si B est une matrice stochastique, alors $b = 1$ et il existe une probabilité p sur Y telle que $pQ = p$.

12 Applications dilatantes

On considère une application $\varphi : [0,1] \longrightarrow [0,1]$, on munit l'intervalle de sa tribu borélienne $\mathcal{B}$ et de la mesure de Lebesgue λ. On suppose qu'il existe une famille finie ou dénombrable (on n'hésitera pas à supposer en première lecture que cette famille est finie) d'intervalles ouverts I_i tels que la restriction de φ à I_i est un difféomorphisme de I_i dans $]0,1[$. On note $\mathcal{S}_0$ l'ensemble des indices possibles i, il est fini ou dénombrable. On note $Y_0 := \cup_i I_i$, $Z_0 = [0,1] - Y_0$, et on suppose que Z_0 est dénombrable. On suppose de plus qu'il existe $a > 1$ et une constante $C > 0$ telles que

$$|\varphi'(x)| \geqslant a > 1 \quad , \quad |\varphi''(x)| \leqslant C$$

pour tout $x \in Y_0$. On ne suppose pas l'application φ continue.

Trois exemples type sont :

L'application $x \mapsto 2x$ sur $]0,1/2[$ et $x \mapsto 2x-1$ sur $]1/2,1[$ (qui, pour la théorie ergodique, est l'application $x \mapsto 2x$ sur $\mathbb{T}$).

L'application tente $x \mapsto 1 - |2x-1|$.

L'application de Gauss $x \mapsto \{1/x\}$ (partie fractionnaire de $1/x$).

Sous les hypothèses précédentes, on a :

Théorème 12.1. *Il existe une fonction Lipschitz $g : [0,1] \longrightarrow]0,\infty)$ telle que la mesure $m = g\lambda$ est invariante. La mesure m est ergodique, et même mélangeante.*

On a alors $S_*^n \delta_x / n \longrightarrow m$ pour λ-presque tout x (puisque les ensembles de mesure nulle pour λ sont les mêmes que les ensembles de mesure nulle pour m).

L'ergodicité de m implique qu'aucune autre mesure invariante n'est absolument continue par rapport à λ.

Exercice 12.1. *Dans le cas de l'application de Gauss $\varphi(x) = \{1/x\}$, montrer que $g = ((\log 2)(1+x))^{-1}$.*

Commençons par quelques définitions. On note Y_1 la réunion des intervalles I_i et Z_1 son complémentaire, qui est dénombrable. On note $Z_n := \cup_{k=0}^n \varphi^{-n}(Z_0)$, et $Z = \cup Z_n$. Ce sont des ensembles dénombrables. On note Y_n le complémentaire de Z_n, et Y le complémentaire de Z. Les points de Y sont ceux dont l'orbite reste dans Y_0. Chaque point de Y a un itinéraire, qui est la suite des indices i_n tels que $\varphi^n(x) \in I_{i_n}$. Les composantes connexes de Y_n sont en bijections avec les itinéraires finis $i_0, i_1, \dots, i_n$. On note $S_n = S_0^{n+1}$ l'ensemble des itinéraires de taille n. En fait, on vérifie par récurrence la propriété suivante :

L'ensemble $I(i_0, i_1, \dots, i_n)$ constitué des points ayant l'itinéraire $i_0, \dots i_n$ est un intervalle ouvert disjoint de Z_n, dont les bords sont dans Z_n, et en restriction duquel φ^n est un difféomorphisme sur $]0,1[$.

On note $\mathcal{A}_0$ la tribu constituée des parties de $[0,1]$ qui sont réunion d'intervalles I_i et d'une partie de Z_0. Autrement dit, ce sont les réunions dénombrables d'intervalles ouverts ou fermés à chaque extrémités et dont les extrémités sont dans Z_0. On note de même $\mathcal{A}_n$ la tribu dont les éléments sont lés réunions d'intervalles à extrémités dans Z_n. C'est une famille croissante de tribus. Pour tout n, l'application φ est mesurable de $([0,1], \mathcal{A}_{n+1})$ dans $([0,1], \mathcal{A}_n)$. En fait, la tribu $\mathcal{A}_{n+1}$ est la tribu engendrée par $\varphi^{-1}(\mathcal{A}_n)$ et $\mathcal{A}_n$. On note finalement $A = \cup_n \mathcal{A}_n$, c'est une algèbre d'ensembles.

Propriété 12.2. *Pour tout itinéraire, on a $\lambda(I(i_0, i_1, \dots, i_n)) \leqslant a^{-n}$. L'ensemble Z est donc dense dans* $[0,1]$*, et l'algèbre $\mathcal{A}$ engendre la tribu borélienne.*

◄ On a vu que $I(i_0, \dots i_n)$ est un intervalle en restriction duquel φ est un difféomorphisme sur $]0,1[$. On a donc soit $\varphi^n \geqslant a^n$ soit $\varphi^n \leqslant -a$ sur cet intervalle. Dans les deux cas,

$$1 = \lambda(\varphi^n(I(i_0, \dots, i_n))) \geqslant a^n \lambda(I(i_0, \dots i_n))$$

ce qui donne la première inégalité. La densité de Z découle du fait que Z_n est a^{-n}-dense. Comme Z est dense, tout intervalle fermé de $[0,1]$ est une intersection dénombrable d'intervalles à extrémités dans Z, et donc d'éléments de $\mathcal{A}$. ►

Montrons d'abord le théorème dans le cas particulier où les branches sont affines, c'est à dire ou φ' est constante sur I_i, et donc égale à $\varphi_i'(x) = 1/\lambda(I_i)$ pour $x \in I_i$. Montrons pour commencer que la mesure de Lebesgue λ est invariante. La préimage d'un borélien B est la réunion des $B_i := \varphi_i^{-1}(B)$. On a $\lambda(B_i) = \lambda(B)\lambda(I_i)$, donc

$$\lambda(\varphi^{-1}(B)) = \sum_i \lambda(B_i) = \lambda(B).$$

Ceci montre l'invariance de λ. Montrons maintenant le mélange. On a vu que

$$\lambda(\varphi^{-1}(B) \cap I_i) = \lambda(B)\lambda(I_i).$$

On démontre alors par récurrence que
$$\lambda(\varphi^{-n}(B) \cap A) = \lambda(B)\lambda(A)$$
pour tout $A \in \mathcal{A}_n$. Ceci montre que la convergence $\lambda(\varphi^{-n}(B) \cap A) \longrightarrow \lambda(B)\lambda(A)$ a lieu pour tout $A \in \mathcal{A}$, et donc pour tout $A \in \mathcal{B}$. La mesure λ est donc mélangeante.

Si f est une fonction strictement positive et Lipschitz sur $[0, 1]$, alors $\log \circ f$ est une fonction Lipschitz.

Lemme 12.3. *Soit $\mu = f\lambda$ une mesure de probabilité. Supposons que f est une fonction Lipschitz et strictement positive. La mesure $\varphi_*\mu$ est admet une densité h strictement positive qui vérifie*
$$\mathrm{Lip}(\log \circ h) \leqslant Ca^{-2} + a^{-1}\mathrm{Lip}(\log \circ f)$$
pour une constant C indépendante de f. Cette inégalité est aussi satisfaite par les densités h_i des mesures $(\varphi_i)_*(\mu_{|I_i})$.

◄ Pour chaque $i \in \mathcal{S}_0$, la restriction φ_i de φ à I_i est un difféomorphisme sur $]0, 1[$. La densité h_i de la mesure $(\varphi_i)_*(\mu_{|I_i})$ est donc $h_i = (f \circ \varphi_i^{-1})/|\varphi' \circ \varphi^{-1}|$. On a alors
$$\mathrm{Lip}(\log \circ h_i) \leqslant \mathrm{Lip}(\log \circ f)\mathrm{Lip}(\varphi_i^{-1}) + \mathrm{Lip}(\log \circ |\varphi_i'| \circ \varphi_i^{-1}) \leqslant a^{-1}\mathrm{Lip}(\log \circ f) + Ca^{-2}$$
où on rappelle que C est la borne de φ''. On a $\varphi_*\mu = \sum_i (\varphi_i)_*(\mu_{|I_i})$ la densité de cette mesure est $h := \sum_i h_i$. Pour montrer que cette somme de fonctions positives converge il suffit de la majorer. Pour ceci, on constate que la fonction h_i, qui est continue, atteint forcément sa valeur moyenne, qui est $\mu(I_i)$. La fonction $\log \circ h_i$ prend donc la valeur $\log(\mu(I_i))$, et donc elle est majorée par $\log(\mu(I_i)) + \mathrm{Lip}(\log \circ h_i)$. On a donc
$$h_i \leqslant \mu(I)e^{\mathrm{Lip}(\log \circ h_i)} \leqslant \mu(I)e^{Ca^{-2}+a^{-1}\mathrm{Lip}(\log \circ f)}$$
et donc $\sum h_i \leqslant \exp(Ca^{-2} + a^{-1}\mathrm{Lip}(\log \circ f))$. Chacune des fonction h_i satisfait
$$h_i(x) \leqslant h_i(y)e^{|y-x|(Ca^{-2}+a^{-1}\mathrm{Lip}(\log \circ f))}$$
pour tous x et y. Il en est donc de même de la somme h, c'est à dire qu'on a $\mathrm{Lip}(\log \circ h) \leqslant a^{-1}\mathrm{Lip}(\log \circ f) + Ca^{-2}$. ►

On peut alors démontrer l'existence de la mesure m du théorème 12.1. On se donne une constante L assez grand pour que $Ca^{-2}+a^{-1}L \leqslant L$. Alors pour toute fonction f vérifiant $\mathrm{Lip}(\log \circ f) \leqslant L$ et $\int f d\lambda = 1$ (par exemple la fonction $f = 1$), on pose $\mu = f\lambda$ et on considère comme d'habitude la suite $S_*^n\mu/n$. Il découle du Lemme précédent que cette mesure admet une densité g_n qui vérifie $\mathrm{Lip}(\log \circ g_n) \leqslant L$. Comme les fonctions g_n prennent la valeur 1, on déduit facilement qu'elles sont equi-Lipschitz. On trouve donc par Ascoli une sous-suite qui converge uniformément vers une limite Lipschitz g. La mesure $m = g\lambda$ est alors une valeur d'adhérence de la suite $S_*^n\mu/n$, la densité g satisfait $\mathrm{Lip}(\log \circ g) \leqslant L$, donc g se prolonge en une fonction Lipschitz strictement positive de $[0, 1]$.

L'existence d'une mesure de probabilité invariante $m = h\lambda$ avec $\mathrm{Lip}(\log h) \leqslant L$ est démontrée.

Propriété 12.4. *Pour tout itinéraire $j = (i_0, \dots, i_n) \in \mathcal{S}_n$, la mesure $m_j = (\varphi_j^n)_*(m_{|I})$, où $I = I(i_0, \dots, i_n)$. Chacune des mesures m_j admet une densité h_j qui vérifie $\mathrm{Lip}(\log \circ h_j) \leqslant L$ et donc*
$$m(I_j)e^{-L} \leqslant h_j \leqslant m(I_j)e^L.$$

◄ On le montre par récurrence. En notant $\tilde{j} = (i_0, \dots, i_{n-1})$, on a $m_j = (\varphi_{i_n})_* m_{\tilde{j}|I_{i_n}}$. Par hypothèse de récurrence, la mesure $m_{\tilde{j}}$ admet une densité $h_{\tilde{j}}$ qui vérifie $\mathrm{Lip}(\log \circ h_{\tilde{j}}) \leqslant L$. Le lemme ci-dessus implique qu'il en est de même de la mesure m_j. ►

Pour étudier l'ergodicité de la mesure m, nous avons besoin d'un résultat général de théorie de la mesure.

Théorème 12.5. *Soit $\mathcal{A}_n$ une suite croissantes de tribus dont la réunion engendre $\mathcal{B}$. Pour toute fonction $f \in L^2([0, 1], m)$, la suite $f_n := E(f|\mathcal{A}_n)$ converge vers f dans L^2.*

Il s'agit de la version la plus facile du théorème de convergence des martingales. La convergence a aussi lieu presque partout, mais la démonstration en est un peu plus difficile (c'est le théorème de convergence ponctuel des martingales).

◄ Notons $H = L^2([0, 1], \mathcal{B}, m)$ et $H_n = L^2([0, 1], \mathcal{A}_n, m)$. La fonction $E(f|\mathcal{A}_n)$ est la projetée orthogonale de f sur H_n. Il suffit donc de démontrer que le sous-espace $\cup H_n$ est dense dans H.

Ceci découle encore d'un argument de classe monotone. Soit F la fermeture de $\cup H_n$. Considérons les parties de $[0, 1]$ dont l'indicatrice est un élément de F. Les éléments de $\cup \mathcal{A}_n$ dont évidemment de tels ensembles. De plus, il est facile de vérifier que ces ensembles constituent une classe monotone. Tout borélien a donc sa fonction indicatrice dans F. Comme F est un sous-espace vectoriel, toute fonction à image finie et dans F. Ces fonctions sont denses dans L^2, donc $F = L^2$. ►

Dans le présent contexte, on déduit :

Corollaire 12.6. *Soit B un borélien tel que $m(B) < 1$ et $\epsilon > 0$. Alors il existe $n \geqslant 0$ et un itinéraire $i_0, \dots, i_n$ tel que $m(B \cap I(i_0, \dots, i_n)) \leqslant \epsilon m(I(i_0, \dots, i_n))$.*

◄ La fonction $E(1_B|\mathcal{A}_n)$ tend vers 1_B dans L^2, donc $\int_{B^c} E(1_B|\mathcal{A}_n) \longrightarrow 0$. Il existe donc une sous-suite telle que

$$E(1_B|\mathcal{A}_{n_k}) \longrightarrow 0$$

presque partout hors de B. On remarque alors que $E(1_B|\mathcal{A}_n)$ est la fonction qui est constante sur chaque composante connexe I de Y_n et y prend la valeur $m(B \cap I)/m(I)$. Pour chaque $x \in Y_n$, notons $J_n(x)$ la composante connexe de Y_n contenant x. L'ensemble $\tilde{Y} \subset Y$ des point x en lesquels $E(1_B|\mathcal{A}_n) = m(B \cap J_n(x))/m(J_n(x))$ pour tout n est de mesure totale. Soit x un point de $\tilde{Y} - B$ sur lequel la convergence $E(1_B|\mathcal{A}_{n_k}) \longrightarrow 0$ a lieu. Pour tout $\epsilon > 0$, il existe donc n_k tel que $m(B \cap J_{n_k}(x))/m(J_{n_k}(x)) = E(1_B|\mathcal{A}_{n_k})(x) \leqslant \epsilon$. ►

Montrons l'ergodicité de la mesure m du théorème 12.1. Soit B un borélien invariant, tel que $m(B) < 1$ (ou ce qui est équivalent, $\lambda(B) < 1$). On considère comme ci-dessus un itinéraire dont l'intervalle $I \in \mathcal{A}_n$ correspondant vérifie $m(B \cap I) \leqslant \epsilon m(I)$. On a alors

$$\epsilon m(I) \geqslant m(B \cap I) = m_I(B) = \int_B h_I d\lambda \geqslant \lambda(B) m(I) e^{-L}$$

et donc $\lambda(B) \leqslant \epsilon e^L$. Comme ceci est vrai pour tout $\epsilon > 0$ que on déduit que $\lambda(B) = 0$, donc $m(B) = 0$. On a montré l'ergodicité de m.

Une petite modification de l'argument ci-dessus permet de montrer que tous les ensembles asymptotiquement mesurables sont de mesure 0 ou 1. On rappelle que B est dit asymptotiquement mesurable si, pour tout $n \geqslant 0$, $\varphi^n(B)$ est mesurable et $B = \varphi^{-N}(\varphi^n(B))$ et que les ensembles asymptotiquement mesurables forment une tribu $\cap_n \varphi^{-n}(\mathcal{B})$.

Propriété 12.7. *Soit B un Borélien qui est asymptotiquement mesurable, alors $m(B) \in \{0, 1\}$.*

◄ On commence la preuve comme ci-dessus en choisissant une composante connexe maximale I de Y_n telle que $m(B \cap I) \leqslant \epsilon m(I)$. On a alors $B \cap I = \varphi_I^{-n}(\varphi^n(B))$, donc $m(B \cap I) = m_I(\varphi^n(B))$. On déduit comme plus haut :

$$\epsilon m(I) \geqslant m(B \cap I) = m_I(\varphi^n(B)) = \int_{\varphi^n(B)} h_I d\lambda \geqslant \lambda(B) m(I) e^{-L}$$

et donc $m(B) = m(\varphi^n(B)) \leqslant \lambda(\varphi^n(B)) e^L \leqslant \epsilon e^{2L}$ pour tout $\epsilon > 0$, donc $m(B) = 0$. ►

Proposition 12.8. *Si m est une mesure de probabilité invariante telle que $m(B) \in \{0, 1\}$ pour tout B asymptotiquement mesurable, alors m est mélangeante.*

◄ On considère l'espace de Hilbert $H = L^2(X, m)$, et les sous-espaces fermés $H_n = L^2(X, \varphi^{-n}(\tau), m)$. Ces sous-espaces forment une suite décroissante, et l'hypothèse est que cette intersection est l'ensemble des fonctions constantes, identifié à $\mathbb{R}$.

Posons $F_n = H_n \cap H_{n+1}^{\perp}$, c'est donc l'orthogonal de H_{n+1} dans H_n. Posons $G_n = \mathbb{R} \oplus F_1 \oplus F_2 \oplus \cdots \oplus F_n$ et $G = \cup_n G_n$. Le sous-espace G est dense dans H.

L'application $f \longmapsto f \circ \varphi^n$ envoie H sur H_n, de sorte que $\int g(f \circ \varphi^n) dm = \int f \int g$ à partir d'un certain rang pour $f, g \in G_n$. Par l'argument classique de densité, le mélange en découle. ►

13 Mélange et théorème de Perron Frobenius

Pour étudier les mesures invariantes, on a été amené à considérer à plusieurs reprises la transformation $m \mapsto \varphi_* m$ sur l'ensemble des mesures de probabilité. On a, par exemple, vu que cette application admet un point fixe si X est un espace métrique compact et φ est continue. Notons que dans ce cas, l'espace $\mathcal{P}$ est un espace métrique compact et φ_* est une application continue. On peut considérer $(\mathcal{P}, \varphi_*)$ comme un système dynamique topologique.

La dynamique la plus simple que l'on pourrait imaginer pour $(\mathcal{P}, \varphi_*)$ est celle d'un unique point fixe qui attirerait tous les autres points. Mais ceci arrive rarement. En effet, pour tout $x \in X$, on a $\varphi_* \delta_x = \delta_{\varphi(x)}$, et une suite de diracs ne peut converger que vers un dirac. Donc la dynamique $(\mathcal{P}, \varphi_*)$ a un unique point fixe qui attire tous les autres points si et seulement si (X, φ) a la même propriété.

Dans ce chapitre, nous allons relier les propriétés ergodiques d'une mesure, et en particulier son caractère mélangeant, à ses propriétés d'attractivité en tant que point fixe de φ_*. Au niveau formel, ce lien est élémentaire : La probabilité invariante m est mélangeante si et seulement si $\int f \circ \varphi^n g dm \longrightarrow \int f dm \int g dm$ pour toutes fonctions continues f et g. Ceci est équivalent à dire que $\varphi_*^n \mu \longrightarrow m$ pour toute probabilité μ admettant une densité continue par rapport à m (ou plus généralement une densité par rapport à m). Autrement dit, la mesure m est mélangeante si et seulement si domaine d'attraction en tant que point fixe de φ_* contient les mesures ayant une densité par rapport à m.

La structure particulière du système $(\varphi_*, \mathcal{P}(X))$ (il s'agit d'une application linéaire sur un ensemble convexe) suggère l'utilisation des distances de Hilbert, qui sont des distances contractées par les applications linéaires. Pour définir cette distance sur le convexe $W \subset E$, On note $\mathcal{H}$ l'ensemble des fonctions sur E qui sont le quotient de deux fonctions affines, et $\mathcal{H}_W$ l'ensemble des fonctions de $\mathcal{H}$ qui sont strictement positives sur W. La distance de Hilbert sur W est définie par

$$\theta_W(x, y) = \sup_{f \in \mathcal{H}_W} \log \circ f(x) - \log \circ f(y) = \log \sup_f \frac{f(x)}{f(y)}.$$

On vérifie facilement que θ_W prend ses valeurs dans $[0, \infty]$ et qu'elle satisfait l'inégalité triangulaire. On notera θ au lieu de θ_W quand il n'y a pas d'ambiguïté. On commence par énoncer la propriété de contraction, qui est une conséquence immédiate de la définition :

Propriété 13.1. *Soit $F(x) : L(x)/l(x)$ une application de E dans E qui est le quotient d'une application linéaire L et d'une forme linéaire l. Supposons que $F(W) \subset W$ (donc en particulier que l ne s'annule pas sur W). Alors*

$$\theta_W(F(x), F(y)) \leqslant \theta_W(x, y)$$

pour tous x et y dans W.

On ajoute une seconde propriété immédiate mais importante :

Propriété 13.2. *Soit $U \subset W$ deux convexes. Alors $\theta_U(x, y) \geqslant \theta_W(x, y)$ pour tous x et y dans U.*

Proposition 13.3. *Si W a la propriété que toute doite affine D intersecte W suivant un intervalle ouvert strictement contenu dans D, alors θ_W est une distance sur W.*

L'exemple le plus simple est celui d'un convexe ouvert dans un espace vectoriel topologique localement convexe. L'ensemble des mesures de probabilité est un exemple délicat. Il est compact et donc ne vérifié pas l'hypothèse ci-dessus. Plus grave, il n'est pas évident de lui définir un intérieur lorsque X n'est pas dénombrable. En effet, si m est une mesure de probabilité, alors il existe une mesure de probabilité μ telle que m est dans le bord de $\mathcal{P} \cap D$, où D est la droite affine engendrée par m et μ. Il suffit pour ceci de prendre pour μ un dirac supporté par un point x tel que $m(\{x\}) = 0$. Nous verrons plus tard qu'il est quand même possible de faire des choses.

La propriété précédente découle de :

Propriété 13.4. *Soient $x, y \in K$, soit D la droite affine contenant x et y. Sur D, soient p et q les bords de l'intervalle $D \cap K$. On se donne une paramétrisation affine de D et on suppose que p et q sont choisis tels que l'ordre des points sur le droite D est p, x, y, q. On suppose que D n'est pas contenue dans K, c'est à dire que $p > -\infty$ ou $q < +\infty$. Alors on a*

$$\theta_W(x, y) = \log \frac{(y-p)(q-x)}{(q-y)(x-p)} \quad , \quad \theta_W(x, y) = \log \frac{q-x}{q-y} \text{ si } p = -\infty \quad , \quad \theta_W(x, y) = \log \frac{y-p}{x-p} \text{ si } q = \infty.$$

◄ Toute fonction affine sur D qui est positive sur $D \cap K$ se prolonge en une fonction affine sur E qui est positive sur K, par le théorème de Hahn Banach. Ceci implique que $\theta_K(x, y) = \theta_{K \cap D}(x, y)$.

On travaille maintenant en restriction à la doite D. Soit l est une fonction affine strictement positive sur $]p, q[$, alors :

Soit $l(x) = a(x - z)$, avec $a > 0$ et $z \leqslant p$, et alors $l(y) \leqslant (x) \leqslant (y - p)/(x - p)$.

Soit $l(x) = a$, alors $l(y)/l(x) = 1 \leqslant (y-p)/(x-p)$.

Soit $l(x) = a(w-x)$ avec $a > 0$ et $w \geqslant q$. Dans ce cas, $l(y)/l(x) \leqslant 1 \leqslant (y-p)/(x-p)$.

Dan tous les cas, on a $l(y)/l(x) \leqslant (y-p)/(x-p)$. De manière similaire, on a $g(x)/g(y) \leqslant (q-x)/(q-y)$. On déduit finalement, avec $f = g/l$, que

$$\frac{f(x)}{f(y)} = \frac{g(x)l(y)}{g(y)l(x)} \leqslant \frac{(y-p)(q-x)}{(q-y)(x-p)}. \blacktriangleright$$

Propriété 13.5. *Soit $\mathcal{F}$ une famille d'applications affines. Considérons le convexe $W = \cap_{f\in\mathcal{F}}\{f > 0\}$. Alors le supremum dans la definition de θ_W peut être pris sur les fonctions $f = l/g$ avec l, g dans $\mathcal{F}$.*

◄ Il suffit de le démontrer pour un intervalle $]p, q[\in \mathbb{R}$. La famille $\mathcal{F}$ contient alors des éléments $l_n = a_n(x - p_n)$ et $g_n = b_n(q_n - x)$, où $p_n < p$ tend vers p, $q_n > q$ tend vers q, et $a_n > 0$, $b_n > 0$. Pour $x < y$ dans $]q, p[$, on a alors $l_n(y)g_n(x)/l_n(x)g_n(y) \longrightarrow \theta(x, y)$ donc $\theta(x, y) \leqslant \sup_{l,g\in\mathcal{F}} l(y)g(x)/l(x)g(y)$. ►

Propriété 13.6. *Soit W un convexe ouvert (resp. borné pour la norme) d'un espace vectoriel normé. Alors la topologie associée à la distance θ sur W est plus forte (resp. plus faible) que la topologie ambiante.*

Soit W un convexe ouvert et borné d'un espace de Banach E, alors la distance θ_W est complète.

◄ Si W est ouvert, pour tout $x \in W$, il existe r tel que $B(x, r) \subset W$. Ceci implique que $\theta_W \leqslant \theta_{B(x,r)}$ sur $B(x, r)$. et donc que

$$\theta(x, y) \leqslant \log\frac{r + |y-x|}{r - |y-x|} \leqslant \frac{2|y-x|}{r - |y-x|}$$

pour tout $y \in B(x, r)$. La convergence vers x pour la norme implique donc la convergence pour θ.

Si W est borné, alors il existe $R > 0$ tel que $W \subset B(x, R)$ pour tout x. Ceci implique que, pour tous x et y,

$$|y-x| \leqslant R\tanh(\theta(x, y)/2).$$

Si W est borné, on conclut donc que toute suite de Cauchy x_n pour θ est de cauchy pour la norme. Elle converge donc vers un point $x \in \bar{W}$. Soit n tel que $\theta(x_n, x_m) \leqslant 1$ pour tout $m \geqslant n$. Si x n'est pas dans W, alors il existe une fonction affine continue l qui est nulle en x et strictement positive sur W. On a alors $\theta_W(x_n, x_m) \geqslant \log f(x_n) - \log f(x_m) \longrightarrow +\infty$ lorsque $m \longrightarrow \infty$, ce qui est une contradiction. ►

Théorème 13.7. *Soit $U \subset W$. deux convexes. Supposons que le diamètre Δ de U pour la distance θ_W est fini. Alors*

$$\theta_U \geqslant \theta_W/\tanh(\Delta/4).$$

En conséquence, si $F = L/l : W \longrightarrow W$ est telle que le diamètre de son image est $\delta < 1$. Alors elle est strictement contractante de rapport $\tanh(\Delta/4)$ *pour la distance θ_W.*

◄ On remarque d'abord que, pour $0 < \delta \leqslant 1/2$, $\theta_{[0,1]}(\delta, 1-\delta) = 2\log(1-\delta) - 2\log\delta$. C'est le diamètre Δ du segment $]\delta, 1-\delta[$. On a donc $\delta = 1/(e^{\Delta/2}+1)$. Tout intervalle de diamètre Δ dans $]0, 1[$ est équivalent, par une transformation homographique, à l'intervalle $]\delta, 1-\delta[$.

Nous devons maintenant montrer que

$$\theta_{]0,1[}(x, y) \leqslant (1-2\delta)\theta_{]\delta,1-\delta[}(x, y) = \tanh(\Delta/4)\theta_{]\delta,1-\delta[}(x, y)$$

pour tous $x \leqslant y$ dans $]\delta, 1-\delta[$. On écrit d'abord

$$\theta_{]0,1[}(x, y) = \log\frac{y}{1-y} - \log\frac{x}{1-x} = \int_x^y \frac{1}{t(1-t)}dt$$

et

$$\theta_{]\delta,1-\delta[}(x, y) = \log\frac{y-\delta}{1-y-\delta} - \log\frac{x-\delta}{1-x-\delta} = \int_x^y \frac{1-2\delta}{(t-\delta)(1-t-\delta)}dt.$$

On remarque que

$$\frac{(t-\delta)(1-t-\delta)}{t(1-t)} = 1 + \frac{\delta^2-\delta}{t-t^2} \leqslant 1 + 4(\delta^2-\delta) = (1-2\delta)^2.$$

Finalement, on obtient

$$\theta_{]0,1[}(x, y) = \int_x^y \frac{1}{t(1-t)}dt \leqslant \int_x^y \frac{(1-2\delta)^2}{(t-\delta)(1-t-\delta)}dt = (1-2\delta)\theta_{]\delta,1-\delta[}(x, y). \blacktriangleright$$

Commençons par étudier un exemple en dimension finie. On considère une matrice stochastique Q à coefficients strictement positifs et l'application $p \mapsto pQ$ sur l'espace $\mathcal{P}(Y) \subset \mathbb{R}^Y$, dont les éléments sont vus comme des matrices lignes. On a vu que cette application admet un point fixe. Les considérations ci-dessus permettent de montrer que ce point fixe est asymptotiquement stable et que son bassin est $\mathcal{P}(Y)$. On conclut que $\rho Q^n \longrightarrow p$ pour tout $\rho \in \mathcal{P}(Y)$, ce qui implique que $Q^n \longrightarrow P$, la matrice dont toutes les lignes sont égales à p. On obtient le résultat suivant sur le caractère mélangeant des mesures de Markov (en notant m_p la mesure de Markov déterminée par la probabilité stationnaire p et la matrice de transition Q)

Théorème 13.8. *Les propriétés suivantes sont équivalentes pour une matrice stochastique Q.*
La matrice Q est irréductible et apériodique.
$Q^n \longrightarrow P$, une matrice stochastique dont toutes les lignes sont égales et à coefficients strictement positifs.
Il existe une mesure invariante p à coefficients strictement positifs telle que la mesure de Markov m_p est mélangeante.

◄ Si Q est à coefficients strictement positifs, alors l'image de l'application $\rho \mapsto \rho Q$ est un compact contenu dans l'intérieur W de $\mathcal{P}(Y)$. Il est donc de diamètre fini, ce qui implique que l'application $\rho \mapsto \rho Q$ est une contraction sur W pour la norme de Hilbert. Ellle a donc un unique point fixe, qui attire tout $\mathcal{P}(Y)$. Ceci implique que $\rho Q^n \longrightarrow P$ pour tout $\rho \in \mathcal{P}(Y)$, et donc que $Q^n \longrightarrow P$. Si Q est seulement supposée irréductible et apériodique, alors il existe N tel que Q^N est à coefficients strictement positifs. Ceci implique que l'application $\rho \mapsto \rho Q^N$ est une contraction pour la norme de Hilbert sur W. Par des arguments déjà vus dans les premiers chapitres, ceci implique que l'application $\rho \mapsto \rho Q$ a un unique point fixe qui attire tout W, et donc qui attire tout $\mathcal{P}(Y)$.

Supposons que $Q^n \longrightarrow P$. Soient C et C' les deux cylindres élémentaires déterminés par les mots finis $y = y_0 \cdots y_l$ et $z = z_0 \cdots z_m$. Pour n assez grand, rappelle l'expression

$$m_p(\sigma^{-n}(C) \cap C') = p(z_0)Q(z_0, z_1) \cdots Q(z_{l-1}, z_l)Q^{n-m-1}(z_m, y_0)Q(y_0, y_1) \cdots Q(y_{l-1}, y_l)$$

et donc, sous notre hypothèse $m_p(\sigma^{-n}(C) \cap C') \longrightarrow m_p(C)m_p(C')$. Cette convergence pour les cylindre entraîne cette convergence pour tous les ensembles mesurables C et C', c'est à dire le mélange de m_p.

Supposons que m_p est ergodique et p à coefficients strictement positifs. Dans ce cas on a la convergence de $m_p(\sigma^{-1}(C) \cap C')$ vers $m_p(C)m_p(C')$ pour tous C et C' mesurables. En appliquant cette convergence lorsque $C' = \pi_0^{-1}(i)$ et $C = \pi_0^{-1}(j)$, on obtient que $p(i)Q^{n-1}(i,j) \longrightarrow p(i)p(j)$ pour tous i, j. Comme $p(i) > 0$ ceci implique que $Q^n \longrightarrow P$.

Finalement, si $Q^n \longrightarrow P$, et si P est à coefficients strictement positifs, alors il existe n tel que Q^n est à coefficients strictement positifs. ►

On se place maintenant dans le contexte du chapitre précédent, avec $X = [0, 1]$, et une application expansive $\varphi : X \longrightarrow X$. Nous avons montré dans le chapitre précédent qu'il existe une mesure invariante m admettant une densité Lipschitz par rapport à la mesure de Lebesgue et que cette mesure est mélangeante. Nous donnons maintenant une nouvelle preuve de ce résultat. Cette méthode est plus quantitative sur les vitesses de convergence.

On considère comme au chapitre précédent l'ensemble des mesures de probabilité engendrées par une densité f satisfaisant $\mathrm{Lip}\log\circ f \leqslant L$. Plus précisément, on considère l'ensemble $W_L \subset \mathcal{P}(X)$ des probabilités dont la densité f satisfait

$$f(y) < f(x)e^{L|y-x|}$$

pour tous $x \neq y$. C'est un ensemble convexe qui vérifie la condition 13.3. On identifie les éléments de W_L à leur densité, c'est à dire à une partie de l'espace de Banach $E = C(X)$. On note θ_L la distance de Hilbert sur W_L.

On constate d'abord que $\|f\|_{C^0} \leqslant e^L$ pour tout $f \in W_L$. Le convexe W_L est donc borné, ce qui implique que la distance θ controle la distance uniforme.

Nous avons démontré dans le chapitre précédent que, si L est assez grand, alors il existe $L' < L$ tel que $\varphi_*(\mathcal{P}_L) \subset \mathcal{P}_{L'}$.

Proposition 13.9. *Si $L' < L$, alors le diamètre de $W_{L'}$ pour θ_L est fini. L'application φ_* est donc une contraction de W_L.*

◄ C'est facile, mais un peu laborieux. On remarque d'abord que le W_L est défini par les fonctions linéaires $f \mapsto e^{L|y-x|}f(y) - f(x)$. On a donc

$$\theta_L(f, g) = \log \sup_{x \neq y, u \neq v} \frac{\left(e^{L|y-x|}f(y) - f(x)\right)\left(e^{L|v-u|}g(v) - g(u)\right)}{\left(e^{L|y-x|}g(y) - g(x)\right)\left(e^{L|v-u|}f(v) - f(u)\right)}.$$

Si f et g sont dans $W_{L'}$, on a alors

$$\begin{aligned}\theta_L(f,g) &\leqslant \log \sup_{x\neq y, u\neq v} \frac{\left(e^{L|y-x|}-e^{-L'|y-x|}\right)f(y)\left(e^{L|v-u|}-e^{-L'|v-u|}\right)g(v)}{\left(e^{L|y-x|}-e^{L'|y-x|}\right)g(y)\left(e^{L|v-u|}-e^{L'|v-u|}\right)f(v)}\\ &\leqslant \log \sup_{x\neq y, u\neq v} \frac{\left(e^{L|y-x|}-e^{-L'|y-x|}\right)\left(e^{L|v-u|}-e^{-L'|v-u|}\right)}{\left(e^{L|y-x|}-e^{L'|y-x|}\right)\left(e^{L|v-u|}-e^{L'|v-u|}\right)}e^{2L'|v-y|} \leqslant 2\log\frac{e^L e^{L'}}{e^L-e^{L'}}. \blacktriangleright\end{aligned}$$

Pour tout $\mu \in W_L$, la suite $\varphi_*^n\mu$ est donc de Cauchy pour θ, donc pour la distance uniforme des densités. Il existe donc une fonction $h \in W_L$ telle que la densité de $\varphi_*^n\mu$ converge uniformément vers h. Cette fonction est unique.

Montrons maintenant comment ce comportement asymptotique implique le mélange. On va montrer que

$$\int f\circ\varphi^n g\,dm \longrightarrow \int f\,dm \int g\,dm$$

pour toutes fonctions f et g dans L^2. Par un argument de densité classique, il suffit de le montrer pour des fonctions f et g continues.

On remarque dans un premier lieu que cette convergence a lieu pour f continue et $g \in W_L$ (L assez grand). En effet dans ce cas, on a $gm = gh\lambda \in W_{2L}$, donc $\varphi_*^n(gm) \longrightarrow m$ et donc, pour f continue,

$$\int f\circ\varphi^n g\,dm = \int f\,d(\varphi_*^n(gm)) \longrightarrow \int f\,dm.$$

Pour de telles fonctions, cette convergence est même exponentiellement rapide, et on peut en estimer le taux (on parle de décroissance exponentielle des corrélations). C'est un des intérets de la méthode de fournir des estimations sur le vitesse de convergence pour les fonctions régulières, alors que la méthodes du chapitre précédent utilisait des fonctions $\mathcal{A}_n$-mesurable, qui sont moins naturelles du point de vue de l'espace X.

Considérons maintenant une fonction g Lipschitz. Alors $g = 1 + g^+ - (1+g^-)$ où g^+ et g^- sont les parties positive et négative de g, donc sont des fonctions Lipschitz. Comme $1+g^\pm$ est une fonction Lipschitz et strictement positive, elle est multiple positive d'un élément de W_L (pour L assez grand). La convergence ci-dessus a donc lieu pour $(1+g^+)$ et pour $(1+g^-)$, donc pour g.

Finalement, on passe comme d'habitude des fonctions Lipschitz aux fonctions continues par densité uniforme des fonctions Lipschitz.