

CLAUDE TRICOT

RUDOLF RIEDI

Attracteurs, orbites et ergodicité

Annales mathématiques Blaise Pascal, tome 6, n° 1 (1999), p. 55-72

<http://www.numdam.org/item?id=AMBP_1999__6_1_55_0>

© Annales mathématiques Blaise Pascal, 1999, tous droits réservés.

L'accès aux archives de la revue « Annales mathématiques Blaise Pascal » (<http://math.univ-bpclermont.fr/ambp/>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
<http://www.numdam.org/>

Attracteurs, orbites et ergodicité

Claude Tricot* Rudolf Riedi †

Résumé

L'attracteur d'un système de fonctions itéré est le support d'une mesure dite *invariante*, ou *auto-similaire* : c'est le point fixe de l'opérateur de Markov dans l'espace métrique des mesures de probabilité à support compact. Un algorithme aléatoire permet la construction de l'attracteur, qui est presque sûrement l'ensemble des points limites d'une orbite. Un théorème ergodique permet de prouver que la fréquence de visite d'un ensemble par cette orbite est presque sûrement égale à la mesure invariante de cet ensemble. Nous donnons ici une démonstration simple de ce résultat connu.

1 Introduction

Dans un espace métrique complet \mathcal{X} , un «système de fonctions itéré» (en anglais IFS) est une famille finie $\{f_1, \dots, f_K\}$ d'applications contractantes pour la distance définie sur \mathcal{X} . Soit h la distance de Hausdorff définie sur l'espace $\mathcal{H}(\mathcal{X})$ des ensembles compacts de \mathcal{X} : avec cette distance, $\mathcal{H}(\mathcal{X})$ est lui-même métrique et complet.

Dans ce nouvel espace l'opérateur défini par

$$F(E) = \bigcup_{i=1}^K f_i(E)$$

est lui-même contractant : il admet un point fixe, qui est un compact de \mathcal{X} noté A_{inv} [3].

Par construction, A_{inv} est *invariant* dans le sens suivant :

$$A_{\text{inv}} = \bigcup_{i=1}^K f_i(A_{\text{inv}}), \tag{1}$$

il est donc la réunion de copies de lui-même. Dans \mathbb{R}^n , on dit qu'il est *auto-similaire* (ou qu'il possède une structure de similitude interne, formule plus élégante mais plus

*Laboratoire de Mathématiques Pures, Université de Clermont-Ferrand II, 63177 Aubière, France

†Dept of ECE, MS 366, Rice University, Houston, TX 77251-1892, USA

longue) si les f_i sont des similitudes, et *auto-affine* si les f_i sont des applications affines.

On appelle A_{inv} l'*attracteur* du système. Les attracteurs servent souvent de modèles en analyse fractale.

Pour construire un tel attracteur, il existe un algorithme très efficace utilisant une suite aléatoire et qu'on appelle le «chaos game» [1, 4]. On se donne K nombres réels p_1, \dots, p_K strictement positifs, et tels que $\cup_{i=1}^K p_i = 1$. En «tirant au hasard» un nombre entier dans l'ensemble $\{1, \dots, K\}$ avec ces probabilités, et ceci une infinité de fois, on réalise une suite $(\sigma_1, \dots, \sigma_n, \dots)$ de valeurs d'une variable aléatoire multinomiale. On fixe ensuite un point de départ x_0 dans \mathcal{X} , et on construit la suite

$$\begin{aligned} x_1 &= f_{\sigma_1}(x_0) \\ x_2 &= f_{\sigma_2}(x_1) \\ &\dots \\ x_{n+1} &= f_{\sigma_{n+1}}(x_n) \end{aligned}$$

Cette suite (x_n) est l'*orbite* de x_0 . Elle ne converge pas en général : elle a plutôt un comportement chaotique, et paraît visiter tous les «sites» de l'attracteur. Plus précisément, on peut démontrer deux résultats :

1. L'ensemble des points limite de (x_n) est dans l'attracteur (par un simple argument de contraction de la distance à A_{inv});
2. Presque sûrement, tout point de A_{inv} est point limite de cette suite (par une application du lemme de Borel-Cantelli).

Le calcul de cette orbite est très rapide, et dans R^2 , 50 000 points suffisent à dessiner précisément l'attracteur, quel que soit le point de départ choisi.

Mais on peut aller plus loin. En effet, l'aspect de la figure obtenue par une orbite aléatoire change beaucoup selon les probabilités choisies. Dans la figure 1 par exemple, le carré aurait été représenté avec un niveau de gris uniforme si on avait pris des probabilités de même valeur $1/4$. Mais le fait d'avoir choisi des valeurs inégales pour les p_i produit des inégalités dans les fréquences de visite locale, d'où cet effet de «texture» dans les niveaux de gris, effet qui paraît aussi mettre en valeur la structure de A_{inv} . L'image indique l'existence d'une mesure «auto-similaire» dont le support est l'attracteur, et dont la valeur sur chaque ensemble est la fréquence de visite dans cet ensemble.

Or on connaît théoriquement l'existence d'une mesure μ_{inv} auto-similaire sur A_{inv} : elle est obtenue [3] comme point fixe dans l'espace $\mathcal{M}(\mathcal{X})$ des mesures de probabilité à support compact dans \mathcal{X} , et vérifie pour tout ensemble mesurable E

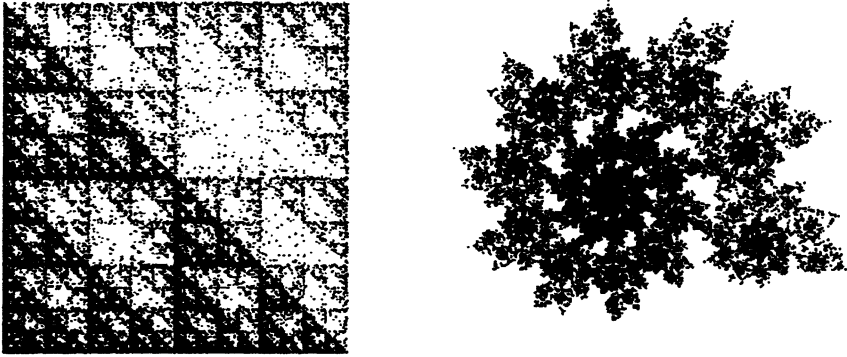


FIG. 1: Les attracteurs de deux systèmes de fonctions itérés. Celui de gauche comporte 4 similitudes de rapport $1/2$; l'attracteur est un carré. Celui de droite comporte deux similitudes. Chaque attracteur a été obtenu par une orbite aléatoire de 50 000 points. Les probabilités sont inégales : remarquer l'effet de texture qui en résulte.

l'équation d'invariance suivante :

$$\mu_{\text{inv}}(E) = \sum_{i=1}^K p_i \mu_{\text{inv}}(f_i^{-1}(E)). \quad (2)$$

On remarque qu'en appliquant l'équation 2 au support de μ_{inv} , on obtient précisément l'équation 1, ce qui montre que le support de μ_{inv} est bien l'attracteur.

La question qui se pose est la suivante :

Cette mesure invariante μ_{inv} est-elle précisément celle représentée par les niveaux de gris ?

dont la réponse, en termes mathématiques, se trouve dans le théorème suivant :

Théorème 1 *Presque sûrement, pour tout ensemble mesurable E la mesure $\mu_{\text{inv}}(E)$ est égale à la fréquence du nombre de points de l'orbite appartenant à E .*

Nous verrons qu'il s'agit d'un théorème ergodique, en fait un cas particulier d'un résultat plus général. Notre démonstration consiste à replacer le problème dans un contexte un peu plus abstrait où l'on peut utiliser la propriété d'ergodicité du *décalage de Bernouilli*. Nous commencerons dans la section 2 par rappeler les notions utiles sur les fonctions mesurables et l'ergodicité, afin de rendre la présentation aussi autonome que possible.

Notons que ce théorème a déjà été démontré dans [2], référence constamment citée par la suite sur cette question. Sans vouloir entrer dans une polémique, disons que la démonstration y est obscure et que les spécialistes préfèrent l'accepter sans vérification. Nous avons cru utile, à des fins pédagogiques, d'en proposer une autre qui a l'avantage d'employer des arguments très simples.

2 Rappels sur la théorie ergodique

2.1 Fonctions préservant la mesure

Un espace mesurable $(\mathcal{X}, \mathcal{A})$ est un espace dans lequel on a défini une σ -algèbre \mathcal{A} . Etant donnés deux espaces mesurables $(\mathcal{X}_1, \mathcal{A}_1)$ et $(\mathcal{X}_2, \mathcal{A}_2)$, une fonction $\mathbf{T} : \mathcal{X}_1 \rightarrow \mathcal{X}_2$ est *mesurable* si pour tout ensemble E de \mathcal{A}_2 , l'ensemble $\mathbf{T}^{-1}(E)$ appartient à \mathcal{A}_1 .

Remarque : Rappelons que \mathbf{T}^{-1} n'est une fonction $\mathcal{X}_2 \rightarrow \mathcal{X}_1$ que si \mathbf{T} est *bijective*. En général, on considère \mathbf{T}^{-1} comme une fonction $\mathcal{A}_2 \rightarrow \mathcal{A}_1$, définie par

$$\mathbf{T}^{-1}(E) = \{x \in \mathcal{X}_1 / \mathbf{T}(x) \in E\}.$$

Soit $\mathbf{T} : \mathcal{X} \rightarrow \mathcal{X}$, on définit les itérées de \mathbf{T} et de \mathbf{T}^{-1} par les relations

$$\mathbf{T}^0 = \mathbf{T}, \mathbf{T}^{n+1} = \mathbf{T}(\mathbf{T}^n), \mathbf{T}^{-n-1} = \mathbf{T}^{-1}(\mathbf{T}^{-n}).$$

Une *mesure* μ sur $(\mathcal{X}, \mathcal{A})$ est une application σ -additive $\mu : \mathcal{A} \rightarrow \mathbb{R}^+$ telle que $\mu(\emptyset) = 0$.

Définition 1 Soit une mesure μ sur $(\mathcal{X}, \mathcal{A})$. La fonction $\mathbf{T} : \mathcal{X} \rightarrow \mathcal{X}$ préserve la mesure μ si pour tout $E \in \mathcal{A}$,

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

Pour montrer que \mathbf{T} préserve la mesure, il suffit de le vérifier sur les ensembles de n'importe quelle famille qui engendre \mathcal{A} .

Exemple 1 : Un *déplacement* de \mathbb{R}^n est une application affine dont la matrice est orthogonale, de déterminant ± 1 , donc bijective. Elle préserve le volume n -dimensionnel. Si μ désigne ce volume, on obtient pour tout borélien E

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

Exemple 2 : Soit $\mathcal{X} = [0, 1]$, avec la mesure binomiale μ déterminée par p, q . La fonction, ou transformation, définie par

$$T(x) = \begin{cases} 2x & \text{si } 0 \leq x \leq 1/2 \\ 2x - 1 & \text{si } 1/2 < x \leq 1 \end{cases}$$

n'est pas bijective. Considérons un intervalle dyadique, du type

$$u = [k2^{-n}, (k+1)2^{-n}], \quad 0 \leq k \leq 2^n - 1.$$

Il est de mesure $\mu(u) = p^N q^{n-N}$, où N est le nombre de 0 dans le développement binaire de $k2^{-n}$, entre les rangs 1 et n . Il a pour image inverse la réunion des deux intervalles :

$$u_1 = [k2^{-n-1}, (k+1)2^{-n-1}], \text{ de mesure } p^{N+1}q^{n-N}, \text{ et}$$

$$u_2 = [1/2 + k2^{-n-1}, 1/2 + (k+1)2^{-n-1}], \text{ de mesure } p^N q^{n+1-N}.$$

Comme $T^{-1}(u) = u_1 \cup u_2$, et $\mu(u) = \mu(u_1) + \mu(u_2)$, T préserve la mesure de tout intervalle dyadique. Or les intervalles dyadiques engendrent la σ -algèbre borélienne de $[0, 1]$. Donc T préserve la mesure binomiale.

2.2 Fonctions mélangeantes et ergodiques

Définition 2 Soit une mesure μ sur (X, \mathcal{A}) . La fonction $T : X \rightarrow X$ est mélangeante (par rapport à μ) si pour tous E, F dans \mathcal{A} ,

$$\lim_k \mu(E \cap T^{-k}(F)) = \mu(E)\mu(F).$$

Ce qui peut se traduire ainsi : L'opération T^{-1} mélange l'ensemble F dans X de telle sorte qu'au bout d'un temps infini, chaque ensemble E en contient une part proportionnelle à sa mesure...

De nouveau, si on veut montrer que T est mélangeante, il suffit de le vérifier sur tous les ensembles E et F appartenant à une famille qui engendre \mathcal{A} .

Définition 3 La fonction T est ergodique si

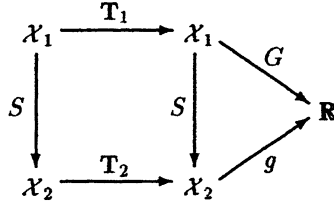
$$E = T^{-1}(E) \implies \mu(E) = 0 \text{ ou } 1.$$

Lorsque T est mélangeante, elle est ergodique : en effet, si $T^{-1}(E) = E$, alors $T^{-k}(E) = E$ pour tout k , donc $\lim_k \mu(E \cap T^{-k}(E)) = \mu(E)$. Si T est mélangeante, on obtient $\mu(E) = \mu(E)^2$, donc $\mu(E) = 0$ ou 1.

Voici la version forte du théorème ergodique dont nous aurons besoin par la suite :

Théorème 2 Soit (X, \mathcal{A}, μ) un espace mesurable tel que $\mu(X) < +\infty$, $G : X \rightarrow \mathbb{R}$ une application μ -intégrable, et $T : X \rightarrow X$ une transformation ergodique qui préserve la mesure. Alors pour μ -presque tout $x \in X$:

$$\lim_{k \rightarrow \infty} \frac{1}{k} \sum_{i=0}^{k-1} G(T^i(x)) = \int_X G d\mu. \quad (3)$$

FIG. 2: Correspondances entre les espaces \mathcal{X}_1 et \mathcal{X}_2 .

2.3 Mesure image

Soit $(\mathcal{X}_1, \mathcal{A}_1)$ et $(\mathcal{X}_2, \mathcal{A}_2)$ deux espaces mesurables, et une application mesurable $S : \mathcal{X}_1 \rightarrow \mathcal{X}_2$. Si $(\mathcal{X}_1, \mathcal{A}_1)$ est muni d'une mesure μ_1 , on peut munir $(\mathcal{X}_2, \mathcal{A}_2)$ d'une mesure en posant pour tout $E \in \mathcal{A}_2$:

$$\mu_2(E) = \mu_1(S^{-1}(E)).$$

On écrit aussi $\mu_2 = \mu_1 \circ S^{-1}$: c'est la *mesure image* de μ_1 .

Soit $T_1 : \mathcal{X}_1 \rightarrow \mathcal{X}_1$ et $T_2 : \mathcal{X}_2 \rightarrow \mathcal{X}_2$ deux applications qui *commuent*, c'est-à-dire telles que

$$S \circ T_1 = T_2 \circ S.$$

Lemme 1 *Supposons que T_1 préserve la mesure μ_1 . Alors T_2 préserve la mesure μ_2 .*

En effet,

$$\begin{aligned} \mu_2(T_2^{-1}(E)) &= \mu_1(S^{-1} \circ T_2^{-1}(E)) \\ &= \mu_1(T_1^{-1} \circ S^{-1}(E)) \\ &= \mu_1(S^{-1}(E)) = \mu_2(E). \end{aligned}$$

Lemme 2 *Supposons que T_1 est ergodique. Alors T_2 est ergodique.*

Car si $T_2^{-1}(E) = E$, alors $T_1^{-1} \circ S^{-1}(E) = S^{-1} \circ T_2^{-1}(E) = S^{-1}(E)$. Donc l'ensemble $S^{-1}(E)$ est invariant dans \mathcal{X}_1 , et $\mu_1(S^{-1}(E)) = 0$ ou 1. Ce sont les deux seules valeurs possibles de $\mu_2(E)$.

Soit une application μ_2 -intégrable $g : \mathcal{X}_2 \rightarrow \mathbf{R}$, et $G = g \circ S$. Alors G est μ_1 -intégrable, et son intégrale est obtenue par la formule de changement de variable suivante :

$$\int_{\mathcal{X}_1} G d\mu_1 = \int_{\mathcal{X}_2} g d(\mu_1 \circ S^{-1}) = \int_{\mathcal{X}_2} g d\mu_2.$$

3 Les décalages de Bernoulli

Exemple 1 On a vu dans la section 2 un exemple de transformation préservant la mesure binomiale μ dans $\mathcal{X} = [0, 1]$. Le même exemple peut être étudié de façon différente, en représentant les points de $[0, 1]$ par leur développement binaire. On considère donc l'espace Σ de toutes les suites

$$\sigma = (\sigma_1, \sigma_2, \dots, \sigma_n, \dots)$$

de 0 et de 1. Si $Y = \{0, 1\}$, on peut écrire Σ sous la forme de l'espace produit

$$\Sigma = \prod_{i=1}^{\infty} Y_i$$

où chaque Y_i est une copie de Y . La mesure ν sur Y déterminée par

$$\nu(\{0\}) = p, \quad \nu(\{1\}) = q$$

fournit une mesure produit P sur Σ , qui à tout ensemble $E(\sigma_1, \dots, \sigma_n)$ de la forme

$$E(\sigma_1, \dots, \sigma_n) = \{\sigma_1\} \times \{\sigma_2\} \times \dots \times \{\sigma_n\} \times \prod_{i=n+1}^{\infty} Y_i$$

attribue la mesure

$$P(E(\sigma_1, \dots, \sigma_n)) = p^{n-N} q^N,$$

avec $N = \sum_{i=1}^n \sigma_i$.

Soit $S : \Sigma \rightarrow [0, 1]$ l'application mesurable qui, à toute suite $(\sigma_n)_{n \geq 1}$, fait correspondre le point x ayant cette suite pour développement dyadique (S n'est pas injective). Chaque ensemble $E(\sigma_1, \dots, \sigma_n)$ a une image par S qui est un intervalle dyadique de $[0, 1]$, noté $I(\sigma_1, \dots, \sigma_n)$, et

$$P(E(\sigma_1, \dots, \sigma_n)) = \mu(I(\sigma_1, \dots, \sigma_n)).$$

De plus,

1. Si $x = S(0, \sigma_2, \sigma_3, \dots)$, alors $x \leq 1/2$, et $2x = S(\sigma_2, \sigma_3, \dots)$.
2. Si $x = S(1, \sigma_2, \sigma_3, \dots)$, alors $x \geq 1/2$, et $2x - 1 = S(\sigma_2, \sigma_3, \dots)$.

On définit maintenant la transformation $T_\Sigma : \Sigma \rightarrow \Sigma$ par

$$T_\Sigma(\sigma_1, \sigma_2, \dots, \sigma_n, \dots) = (\sigma_2, \sigma_3, \dots, \sigma_{n+1}, \dots),$$

qui consiste simplement à éliminer la première composante σ_1 de la suite. Pour tout $\sigma \in \Sigma$.

$$T \circ S(\sigma) = S \circ T_\Sigma(\sigma)$$

où T est la transformation de $[0, 1]$ définie dans l'exemple précédent. Cette égalité montre que T dans $[0, 1]$ commute avec T_Σ dans Σ (section 2). On appelle T_Σ le *décalage de Bernoulli* dans Σ .

a) On remarque que

$$T_\Sigma^{-1}(E(\sigma_1, \dots, \sigma_n)) = Y_1 \times \{\sigma_1\} \times \{\sigma_2\} \times \dots \times \{\sigma_n\} \times \prod_{i=n+2}^{\infty} Y_i$$

qui a la même mesure que $E(\sigma_1, \dots, \sigma_n)$. Ceci entraîne que T_Σ préserve la mesure P . On le savait par ailleurs puisque T_Σ commute avec T .

b) De plus, T_Σ est mélangeante : Prenons deux ensembles $E_1 = E(\sigma_1, \dots, \sigma_n)$ et $E_2 = E(\sigma'_1, \dots, \sigma'_m)$. On sait que

$$T_\Sigma^{-k}(E_2) = \left(\prod_{i=1}^k Y_i \right) \times E_2,$$

de sorte que pour tout $k > n$:

$$E_1 \cap T_\Sigma^{-k}(E_2) = \prod_{i=1}^n \{\sigma_i\} \times \prod_{i=n+1}^k Y_i \times \prod_{j=1}^m \{\sigma'_j\} \times \prod_{i=m+n+1}^{\infty} Y_i.$$

On en déduit que

$$P(E_1 \cap T_\Sigma^{-k}(E_2)) = p^{m+n-N} q^N,$$

avec $N = \sum_{i=1}^n \sigma_i + \sum_{j=1}^m \sigma'_j$. C'est aussi la valeur de $P(E_1)P(E_2)$. La transformation T_Σ est donc ergodique. D'après le Lemme 2, il en est de même de T .

Exemple 2 On peut définir le décalage de Bernoulli dans un contexte plus général. Soit un ensemble fini

$$Y = \{1, \dots, N\},$$

$N \geq 2$, et l'espace produit

$$\Sigma = \prod_{i=1}^{\infty} Y_i$$

où chaque Y_i est une copie de Y . Soit p_1, p_2, \dots, p_N des nombres réels > 0 tels que $\sum p_i = 1$. Les p_i déterminent une mesure sur Y , qui engendre par produit une mesure P sur Σ . Cette mesure correspond à une mesure *multinomiale* sur $[0, 1]$. Chaque ensemble $E(\sigma_1, \dots, \sigma_n)$ de la forme

$$E(\sigma_1, \dots, \sigma_n) = \{\sigma_1\} \times \{\sigma_2\} \times \dots \times \{\sigma_n\} \times \prod_{i=n+1}^{\infty} Y_i$$

a pour mesure

$$P(E(\sigma_1, \dots, \sigma_n)) = p_{\sigma_1} \dots p_{\sigma_n}.$$

La transformation T_Σ est définie par décalage d'indices, comme précédemment.

Exemple 3 On reprend les mêmes notations. On considère l'ensemble Ω des doubles suites

$$\omega = (\dots, \omega_{-2}, \omega_{-1}, \omega_0, \omega_1, \omega_2, \dots)$$

où ω_i appartient à Y . En partageant ω en deux suites $(\omega_0, \omega_{-1}, \omega_{-2}, \dots)$ et $(\omega_1, \omega_2, \dots)$, on voit que Ω peut s'écrire sous la forme du produit cartésien $\Sigma_1 \times \Sigma_2$, où chaque Σ_i est une copie de Σ .

Si P_i est la mesure produit sur Σ_i , alors Ω est muni de la mesure produit $\mu_\Omega = P_1 \times P_2$. La σ -algèbre est engendrée par les ensembles du type

$$\left(\prod_{i=-\infty}^{n-1} Y_i \right) \times \{\omega_n\} \times \{\omega_{n+1}\} \times \dots \times \{\omega_m\} \times \left(\prod_{i=m+1}^{+\infty} Y_i \right),$$

où $n, m \in \mathbb{Z}$, $n \leq m$. Ces ensembles sont de mesure $p_{\omega_n} p_{\omega_{n+1}} \dots p_{\omega_m}$.

On considère la fonction $T_\Omega : \Omega \rightarrow \Omega$ définie par

$$T_\Omega(\dots, \omega_{-2}, \omega_{-1}, \omega_0, \omega_1, \omega_2, \dots) = (\dots, \omega_{-1}, \omega_0, \omega_1, \omega_2, \omega_3, \dots)$$

c'est-à-dire, de nouveau, un simple décalage des indices vers la gauche :

$$(T_\Omega(\omega))_n = \omega_{n+1}.$$

Cette transformation préserve la mesure : en effet on vérifie facilement sur un ensemble du type précédent que $\mu_\Omega(T_\Omega^{-1}(E)) = \mu_\Omega(E)$, et cette égalité est donc vraie pour tout ensemble de la σ -algèbre. Elle est mélangeante : cela se démontre avec les mêmes arguments que pour le décalage de Bernoulli sur Σ . Elle est donc, également, ergodique. Elle s'appelle aussi *décalage de Bernoulli*, sur l'espace Ω .

4 L'ergodicité de la mesure invariante

Soit un espace mesurable $(\mathcal{X}, \mathcal{A})$, et $\mathcal{M}(\mathcal{X})$ l'ensemble des mesures de probabilité à support compact de \mathcal{X} . Soit un système de fonctions itéré $\{f_1, f_2, \dots, f_K\}$, et des probabilités non nulles p_1, \dots, p_K . L'attracteur du système est A_{inv} , tel que

$$A_{\text{inv}} = \bigcup_{i=1}^K f_i(A_{\text{inv}}).$$

Selon la méthodologie décrite en [3], on peut définir dans l'espace métrique, complet $\mathcal{M}(\mathcal{X})$ un opérateur contractant M , défini pour tout $\mu \in \mathcal{M}(\mathcal{X})$ par

$$\forall E \in \mathcal{A}, M(\mu)(E) = \sum_{i=1}^K p_i \mu(f_i^{-1}(E)). \tag{4}$$

C'est l'opérateur de Markov, dont le point fixe est une mesure notée μ_{inv} . Son support est A_{inv} , et elle vérifie l'équation de similarité suivante :

$$\forall E \in \mathcal{A}, \mu_{\text{inv}}(E) = \sum_{i=1}^K p_i \mu_{\text{inv}}(f_i^{-1}(E)). \quad (5)$$

Notations

1. Comme dans la section 3, Σ est l'espace des suites

$$\sigma = (\sigma_1, \sigma_2, \dots, \sigma_n, \dots), \sigma_i \in \{1, \dots, K\}.$$

Il est muni de la mesure P induite par les p_i .

2. Pour tout $x_0 \in \mathcal{X}$, $\sigma \in \Sigma$, on appelle *orbite de x_0* la suite

$$(x_0, \sigma) = (x_n(x_0, \sigma))_{n \geq 0},$$

où $x_0(x_0, \sigma) = x_0$, et $x_n(x_0, \sigma) = f_{\sigma_n}(x_{n-1}(x_0, \sigma))$.

On sait que l'ensemble des points limite de cette orbite est l'ensemble A_{inv} lui-même, pour presque tout σ .

3. Tout point $x \in A_{\text{inv}}$ est déterminé par au moins une *adresse*, c'est-à-dire une suite $\sigma = (a_1, a_2, \dots)$ de Σ telle que pour tout compact K de \mathcal{X} ,

$$\{x\} = \lim_{n \rightarrow \infty} f_{a_1}(f_{a_2}(\dots(f_{a_n}(K))\dots)).$$

On prendra par la suite $K = A_{\text{inv}}$ pour fixer les idées.

On en déduit qu'il existe une application surjective $S : \Sigma \rightarrow A_{\text{inv}}$, telle que $S(\sigma) = x$.

Théorème 3 (Propriété ergodique de μ_{inv}) Soit une fonction μ_{inv} -intégrable $g : \mathcal{X} \rightarrow \mathbf{R}$. Pour μ_{inv} -presque tout x_0 , et pour P -presque tout σ ,

$$\lim_{k \rightarrow \infty} \frac{1}{k} \sum_{i=0}^{k-1} g(x_i(x_0, \sigma)) = \int_{\mathcal{X}} g d\mu_{\text{inv}}. \quad (6)$$

Nous verrons que le même résultat est vrai pour tout x_0 , si g vérifie des conditions supplémentaires (voir les sections 6 et 7).

Avant la démonstration de ce théorème, il faut établir que l'application surjective S fait correspondre entre elles les mesures P sur Σ et μ_{inv} sur A_{inv} .

Lemme 3 L'application S est mesurable, et

$$\mu_{\text{inv}} = P \circ S^{-1}. \quad (7)$$

Démonstration Il suffira de prouver que la mesure $\nu = P \circ S^{-1}$ est une mesure invariante par l'opérateur de Markov M . Remarquons d'abord que le support de ν est inclus dans A_{inv} , donc qu'il est compact ; de même pour $M(\nu)$.

Soit $E \in \mathcal{A}$, et $\sigma \in \Sigma$. Soit i la première composante de σ . En utilisant le décalage de Bernoulli T_Σ sur Σ , $\sigma = (i, \sigma')$, où $\sigma' = T_\Sigma(\sigma)$.

On remarque que σ appartient à $S^{-1}(E)$ si le point $f_i(S(\sigma'))$ est dans E , donc, par continuité de f_i , si $\sigma' \in S^{-1}(f_i^{-1}(E))$. On en déduit

$$\cup_{i=1}^N \{(i, \sigma') / \sigma' \in S^{-1}(f_i^{-1}(E))\} \subset S^{-1}(E).$$

En prenant la mesure P de ces ensembles :

$$\sum_{i=1}^N p_i P(S^{-1}(f_i^{-1}(E))) \leq P(S^{-1}(E)),$$

ce qui peut s'écrire aussi

$$M(\nu)(E) \leq \nu(E),$$

pour tout E . Etant de même poids, ces deux mesures sont égales. Par unicité, on obtient que $\nu = \mu_{\text{inv}}$.

Démonstration du théorème 3 On va utiliser directement le théorème ergodique 2 dans l'espace Ω des double suites

$$\omega = (\dots, \omega_{-2}, \omega_{-1}, \omega_0, \omega_1, \omega_2, \dots)$$

considéré dans la section 3, muni de la mesure μ_Ω , et du décalage de Bernoulli T_Ω . L'argument crucial de la démonstration consiste à partager chaque double suite en deux suites, dont l'une détermine le point de départ de l'orbite (c'est l'adresse), et l'autre détermine les images successives de ce point. On considère donc

- La suite $(\dots, \omega_{-2}, \omega_{-1}, \omega_0)$, que l'on écrit en la changeant de sens $(a_1, a_2, \dots, a_n, \dots)$, avec $a_n = \omega_{-n+1}$. Son image par S donne un point de A_{inv} noté $x_0(\omega)$. Ainsi

$$x_0(\omega) = \lim_{n \rightarrow \infty} f_{\omega_0}(f_{\omega_{-1}}(\dots(f_{\omega_{-n}}(A_{\text{inv}}))\dots)).$$

- La suite $(\omega_1, \omega_2, \dots, \omega_n, \dots)$ que l'on note $\sigma(\omega)$.

La donnée d'une double suite ω est donc équivalente à celle de l'orbite $(x_0(\omega), \sigma(\omega))$. On définit une application $U : \Omega \rightarrow \mathcal{X}$, telle que

$$U(\omega) = x_0(\omega).$$

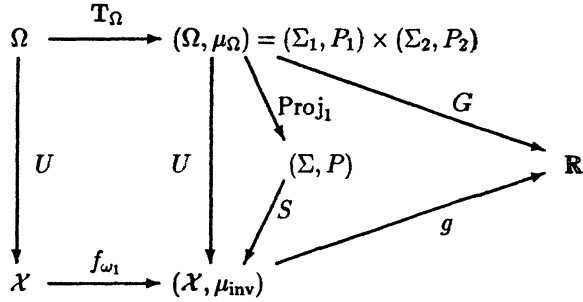


FIG. 3: Correspondances entre les espaces Ω et \mathcal{X} , étant donné ω .

a) Montrons d'abord que pour tout $\omega \in \Omega$,

$$f_{\omega_1} \circ U(\omega) = U \circ T_\Omega(\omega).$$

En effet, T_Ω effectue le décalage d'indices $(T(\omega))_n = \omega_{n+1}$, de sorte que

$$x_0(T_\Omega(\omega)) = \lim_{n \rightarrow \infty} f_{\omega_1}(f_{\omega_0}(f_{\omega_{-1}}(\dots(f_{\omega_{-n}}(A_{\text{inv}}))\dots))).$$

Par continuité de f_{ω_1} , on obtient

$$x_0(T_\Omega(\omega)) = f_{\omega_1}(x_0(\omega)),$$

ce qui prouve l'égalité cherchée. On voit que la transformation T_Ω ne change pas l'orbite elle-même, elle ne fait que remplacer le point de départ $x_0(\omega)$ par sa première image $x_1(x_0(\omega), \sigma(\omega))$.

On peut aussi écrire :

$$U \circ T_\Omega(\omega) = x_1(x_0(\omega), \sigma(\omega)),$$

et pour tout $i \geq 1$:

$$U \circ T_\Omega^i(\omega) = x_i(x_0(\omega), \sigma(\omega)).$$

b) L'application U permet aussi d'établir une correspondance entre les mesures sur Ω et sur \mathcal{X} . On a :

$$\mu_{\text{inv}} = \mu_\Omega \circ U^{-1}.$$

En effet, on peut écrire $\Omega = \Sigma_1 \times \Sigma_2$. Si Proj_i désigne la projection de Ω sur Σ_i , on obtient $U = S \circ \text{Proj}_1$, et pour tout $E \in \mathcal{A}$:

$$U^{-1}(E) = S^{-1}(E) \times \Sigma_2.$$

La mesure $\mu_\Omega(U^{-1}(E))$ vaut donc $P(S^{-1}(E))P(\Sigma_2) = P(S^{-1}(E))$, qui est égal à $\mu_{\text{inv}}(E)$ d'après le lemme 3.

c) Soit $G = g \circ U$: C'est une application $\Omega \rightarrow \mathbf{R}$, à laquelle on va appliquer le théorème ergodique. Tout d'abord, pour tout $i \geq 0$,

$$G(\mathbf{T}_\Omega^i(\omega)) = g(U(\mathbf{T}_\Omega^i(\omega))) = g(x_i(x_0(\omega), \sigma(\omega))).$$

De plus,

$$\int_\Omega G d\mu_\Omega = \int_\Omega g \circ U d\mu_\Omega = \int_{\mathcal{X}} g d(\mu_\Omega \circ U^{-1}),$$

par changement de variable. Comme $\mu_\Omega \circ U^{-1} = \mu_{\text{inv}}$, cela donne

$$\int_\Omega G d\mu_\Omega = \int_{\mathcal{X}} g d\mu_{\text{inv}}.$$

Le théorème ergodique implique que

$$\lim_{k \rightarrow \infty} \frac{1}{k} \sum_{i=0}^{k-1} g(x_i(x_0(\omega), \sigma(\omega))) = \int_{\mathcal{X}} g d\mu_{\text{inv}}.$$

pour μ_Ω -presque tout ω . Tout ensemble de Ω de mesure μ_Ω nulle se projette par U sur un ensemble de \mathcal{X} de mesure μ_{inv} nulle, et par Proj_2 sur un ensemble de Σ de mesure P nulle. On peut donc écrire aussi bien :

$$\lim_{k \rightarrow \infty} \frac{1}{k} \sum_{i=0}^{k-1} g(x_i(x_0, \sigma)) = \int_{\mathcal{X}} g d\mu_{\text{inv}}.$$

pour μ_{inv} -presque tout x_0 et pour P -presque tout σ .

5 Démonstration dans un autre espace

On peut préférer travailler dans l'espace produit

$$\Omega' = \mathcal{X} \times \Sigma,$$

muni de la mesure $\mu_{\Omega'} = \mu_{\text{inv}} \times P$. C'est une approche plus directe, puisque chaque orbite est donnée par un point de départ (dans \mathcal{X}) et une suite (dans Σ). Cependant les arguments qui seront utilisés sont en fait équivalents à ceux de la section 4.

On définit une transformation $\mathbf{T}_{\Omega'} : \Omega' \rightarrow \Omega'$ en posant

$$\mathbf{T}_{\Omega'}(x_0, \sigma) = (x_1(x_0, \sigma), \mathbf{T}_\Sigma(\sigma)) = (f_{\sigma_1}(x_0), (\sigma_2, \sigma_3, \dots)).$$

Cette transformation change simplement le point de départ de l'orbite.

Proposition 1 *La transformation $\mathbf{T}_{\Omega'}$ préserve la mesure $\mu_{\Omega'}$, et elle est ergodique.*

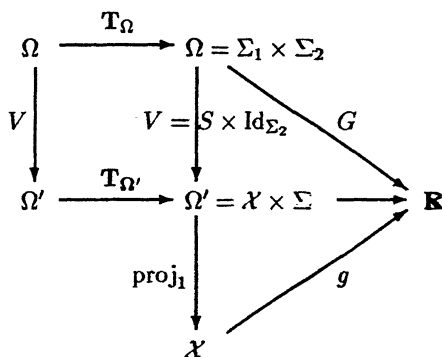


FIG. 4: Correspondances entre les espaces Ω , Ω' et \mathcal{X} .

Démonstration On définit une application

$$V = S \times \text{Id}_{\Sigma_2} : \Omega = \Sigma_1 \times \Sigma_2 \longrightarrow \Omega'$$

en posant $V(\omega) = (x_0(\omega), \sigma(\omega))$. On voit que l'on a déjà utilisé V implicitement dans la démonstration précédente. On a

$$T_{\Omega'} \circ V = V \circ T_\Omega,$$

ce qui veut dire que $T_{\Omega'}$ est l'image dans Ω' de T_Ω dans Ω , au moyen de V . De plus, $\mu_{\Omega'} = P \circ S^{-1} \times P$ (lemme 3) est égale à $\mu_\Omega \circ V^{-1}$. Donc $\mu_{\Omega'}$ est bien la mesure image de μ_Ω . Les résultats de la section 2 démontrent que $T_{\Omega'}$ préserve la mesure, et qu'elle est ergodique, tout comme T_Ω .

Soit maintenant proj_1 la projection $\Omega' = \mathcal{X} \times \Sigma \longrightarrow \mathcal{X}$, telle que $\text{proj}_1(x_0, \sigma) = x_0$. On a $\mu_{\text{inv}} = \mu_{\Omega'} \circ \text{proj}_1^{-1}$. On pose $G' = g \circ \text{proj}_1$. C'est une application $\Omega' \longrightarrow \mathbf{R}$, à laquelle on va appliquer le théorème ergodique. Pour tout $i \geq 1$,

$$G'(T_{\Omega'}^i(x_0, \sigma)) = G'((x_i(x_0, \sigma), T_\Sigma^i(\sigma))) = g((x_i(x_0, \sigma))).$$

De plus,

$$\int_{\Omega'} G' d\mu_{\Omega'} = \int_{\Omega'} g \circ \text{proj}_1 d\mu_{\Omega'} = \int_{\mathcal{X}} g d(\mu_{\Omega'} \circ \text{proj}_1^{-1}) = \int_{\mathcal{X}} g d\mu_{\text{inv}}.$$

Le théorème ergodique implique que

$$\lim_{k \rightarrow \infty} \frac{1}{k} \sum_{i=0}^{k-1} g(x_i(x_0, \sigma)) = \int_{\mathcal{X}} g d\mu_{\text{inv}}$$

pour $\mu_{\Omega'}$ -presque tout (x_0, σ) , donc pour μ_{inv} -presque tout x_0 et pour P -presque tout σ .

6 Cas où la fonction g est continue

Il n'est guère possible d'améliorer le théorème 3 en ce qui concerne les suites $\sigma = (\sigma_n)$. En effet, certaines orbites ne donneront pas le résultat annoncé : si par exemple σ est constante à partir d'un certain rang (l'une des fonctions f_i est choisie systématiquement à chaque étape), alors l'orbite converge vers un point fixe, et la fréquence de visite d'un ensemble ouvert est soit 0, soit 1. Mais ces orbites particulières sont négligeables, au sens de la mesure P .

En revanche, il est intéressant de chercher un résultat identique à celui du théorème 3 qui soit valable pour tout point de départ x_0 (et non plus pour presque tout x_0). Ceci impose une condition supplémentaire sur la fonction g .

Théorème 4 Soit une fonction continue $g : X \rightarrow \mathbb{R}$. Pour tout $x_0 \in X$, et pour P -presque tout σ ,

$$\lim_{k \rightarrow \infty} \frac{1}{k} \sum_{i=0}^{k-1} g(x_i(x_0, \sigma)) = \int_X g d\mu_{\text{inv}}. \quad (8)$$

La démonstration se fait en plusieurs étapes.

Lemme 4 On suppose g uniformément continue et bornée. Pour tous $x_0, y_0 \in X$, et pour tout $\sigma \in \Sigma$,

$$\lim_{k \rightarrow \infty} \frac{1}{k} \sum_{i=0}^{k-1} (g(x_i(x_0, \sigma)) - g(x_i(y_0, \sigma))) = 0.$$

Démonstration Pour tout $\epsilon > 0$, il existe $\eta(\epsilon)$ tel que

$$d(x, y) \leq \eta(\epsilon) \implies |g(x) - g(y)| \leq \epsilon.$$

Posons $x_n = x_n(x_0, \sigma)$ et $y_n = x_n(y_0, \sigma)$ pour faire court. Pour tout n , la distance $d(x_n, y_n)$ est au plus $\rho^n d(x_0, y_0)$, où ρ est le rapport de contraction maximum des f_i . Soit $N(\epsilon)$ tel que $\rho^{N(\epsilon)} d(x_0, y_0) \leq \eta(\epsilon)$. Soit $k > N(\epsilon)$. On a :

$$\left| \sum_{i=0}^{k-1} (g(x_i) - g(y_i)) \right| \leq \sum_{i=0}^{N(\epsilon)-1} |g(x_i) - g(y_i)| + \sum_{i=N(\epsilon)}^{k-1} |g(x_i) - g(y_i)|,$$

donc

$$\frac{1}{k} \sum_{i=0}^{k-1} |g(x_i) - g(y_i)| \leq \frac{2N(\epsilon)}{k} \sup_X |g| + \frac{k - N(\epsilon)}{k} \epsilon.$$

En faisant tendre k vers $+\infty$, et ensuite ϵ vers 0, on obtient le résultat cherché.

Lemme 5 Soit (K_n) une suite de compacts convergeant (au sens de la distance de Hausdorff) vers un compact K . Alors l'ensemble $H = K \cup (\cup_n K_n)$ est compact.

Démonstration Il faut montrer d'abord que si V est un ouvert contenant K , alors tous les K_n sont dans V à partir d'un certain rang. Ensuite, on prend un recouvrement ouvert de H , et on en tire un recouvrement fini.

Démonstration du théorème On choisit une orbite (x_0, σ) qui vérifie l'équation

$$\lim_{k \rightarrow \infty} \frac{1}{k} \sum_{i=0}^{k-1} g(x_i(x_0, \sigma)) = \int_{\mathcal{X}} g d\mu_{\text{inv}}.$$

Soit y_0 un point quelconque de \mathcal{X} . On définit par récurrence la suite de compacts $K_0 = \{x_0, y_0\}$, $K_1 = \cup_i f_i(K_0)$, ..., $K_{n+1} = \cup_i f_i(K_n)$. Cette suite est l'orbite de K_0 par les applications contractantes f_i . Elle converge vers A_{inv} . Les f_i « agissent » sur $H = A_{\text{inv}} \cup (\cup_n K_n)$, en ce sens que $f_i(H) \subset H$. Or H est compact (lemme 5), donc g est uniformément continue sur H , et bornée. Comme x_0 et y_0 sont dans H , on en déduit (lemme 4) que

$$\lim_{k \rightarrow \infty} \frac{1}{k} \sum_{i=0}^{k-1} (g(x_i(x_0, \sigma)) - g(x_i(y_0, \sigma))) = 0.$$

Par conséquent,

$$\lim_{k \rightarrow \infty} \frac{1}{k} \sum_{i=0}^{k-1} g(x_i(y_0, \sigma)) = \int_{\mathcal{X}} g d\mu_{\text{inv}}.$$

Corollaire 1 Soit (g_n) une famille croissante de fonctions continues sur \mathcal{X} . On suppose qu'il existe une fonction intégrable g^* telle que pour tout x ,

$$\sup_n g_n(x) \leq g^*(x).$$

Alors (g_n) tend vers une fonction intégrable g qui a la propriété suivante : pour tout x_0 , et pour P -presque tout σ ,

$$\liminf_{k \rightarrow \infty} \frac{1}{k} \sum_{i=0}^{k-1} g(x_i(x_0, \sigma)) \geq \int_{\mathcal{X}} g d\mu_{\text{inv}}.$$

Démonstration Pour tout x , la croissance de $g_n(x)$ et l'inégalité $g_n(x) \leq g^*(x)$, impliquent que la limite $g(x)$ existe. Le théorème de convergence dominée de Lebesgue montre que

$$\int_{\mathcal{X}} g_n d\mu_{\text{inv}} \longrightarrow \int_{\mathcal{X}} g d\mu_{\text{inv}}.$$

Pour tout n , pour tout x_0 , il existe $\Sigma_n \subset \Sigma$, $P(\Sigma_n) = 1$, tel que pour tout $\sigma \in \Sigma_n$,

$$\lim_{k \rightarrow \infty} \frac{1}{k} \sum_{i=0}^{k-1} g_n(x_i(x_0, \sigma)) = \int_{\mathcal{X}} g_n d\mu_{\text{inv}}.$$

Comme $g_n \leq g$, on en déduit que pour tout n ,

$$\liminf_{k \rightarrow \infty} \frac{1}{k} \sum_{i=0}^{k-1} g(x_i(x_0, \sigma)) \geq \int_{\mathcal{X}} g_n d\mu_{\text{inv}}.$$

On fait ensuite tendre n vers l'infini : on obtient le résultat voulu pour tout x_0 , et pour tout $\sigma \in \bigcap_n \Sigma_n$, donc pour P -presque tout σ .

Le résultat symétrique est vrai aussi :

Corollaire 2 *Soit (g_n) une famille décroissante de fonctions continues sur \mathcal{X} . On suppose qu'il existe une fonction intégrable g^* telle que pour tout x ,*

$$\inf_n g_n(x) \geq g^*(x).$$

Alors (g_n) tend vers une fonction intégrable g qui a la propriété suivante : pour tout x_0 , et pour P -presque tout σ ,

$$\limsup_{k \rightarrow \infty} \frac{1}{k} \sum_{i=0}^{k-1} g(x_i(x_0, \sigma)) \leq \int_{\mathcal{X}} g d\mu_{\text{inv}}.$$

7 Fréquences de visites

Pour tout $x_0 \in \mathcal{X}$, $\sigma \in \Sigma$, $k \geq 1$, et $E \subset \mathcal{X}$, on note

$$N_{(x_0, \sigma)}(E, k)$$

le nombre d'entiers i , $0 \leq i \leq k - 1$, tels que $x_i(x_0, \sigma) \in E$. C'est le nombre de visites de E effectuées par l'orbite entre les temps 0 et $k - 1$.

La fréquence de visite de E par les points de l'orbite est en relation avec la valeur de $\mu_{\text{inv}}(E)$. Voici un résultat dans ce sens :

Corollaire 3 *Soit $E \in \mathcal{A}$. Pour μ_{inv} -presque tout $x_0 \in \mathcal{X}$, et pour P -presque tout σ ,*

$$\lim_{k \rightarrow \infty} \frac{1}{k} N_{(x_0, \sigma)}(E, k) = \mu_{\text{inv}}(E).$$

Démonstration Poser $g = 1_E$, la fonction indicatrice de E , et appliquer le théorème 3.

Mais pour certains ensembles, la fréquence de visite ne dépend pas du point départ.

Théorème 5 *On suppose que l'espace \mathcal{X} est localement compact. Soit un ouvert E de \mathcal{X} dont la fermeture \bar{E} est compacte. Alors pour tout x_0 , et pour P -presque tout σ ,*

$$\mu_{\text{inv}}(E) \leq \liminf_{k \rightarrow \infty} \frac{1}{k} N_{(x_0, \sigma)}(E, k) \leq \limsup_{k \rightarrow \infty} \frac{1}{k} N_{(x_0, \sigma)}(\bar{E}, k) \leq \mu_{\text{inv}}(\bar{E}).$$

Démonstration Comme \mathcal{X} est localement compact, la fonction indicatrice 1_E est une borne supérieure de fonctions continues, donc une limite croissante de fonctions continues. On applique alors le corollaire 1 pour obtenir

$$\mu_{\text{inv}}(E) \leq \liminf_{k \rightarrow \infty} \frac{1}{k} N_{(x_0, \sigma)}(E, k),$$

indépendamment du point de départ. De même, $1_{\bar{E}}$ est borne inférieure de fonctions continues, et on applique le corollaire 2.

Remarque On obtient donc l'égalité

$$\lim_{k \rightarrow \infty} \frac{1}{k} N_{(x_0, \sigma)}(E, k) = \mu_{\text{inv}}(E)$$

pour tout x_0 et presque tout σ , lorsque E est un ouvert dont la frontière $\bar{E} - E$ est de mesure μ_{inv} nulle. Une orbite aléatoire, partant d'un point quelconque, visitera E avec une fréquence égale à sa mesure. Dans ce sens, une orbite aléatoire nous donne, avec probabilité 1, une véritable représentation de μ_{inv} .

Références

- [1] M. Barnsley, *Fractals Everywhere*, Academic Press (1988).
- [2] J. Elton, An Ergodic Theorem for Iterated Maps, *Journal of Ergodic theory and Dynamical Systems*, **7** (1987), 481-488.
- [3] J. Hutchinson, Fractals and Self-similarity, *Indiana University Journal of Mathematics*, **30** (1981), 713-747.
- [4] H.-O. Peitgen, *Chaos and Fractals*, Springer-Verlag (1990).