Théorie Ergodique

Notes de Cours (M2R)

Xavier Bressaud

1[er] octobre 2010

1 Introduction

1.1 Premières impressions

1.1.1 Hypothèse ergodique

(Boltzman) : en deux mots, les moyennes temporelles coincident avec les moyennes spatiales. Formaliser cette idée et développer un cadre dans lequel travailler. Un bon cadre est celui des *systèmes dynamiques*. Les concepts que nous développerons permettent d'étudier ces sytèmes. Définition topologique + Mesure invariante. Résultats dans ce cadre : valable à isomorphisme. Inclus les processus stochastiques (mais pas objet ppal).

1.1.2 Application unimodale

Un joli exemple de famille de systèmes dynamiques :

$$T_a : x \mapsto ax(1-x).$$

Pour des valeurs de a variant entre 0 et 4, T_a définit une application de $[0,1]$ dans lui même. Pour étudier la suite des $T^n x$, un vieux reflexe (nota : $x_{n+1} = T(x_n)$) nous pousse à chercher les points fixes (et comme il y en a...). Bon, pour a petit, tout roule. Un seul point fixe et une dérivée plus petite que 1 en module, ça rentre dans un théorème. Passé un moment, un autre point fixe apparaît, mais ça roule toujours. On ne va plus vers 0, mais vers le "nouveau" autour duquel ça continue de contracter. Bien. Quand la dérivée au point fixe devient plus grande que 1 (plus petite que -1, pour être exact) ça se corse, mais pas trop. On sait faire. On trouve vite une orbite de période 2 telle que le long la dérivée est majorée. Avançons maintenant prudemment ; l'orbite de longueur 2 laisse place à une orbite de longueur 4, puis à une de longueur 8, mais ça va de plus en plus vite. C'est la cascade de doublement de période. Arrive une valeur du paramètre a à partir de laquelle on ne peut plus raisonner suivre les orbites périodiques. A vue de nez, c'est le chaos. En pratique, le comportement du système va maintenant dépendre de manière très irrégulière du paramètre. Pour certaines valeurs, une analyse complète est possible. Par exemple le paramètre extrémal ($a = 4$) est relativement gentil. Variété des comportements possibles : concentration de la dynamique sur un Cantor (attracteur), dynamiques infiniment renormalisables, existence d'une mesure invariante absolument continue par rapport à Lebesgue, combinatoires variées des orbites, théorèmes limites pour certaines observables... (un mot aussi sur ce qui se passerait dans le plan complexe). Autres questions : Que dire de $d(T^n x, T^n y)$? Tend vers 0 ? Oscille ? A quelle vitesse grossit si x et y sont proches ?

1.1.3 Mécanique céleste (ou pas)

Nombreux phénomènes régis par des lois (locales) donnant lieu par exemple à des équations différentielles. Le système est décrit par un point x de l'espace des phases X à l'instant 0 ; il est représenté par un point $f^t(x)$ à l'instant t. Lorsque le "mouvement" est indépendant du temps, on a la propriété de semi-groupe : $f^{t+s} = f^t \circ f^s$. Mouvement des planètes, des partcules dans un gaz, d'un solide, ... Notre objet sera de définir un cadre abstrait rendant compte des propriétés asymptotiques de tels systèmes. Idée de section de Poincarré. Point de vue topologique (voire structure différentiable) et de la théorie de la mesure (volume). Jusqu'à la météo. Idée d'attracteur, éventuellement strange. Idée de flot géodésique sur des surfaces simples ; en particulier modèles billard.

1.1.4 Développement décimal

Soit x un nombre réel (allez, entre 0 et 1 exclu). Regardons son développement en base 10. Par ailleurs considérons l'application définie sur $[0, 1($ qui à x associe la partie fractionnaire de $10x$. On est d'accord pour dire que la suite des premiers digits des $T^n x$ a à voir avec la liste des digits de x. Qu'est-ce alors qu'un nombre normal ? C'est un nombre pour lequel l'hypothèse de Boltzman est satisfaite. C'est le cas de presque tous les nombre sous la mesure de Lebesgue. Non ? Bon, ça ne fait pas de mal. Mais on peut faire des choses plus rigolotes, si par exemple on s'intéresse au développement en fractions continues du réel x. Derrière, il y a l'application de Gauss (ou/et celle de Farey) et les inductions de rotations.

1.1.5 Théorie de l'information

On reçoit une suite de signaux (dans un alphabet, fini, disons). Quel type de structure peut-on détecter ? Périodique ? Complètement "aléatoire" ? Mesure du désordre ? Comparaison ? Notion d'isomorphisme. Plus fin : si on sait, par exemple, que c'est la suite des bords cognés par une boule de billard (bien huilée) peut-on espérer reconstruire la forme du billard ? Chaine de Markov, chaine d'ordre infini ? Codages, dynamique symbolique. Le décalage ? (et aussi dynamique symbolique en plus basse complexité)

1.1.6 Longueur des corrélations

Suite de va iid, chaine de Markov, chaine d'ordre infini. De l'autre coté : suite périodique, orbite de rotation et isométries par morceaux (pendant géométrique), odomètres, application standard. Ordre et désordre. Corrélations. Entropie. Automates cellulaires et systèmes de particules (le coup des feux de forêts : aléa dans la condition initiale).

Exercices 1.1. *Nombre de* 7 *comme premier chiffre du développement décimal de* 2^N*. A relier à la rotation d'angle* $\log_{10} 2$*. Trouver des orbites periodiques dans un billard triangulaire. Boulanger, fat (et thin) baker ainsi que leurs attracteurs. Automorphisme linéaires du tore. Entropie d'un sous-shift de type fini. Adhérence d'une orbite. Reconstruction d'une chaine de Markov (sachant que c'en est une) à partir d'une observation. Ecriture d'un processus stochastique comme un décalage.*

1.2 Cadre abstrait. Principaux résultats

1.2.1 Système dynamique

Un *système dynamique topologique discret* est un couple (X, T) où X est un espace topologique (en général compact) et T une application (en général continue et bijective) de X

dans lui-même. La *théorie ergodique* s'intéresse au comportement asymptotique des *orbites* $\{T^n(x); n \geq 0\}$ du système.

1.2.2 Mesure invariante

Si on munit X de la tribu borelienne, on peut en particulier s'interroger sur l'existence d'une mesure m de probabilité T*-invariante* (au sens où, pour tout borélien A, $m(T^{-1}A) = m(A)$) et chercher à décrire le comportement asymptotique *statistique* du système (sous la mesure m). Le triplet (X, T, m) est appelé *système dynamique (mesuré)*. Un premier résultat garantit l'existence d'une mesure invariante. La question de leur nombre et de leur classification pour un système donné est souvent intéressante. Convergence partout ?

1.2.3 Récurrence

Il va s'agir de comparer la mesure d'un ensemble (disons A) avec le temps passé par une orbite dans cet ensemble. Un premier résultat simple mais important valable dans ce cadre est le théorème de récurrence de Poincarré qui dit essentiellement que :

Théorème 1.1. *Si* $m(A) > 0$*, alors* $\inf(n > 0 : T^n x \in A) < \infty$ *est fini* m*-p.s.*

Il est alors naturel de se demander comment se comporte asymptotiquement le nombre de passages $\sharp\{n \in \mathbb{N} : T^n x \in A\}$ dans le borelien A. Sinon : intervalles errant, ensemble non errant, ensemble ω-limite, attracteurs, ...

1.2.4 Ergodicité

Un ensemble A est dit *invariant* si $T^{-1}(A) = A(\text{mod}0)$ (i.e. $m(A\Delta T^{-1}A) = 0$).La *tribu* $\mathfrak{I}$*des invariants* joue un rôle clef. Un système est dit *ergodique* si cette tribu est triviale (i.e. $\forall A \in \mathfrak{I}, m(A) = 0\text{ou}1$).Le théorème ergodique s'énnonce ainsi :

Théorème 1.2. *Soit* $f \in L^1(m)$,

$$\frac{1}{n}\sum_{k=0}^{n-1} f \circ T^k \longrightarrow_{n\to\infty} \mathbb{E}_m(f|\mathfrak{I}), m - p.s.$$

Si le système dynamique (X, T, m) est ergodique alle théorème ergodique devient :

$$\frac{1}{n}\sum_{k=0}^{n-1} f \circ T^k \longrightarrow_{n\to\infty} \int_X f dm, m - p.s.$$

qui exprime l'idée que les moyennes temporelles convergent vers la moyenne spatiale. Exprimé pour f indicatrice d'un ensemble A,

$$\frac{1}{n}\sharp\{n \in \mathbb{N} : T^n x \in A\} \longrightarrow_{n\to\infty} m(A), m - p.s.$$

Différentes caractérisations de l'ergodicité en partculier à l'aide des corrélations (voir ci dessous). Le point de vue spectral est important lui aussi. Noter que $Uf = f \circ T$ définit un opérateur unitaire de L^2_m. Action sur les Boréliens / sur les fonctions L^2. $\mathcal{B}$ et $L^2(\mathcal{B})$.Théorème ergodique de Von Neuman (convergence dans L^2) et caractérisation spectrale de l'ergodicité (simplicité de la valeur propre 1). Décomposition ergodique.

1.2.5 Mélange

Les corrélations $m(A \cap T^{-n}B) - m(A)m(B)$ mesurent l'indépendance asymptotique. L'ergodicité entraine la convergence en moyenne. Les propriétés de mélange du système décrivent une hierarchie de modes de convergence. La situation la plus forte est l'isomorphisme avec un système de Bernoulli (suite de v.a. iid). Cela mène vers l'idée de classification des systèmes dynamiques.

1.2.6 Isomorphisme

Deux systèmes dynamiques (X, T, μ) et (Y, S, ν) sont *isomorphes* si il existe une application bijective bi-mesurable φ de (X, μ) vers (Y, ν) telle que $S \circ T = T \circ S$. On parle de *facteur* si φ n'est pas bijective. Les propriétés sus-évoquées sont stables par un tel isomorphisme. Une caractéristique clef de la théorie ergodique (en opposition à l'étude des processus) est l'intérêt justement pour les propriétés stables par isomorphisme et donc la classification des systèmes modulo cette équivalence. Dans cet esprit, la recherche d'invariants est une idée standard. Le type spectral de l'opérateur U_f en est un. Un autre très important est l'

1.2.7 Entropie

L'idée initiale est de mesurer le désordre. Notion d'entropie topologique : comptant les orbites de longueur n ($\epsilon-$distinguables, pour ϵ petit) et n grand ; en fait taux de croissance exponantiel. Moins pertinent d'un point de vue métrique. Notion d'information. Entropie d'une partition $\mathcal{P}$: $(H_m(\mathcal{P}) = \sum_i m(P_i) \log m(P_i))$, puis pour dire les choses vite, entropie de cette partition "conditionellement" au passé : quelle information apporte le présent sachant le passé (au filtre d'une partititon). Un petit supremum sur les partitions nous définit un bel invariant. Rapprocher de l'idée de compter les cylindres. Théorème de Shannon. Approfondissons un chouia sur l'idée de codage par une partition :

1.2.8 Codage

Donnée une partition $\mathcal{P} = \{P_i; i \in I\}$. L'application φ qui à un point x de X associe l'indice de l'élément de la partition contenant x permet d'associer à x la suite bi-infinie $\varphi(T^n x)$. Cette application de X dans $I^{\mathbb{Z}}$ définit un facteur. Considérons le décalage sur l'ensemble des suites de I muni de la mesure image. Une telle application est un codage. Théorème : Dans un contexte assez général, il existe un *générateur*, i.e. une partition (dénombrable) qui assure un codage essentiellement bijectif de (X, T, m). Cela signifie qu'il suffit d'étudier le décalage. Théorème d'isomorphisme pour les Bernoulli : deux décalages de Bernoulli sont isomorphes ssi ils ont même entropie.

1.2.9 Processus stochastiques

Soit $(X_n)_{n \in \mathbb{Z}}$ un processus stochastique stationaire à valeurs dans I défini sur son espace de probabilités canonique $(\Omega, \mathcal{A}, \mathbb{P}) = (I^{\mathbb{Z}}, \mathcal{J}^{\otimes \mathbb{Z}}, \mathbb{P})$, i.e. $X_n(\omega) = \omega_n$. Considérons le décalage sur Ω défini par $(T\omega)_n = w_{n+1}$. Le triplet $(\Omega, T, \mathbb{P})$ est un système dynamique mesuré. L'exemple le plus simple est celui d'une suite de va iid à valeurs dans un ensemble fini : on parle de schéma de Benoulli.

1.2.10 Théorèmes limites

Vitesse de décroissance des corrélations. Dépend de la classe d'observable (pas d'invariant ; interessant si on a une structure, par exemple differentielle). Etude des $\sum_{k=0}^{n-1} f \circ T^k$ comme suite de va. Permet alors de montrer des théorèmes limites (TCL, limite locale, résultats de grandes déviations, loi limites de temps d'entrée dans des événements asymptotiquement rares, ...). Notion d'hyperbolicité. Difféo Anosov : dilate ou contracte uniformément. Produits de matrices. Théorème d'Ossedelecs. Formalisme thermodynamique.

1.2.11 Autres cadres

Transformations non-singulières et mesures infinies. Action de groupes. Flots, sections et suspensions. Hypothèses de compacité, de continuité et de bijectivité.

1.2.12 Idées Clefs

Dynamique topologique, systèmes dynamiques et utilisation d'une mesure ; cadre abstrait ; flots et section. Notion d'isomorphisme en théorie ergodique ; force et faiblesse. En particulier codage et importance du shift. Variétés des cadres d'application (mécanique, informatique, modélisation, arithmétique, processus, géométrie des groupes...).

1.3 Plan

Table des matières

2 Dynamique topologique

2.1 Systèmes dynamiques topologiques

Soit X un espace topologique (compact). Un système dynamique topologique discret (X, T) est une application (continue et bijective) $T : X \to X$.

Plus basique : orbite $\mathcal{O}_T(x) = \{T^n x; n \in \mathbb{N}\}$ et son adhérence, orbite périodique : existence de p tel que $T^p x = x$, ensemble invariant strict ou pas. ensemble limite : $\bigcap_{n\geq 0} \bigcup_{k\geq n} T^k x$ (check : fermeture ?), ensemble non errant ? ; attracteur et bassin d'attraction.

Conjugaison Deux systèmes (X, T) et (Y, S) sont (topologiquement) *conjugués* s'il existe une bijection bicontinue $h : X \to Y$, telle que $h \circ T = S \circ h$. Le système (Y, S) est un *facteur* si l'application h est continue mais pas bijective. Les propriétés topologiques d'un système sont stables par conjugaison.

- Une partie A de X est dite *invariante* si $T^{-1}A = A$. ou bien $T^{-1}A \subset A$. Par exemple l'adhérence d'une orbite est un fermé invariant.

  Démonstration. $A = \overline{\{T^n x; n \geq 0\}}$ est fermé. Si est dans A il s'ecrit comme limite d'une suite $x = \lim_n T^{k_n}(x_0)$ alors $T(x) = \lim_n T^{k_n+1}(x_0)$ par continuité de T. □

- Le système est dit *minimal* si toutes les orbites sont denses, ou, ce qui est équivalent, si les seuls fermés invariants sont X et $\emptyset$.

  Démonstration. $\Rightarrow$: Si toutes les orbites sont denses, considérons un fermé F invariant non vide. Soit $x \in F$; par l'invariance, son orbite est dans F et par fermeture, son adhérence aussi. Mais, comme l'orbite est dense, son adhérence est X. $\Leftarrow$: L'orbite d'un point x est un fermé invariant ; c'est donc X et on conclut que l'orbite est dense. □

- Le système est dit *transitif* s'il admet une orbite dense. La transitivité est équivalente à l'existence, pour tout couple d'ouverts B, C, d'un entier $n \geq 0$ tel que $T^{-n}B \cap C \neq \emptyset$.

  Démonstration. Exercice. □

- On dit que le système a la propriété de *mélange (topologique)* si pour tout couple d'ouverts B, C, il existe un entier $N \geq 0$ tel que $\forall n \geq N$, $T^{-n}B \cap C \neq \emptyset$. Remarque : mélange $\Rightarrow$transitivité.

Remarque. On peut en général se contenter de compacité locale et autoriser des discontinuités raisonnables. Si le système n'est pas inversible, on définira son *extension naturelle.*

Premiers exemples Interval exchange transformations, Anosov, Rotations, Piecewise expanding maps... and of course, the shift (S, T) on a subshift S. Action d'un groupe. Sur un Cantor. Homéo d'un cantor. Trouver des ensembles invariants.

Sensibilité aux conditions initiales. Soit (X, d) un espace métrique. On dit qu'un système (X, T) est *sensible au conditions initiales* s'il existe une constante $c > 0$ telle que pour tout $\epsilon > 0$ et tout $x \in X$, il existe un entier n et un point $y \in X$ avec $d(x, y) < \epsilon$ et $d(T^n x, T^n y) > c$. Expansivité. Equicontinuité. Couples asymptotiques.... Exposants de Lyapunov. Pour des systèmes différentiables on s'intéresse au comportement asymptotique de $\frac{1}{n} \log ||D_x T^n(v)||$ pour x pour $x \in X$ et $v \in \mathcal{T}X$.

Induction On définit le *temps de retour* (d'entrée) τ_A et (sur $\tau_A < \infty$) l'application de premier retour dans un borelien A (ou *application induite*) :

$$\tau_A(x) = \inf\{n > 0 : T^n(x) \in A\} \text{ et } T_A(x) = T^{\tau_A(x)}(x).$$

Retours dans le cas minimal. Notation de l'application induite quand cela a du sens.

Convergence des moyennes ergodiques ? Pour certains points, en tous points, vers quelle limite. Ex : si deux valeurs différentes, + transitivité, alors toutes les valeurs ? Action sur une mesure, Existence d'une mesure invariante. Nombre de mesures invariantes ? Unique ergodicité.

2.2 Dynamique symbolique

Quelques bases de dynamique symbolique Alphabet A. L'ensemble $A^* = \cup_{n \geq 0} A^n$ est l'ensemble des mots (finis) sur A, y compris le mot vide. Structure de monoïde libre. Facteur, prefixe, suffixe. Notion de langage.

Shift Soit A un ensemble fini. Par exemple, $A = \{0, 1\}$. On appelle *shift* et on note $\mathcal{X} = \mathcal{X}_A$ l'ensemble $A^{\mathbb{N}}$. On le munit de la topologie (discrète) produit et de la tribu borélienne correspondante. Elles sont engendrées par les *cylindres* $\{c(n, u), n \in \mathbb{N}, u \in A^n\}$, où $c(n, u) = \{x \in \mathcal{X} : x_i = u_i, i = 1, \ldots, n\}$.

Définition 2.1. On appelle décalage l'application continue

$$\begin{array}{rlcl} T: & \mathcal{X} & \to & \mathcal{X} \\ & (x_n)_{n \in \mathbb{N}} & \mapsto & (x_{n+1})_{n \in \mathbb{N}}. \end{array}$$

Cette application n'est pas bijective. Tout point $x \in \mathcal{X}$ a exactement $|A|$ antécédants, $\{ax, a \in A\}$. Shift bilatère.

Définition 2.2. On appelle sous-shift toute partie fermée de $\mathcal{X}$ invariante par T.

Caractérisation d'un sous-shift par son langage., i.e. l'ensemble des mots finis qui apparaissent.

Codage Soit $(X, \widetilde{T})$ un SDT et $\mathcal{P} = \{P_a, a \in A\}$ partition de X. On définit $i : X \to A$ telle que $x \in P_{i(x)}$, et

$$\begin{array}{rll}\varphi : & X & \to \mathcal{A}^{\mathbb{Z}^+} \\ & x & \mapsto (i(\widetilde{T}^n(x)))_{n\geq 0}\end{array}.$$

On dit que $(\overline{\varphi(X)}, T)$ est le *codage* de $(X, \widetilde{T})$ par $\mathcal{P}$. En général, $\varphi(X)$ n'est pas fermé ; il se peut que des suites symboliques ne codant aucune orbite du système initial apparaissent à la fermeture (exemples ?).

Partitions markoviennes Soit (X, T) un système dynamique. Considérons une partition $\mathcal{P} = (P_1, \ldots, P_n)$ de X. On dit qu'elle est *markovienne* si elle a la propriété suivante : Pour chaque i, l'image de l'atome P_i est une union d'atomes : $T(P_i) = \cup_{j\in K_i} P_j$, où les K_j sont des sous ensembles de $\{1, \ldots, n\}$. L'idée est que le codage de l'application par cette partition soit un sous-shift de type fini, au sens suivant.

Sous-shift de type fini

Définition 2.3. Soit L une partie finie de A^2. On appelle *sous-shift de type fini* (SFT) le sous-ensemble $\mathcal{S}$ de $\mathcal{X}$:

$$\mathcal{S} = \{x \in \mathcal{X}; \forall i \in \mathbb{N}, x_i x_{i+1} \notin L\}.$$

Considérons le graphe de sommets A et d'arêtes L et appelons M sa matrice d'incidence. La matrice est dite *irréductible* si le graphe est fortement connexe. Sa *période* est le PGCD des longueurs des cycles du graphe. Elle est dite *apériodique* si sa période est 1. Elle est dite *primitive* si elle admet une puissance dont toutes les entrées sont strictement positives. Une matrice irréductible apériodique est primitive.

Théorème 2.1. *(Perron-Frobenius). Une matrice positive et primitive admet une valeur propre de module maximal réelle, simple, avec des vecteurs propres à coordonnées strictement positives.*

Démonstration. Exercice. □

Exemple 2.1. *Le sous-shift de type fini correspondant à $A = \{0, 1\}$ et $L = \{11\}$ est noté $\mathcal{S}_\varphi$.*

C'est l'ensemble des suites de $\{0, 1\}^{\mathbb{N}}$ ne contenant pas deux 1 successifs. Il est irréductible et apériodique.

Exercices 2.4. *Equlivalence (topologique) des SFT. Eclatement et fusion de sommets.*

Définition 2.5. Soit $x \in \mathcal{X}$. On appelle orbite de x l'ensemble $\Omega = \{T^n(x); n \in \mathbb{N}\}$. Son adhérence, $\overline{\Omega}$ est un sous-shift.

On étend naturellement T — et les définitions ci dessus — à l'ensemble $A^{\mathbb{Z}}$ des suites infinies bilatères. Le décalage est alors une application bijective.

Plus Application transitivité et melange. Recurrence uniforme et minimalité. Mesure (invariante) explicite. Fréquence et mesures invariantes.

Complexité La fonction de *complexité* d'un mot infini u (ou d'un sous-shift) de langage $\mathcal{L}$ est la suite $p_{\mathcal{L}}(n)$ qui compte pour chaque entier n le nombre de mots de longueur n dans le langage $\mathcal{L}$. Cette suite satisfait les propriétés suivantes :

Proposition 2.2. *Soit u un mot infini.*

(i) $\forall n \in \mathbb{N},\ 1 \leq p(n) \leq |A|^n$.

(ii) *La suite p est croissante.*

(iii) $\forall n \in \mathbb{N}, p(n+1) - p(n) = \sum_{w \in F_n(u)} (d(w) - 1) = \sum_{w \in F_n(u)} (g(w) - 1) \leq p(m)p(n)$. *(Noté $s(n)$).*

(iv) $\forall m, n \in \mathbb{N}, p(m+n) \leq p(m)p(n)$.

Démonstration. i trivial. iii implique ii car $d(w) \geq 1$. iii vient de ce que $p(n) = \sum_{w \in F_n(u)} 1$ et $p(n+1) = \sum_{w \in F_n(u)} d(w)$. □

Remarque. On peut limiter la somme $\sum_{w \in F_n(u)} d(w)$ aux facteurs speciaux a droite. Dans le cas binaire, $|A| = 2$, $p(n+1) - p(n)$ est exactement le nombre de facteurs speciaux a droite.

Théorème 2.3 (Morse et Hedlund, 1940)**.** *Soit p est strictement croissante et, dans ce cas, le mot u n'est pas ultimement periodique, soit p est uniformement bornée et le mot u est ultimement periodique.*

Démonstration. Si u est ultimement periodique, $u = x\underline{y}$. Soit $c = |xy|$. Tout facteur apparaissant a la position $n \geq c$ apparait aussi en $n - |y|$. Donc $p(n) \leq c$.

Supposons maintenant qu'il existe n_0 pour lequel $p(n_0) = p(n_0 + 1)$. Alors il n'y a pas de facteur special a droite de longueur n_0 : tout mot de longueur n_0 admet un prolongement unique. Soient n_1 et n_2 deux occurences d'un facteur de longueur n_0. Alors, par recurence sur n, $u_{n_1+n} = u_{n_2+n}$. La suite est ultimement periodique. □

Si u n'est pas ultimement periodique, on a $p(n) \geq n + 1$. On appelle suite *sturmienne* une suite dont la complexité satisfait pour tout n, $p(n) = n + 1$.

Entropie L'entropie (topologique) d'un mot infini ou d'un sous-shift est le taux de croissance exponentielle de la fonction de complexité. Plus formellement :

Proposition 2.4 (Entropie topologique)**.** *La limite*

$$h_{top}(\mathcal{L}) = \lim_{n \to +\infty} \frac{1}{n} \log p_{\mathcal{L}}(n)$$

existe. Elle est appelée Entropie Topologique *de $\mathcal{L}$ (ou de u mot infini de langage $\mathcal{L}$ ou de $\mathcal{S}$ sous-shift de langage $\mathcal{L}$).*

Démonstration. D'apres (iv), $\log p(m+n) \leq \log p(m) + \log p(n)$. Soit d fixé. Ecrivons, $n = qd + r$ avec $0 \leq r < d$, et $\log p(n) \leq q \log p(d) + \log p(r)$. D'ou, $\frac{1}{n} \log p(n) \leq \frac{\log p(d)}{d} + \frac{\log p(r)}{n}$. Donc la limsup est bornée par $\frac{\log p(d)}{d}$ pour tout d et donc par la liminf de cette dernière suite (qui est la même). La limite existe. □

Exercices 2.6. *Entropie d'un SFT. Sofique différent de SFT. Non-minimalité d'un pile ou face. Convergence de l'entropie topologique d'un sous-shift. Exemple : Fibonacci.*

2.3 Exemples

2.3.1 Isométries et isométries par morceaux

Les isométries les plus simples sont les rotations. Nous y penserons comme à des isométries “par morceaux” de l'intervalle :

Exemple 2.2 (Translations sur un groupe abelien $(\Gamma, .)$)**.**

$$\begin{array}{rccc} T_a : & \Gamma & \to & \Gamma \\ & g & \mapsto & a.g \end{array}$$

Exemple 2.3 (Rotations)**.**

$$\begin{array}{rccc} T_\alpha : & \mathbb{U} & \to & \mathbb{U} \\ & z & \mapsto & e^{2i\pi\alpha} z. \end{array}$$

Exemple 2.4 (Rotations (comme isométrie par morceaux))**.**

$$\begin{array}{rlcl} T_\alpha : & [0,1] & \to & [0,1] \\ & x & \mapsto & x + \alpha (mod\ 1). \end{array}$$

Remarque. Si α est rationnel, le système est périodique. Pour α irrationnel, le système est minimal. La mesure de Lebesgue, invariante, est ergodique mais n'a pas la propriété de mélange faible. Ces transformations sont d'entropie nulle.

Isométries par morceaux Soit X un domaine de $\mathbb{R}^d$ et $\mathcal{P} = (P_1, \ldots, P_n)$ une partition de X. Une isométrie par morceaux de X de partition $\mathcal{P}$ est une application T de X dans X dont la restriction à chaque atome P_i est une isométrie. Buzzi [**?**] a montré que ces transformations sont encore d'entropie nulle. Le cas de translations sur le tore appartient à ce cadre. Mais l'étude d'exemples utilisant la richesse du groupe des isométries de $\mathbb{R}^d$, i.e. contenant des rotations, montre que les dynamiques que l'on peut obtenir sont encore mal comprises. Pour certains exemples simples, l'espace se scinde en trois régions : l'une sur laquelle la dynamique est périodique (iles périodiques éventuellement de périodes variables), l'une sur laquelle la dynamique est mal définie (frontières) et une troisième sur laquelle à lieu une dynamique minimale. Lorsque le domaine et les éléments de la partition sont des polygones de $\mathbb{R}^2$, on parlera d'isométries par morceaux *polygonales* du plan. Les *échanges d'intervalles* sont les isométries par morceaux (directes) bijectives du segment. Les rotations sont les échanges de deux intervalles. Mentionnons aussi les odomètres qu'on peut voir comme des translations sur les entiers p-adiques :

Exemple 2.5 (Odomètre)**.** *L'odomètre de base p est l'application définie sur $\mathcal{X} = (\mathbb{Z}/p\mathbb{Z})^{\mathbb{N}}$ par addition de 1 sur la première coordonnée et transmission de retenue à droite. C'est-à-dire que, pour $x \in \mathcal{X}$, si m est le plus petit indice tel que $x_m \neq p-1$,*

$$T(x) = (0, \ldots, 0, x_m + 1, x_{m+1}, \ldots). \tag{1}$$

2.3.2 Applications dilatantes de l'intervalle

Linéaires par morceaux Soit (I, λ) l'intervalle $I = [0,1]$ muni de la mesure de Lebesgue λ.

Exemple 2.6. *Doublement de l'angle. (continue sur le cercle)*

Exemple 2.7 (β-shift)**.** *Soit $\beta > 1$ un réel. On appelera β-shift, l'application T définie sur $I = [0,1]$ par*

$$T(x) = \beta x (mod\ 1).$$

En général, elle n'est pas markovienne. Mais si $\beta = k$ est entier, alors elle est markovienne puisque chaque intervalle de continuité est envoyé sur I tout entier. Remarquons que dans ce cas, le système (I, λ, T) est conjugué au décalage sur $A = \{1, \ldots, k\}$ muni de la mesure produit uniforme ($p_i = 1/k$). Si β est algébrique, alors on peut trouver une partition markovienne et conjuguer le système a un SFT. En particulier si $\beta = \varphi = \frac{1+\sqrt{5}}{2}$, le codage correspondant à la partition $\{[0, 1/\varphi],]1/\varphi, 1]\}$ est le sous-shift de type fini correspondant à $L = \{11\}$.

Le codage naturel d'une application dilatante affine par morceaux de l'intervalle muni de l'image de la mesure de Lebesgue est une chaîne de Markov. La recherche d'une mesure invariante pour la chaîne de Markov et celle d'une piac pour l'application de l'intervalle sont directement reliées. À partir d'une telle application on peut définir une autre chaîne de Markov, en inversant le sens du temps. On peut définir un processus (à valeurs dans I) en choisissant pour $x \in I$ son successeur $y \in \{T^{-1}x\}$ avec une probabilité proportionnelle à $\frac{1}{|T'(y)|}$. Cette probabilité ne dépendant que des intervalles (de codage) dans lesquels se trouvent x et y, la suite des codes des points obtenus est une chaîne de Markov. Sous la mesure invariante, il s'agit finalement de la même chaîne de Markov regardée en temps renversé.

Dilatation uniforme Un autre exemple est celui d'une application uniformément dilatante de l'intervalle dont chaque intervalle de continuité est envoyé bijectivement sur I tout entier —mais pas nécessairement de manière affine. Le codage de l'application est alors un décalage plein. Si l'on pose

$$\phi(\underline{i}(x)) = -\log |f'(x)|,$$

on peut chercher les états d'équilibre associés à ϕ. Cela donne les mesures invariantes absolument continues par rapport à la mesure de Lebesgue. Heuristiquement, si on sait qu'on est dans un certain intervalle V et qu'on veut savoir d'où on vient, on peut regarder les mesures des preimages de l'intervalle V. Mais la mesure des preimages de V est donnée par les inverses des dérivées sur ces preimages.

Dilatation non uniforme Introduites par Pommeau et Manneville [**?**], les applications de la famille suivante sont dilatantes, mais ont un point fixe neutre (indifférent) en 0. Pour $\gamma > 0$, on définit

Exemple 2.8 (Intermittente)**.**

$$\begin{array}{llll} T_\gamma : & [0,1] & \to & [0,1] \\ & x & \mapsto & x + x^{1+\gamma} (mod\ 1). \end{array}$$

Remarque. Cette application est régulière. Ses propriétés statistiques sont intimement liées à la vitesse à laquelle les orbites passant près du point fixe neutre s'en éloignent. Cette vitesse est controlée par le paramètre γ : plus il est grand, moins le point fixe est répulsif et donc plus les orbites typiques passent de temps près de 0. Pour analyser les conséquences de ce phénomène, il est commode de travailler sur une "approximation" localement affine de cette application. Soit $(c_n)_{n\geq 0}$ une suite décroissant strictement vers 0 avec $c_0 = 1$. Ces points délimitent une partition (mod λ) de I en intervalles $I_n := (c_{n+1}, c_n)$, $n \geq 0$. On peut aussi noter $b_n = |I_n| = c_{n+1} - c_n$. Une suite (b_n) avec $\sum_{n\geq 0} b_n = 1$ détermine ainsi une suite (c_n). La définition suivante prend son intérêt si $c_{n+1}/c_n \to 1$.

Exemple 2.9 (Modèle de Wang)**.** *On appelle modèle de Wang de suite (b_n) l'application T définie sur I qui est affine sur chaque I_n et envoie I_0 sur I et I_n sur I_{n-1} pour tous $n \geq 1$.*

Remarque. Nous considèrerons cette application, introduite par Wang [?], comme modèle localement affine de l'intermittente. On peut lui associer une chaîne de Markov à espace d'états dénombrable correspondant au codage par la partition $\{I_n : n \geq 0\}$. Le cas où $b_n \simeq bn^{-(1+\alpha)}$, soit $c_n \simeq cn^{-\alpha}$ correspond dans le cadre lisse à T_γ, avec $\gamma = \frac{1}{\alpha}$. Pour $\alpha > 1$, ces applications admettent une piac. En revanche si $0 < \alpha < 1$ la seule miac est une mesure infinie.

Considérons l'application de Gauss (des fractions continues) définie par :

Exemple 2.10 (Fractions continues)**.**

$$\begin{array}{rccc} G: & [0,1] & \to & [0,1] \\ & x & \mapsto & \frac{1}{x} - \lceil \frac{1}{x} \rceil. \end{array} \tag{2}$$

On pose en outre $G(0) = 0$.

Remarque. Notons qu'elle est régulière sur les intervalles $]\frac{1}{n+1}, \frac{1}{n}]$ et qu'ils sont tous envoyés sur $[0,1[$. Elle n'est pas uniformément dilatante, mais G^2 l'est. Elle joue un rôle comme exemple d'application "hyperbolique" avec une infinité de branches. En particulier, elle admet une piac dont la densité est $g(x) = (\log 2(1+x))^{-1}$. Mais elle est importante aussi pour la description des propriétés arithmétiques des nombres réels qu'elle fournit. Si on code l'application avec sa partition (infinie) de monotonie, le nom d'un point $x \in I$ est exactement son développement en fractions continues. Par ce biais, elle est directement liée à l'induction des rotations d'angles irrationnels.

Mentionnons aussi l'*application de Farey*, définie par

Exemple 2.11 (Farey)**.**

$$F(x) = \left\{ \begin{array}{ll} \frac{x}{1-x} & si\ x < \frac{1}{2} \\ \frac{1-x}{x} & si\ x > \frac{1}{2}. \end{array} \right. \tag{3}$$

Remarque. Remarquons que l'application de Farey a un point fixe neutre. Au voisinage de 0, elle s'écrit sous la forme $F(x) = x + x^2 + o(x^2)$. Elle admet une miac de densité $f(x) = x^{-1}$ qui n'est pas intégrable. L'application de Gauss est une induite de l'application de Farey au sens où $G(x) = F^{r(x)}$ si $r(x) = n$ quand $x \in]\frac{1}{n+1}, \frac{1}{n}]$.

Exercices 2.7. *Montrer que Gauss est presque une induite de Farey. Induction des rotations.*

On considère la famille suivante, pour $a \in [0,2]$,

Exemple 2.12 (Famille quadratique)**.**

$$\begin{array}{rccc} T_a: & [0,1] & \to & [0,1] \\ & x & \mapsto & 1 - ax^2. \end{array}$$

Pour $a = 2$ elle est conjuguée à une application markovienne affine par morceaux. Pour $a \neq 2$ proche de 2, l'orbite du point singulier est plus difficile à contrôler. Cette famille a été l'objet d'une attention particulière comme exemple d'application unimodale non uniformément hyperbolique. On appelle paramètres de Collet-Eckmann les valeurs de a pour lesquels la dérivée le long de l'orbite du point critique croît exponentiellement. Pour ces paramètres, cette application se comporte essentiellement, du point de vue statistique, comme une application dilatante.

Exemple 2.13. *Produits croisés et cocycles. Marches Aléatoires. Exemple de Kalikow.*

2.3.3 Dimensions supérieures

Exemple 2.14 (Boulanger)**.** *On définit une application* T *de* $X = [0,1] \times [0,1]$ *dans lui même par*

$$T(x,y) = \begin{cases} (2x, \frac{1}{2}y) & si\ x < \frac{1}{2} \\ (2x-1, \frac{1}{2}y + \frac{1}{2}) si & si\ x \geq \frac{1}{2}. \end{cases}$$

La partition $\{[0,1/2[\times[0,1],[1/2,1]\times[0,1]\}$ est markovienne. Cette application munie de la mesure de Lebesgue sur le carré est conjuguée au shift bilatère.

Difféomorphismes Anosov Soit $\mathcal{M}$ une variété riemannienne de dimension d ($d \geq 2$) munie du volume riemannien m et de la distance riemannienne $d(\cdot,\cdot)$. Un difféomorphisme T de $\mathcal{M}$ est dans $\mathcal{C}^{(1+\alpha)}(\mathcal{M},\mathcal{M})$ si la différentielle D_xT au point $x \in \mathcal{M}$ dépend α–Hölder continument de x. On notera alors D la constante de distorsion

$$e^{-Dd(x,y)^\alpha} \leq \frac{|det(D_xT)|}{|det(D_yT)|} \leq e^{Dd(x,y)^\alpha}, \tag{4}$$

On dit que c'est un difféomorphisme *Anosov* s'il existe une décomposition invariante de l'espace tangent $\mathcal{T}_x\mathcal{M} = E^s(x) \bigoplus E^u(x)$ en une direction stable et une direction instable c'est à dire avec $||DT|_{E^s}||_\infty < 1$ et $||DT^{-1}|_{E^u}||_\infty < 1$, et si ces espaces dépendent α-Hölder continument de x. On notera d_s et d_u les dimensions des espaces stables et instables, respectivement. Pour tout $x \in \mathcal{M}$, on définit les variétés stable et instable (globales) du difféomorphisme en posant,

$$W^s(x) = \{y \in \mathcal{M} : d(T^nx, T^ny) \longrightarrow_{n\to\infty} 0\}, \text{et}, W^u(x) = \{y \in \mathcal{M} : d(T^{-n}x, T^{-n}y) \longrightarrow_{n\to\infty} 0\}.$$

Pour une étude détaillée de leurs propriétés, renvoyons à Mañe [**?**]. Les difféomorphismes Anosov admettent une partition markovienne. Elle n'est en général pas explicite, mais permet théoriquement de ramener l'étude de leurs propriétés statistiques à celle d'un sous-shift de type fini. L'exemple suivant est un cas particulièrement simple dans la mesure où sa différentielle est constante. Considérons la matrice :

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

Exemple 2.15 (Un automorphisme du tore)**.**

$$\begin{array}{rccc} T: & \mathbb{T}^2 & \to & \mathbb{T}^2 \\ & v & \mapsto & Mv(mod\,\mathbb{T}^2). \end{array}$$

Pour cet exemple, on peut trouver des partitions markoviennes simples. Pour certaines d'entre elles, le sous-shift de type fini associé est $\mathcal{S}_\varphi$ (Exemple 2.1). On peut exploiter dans des cadres plus généraux l'idée d'*hyperbolicité* qui traduit le fait que l'espace se sépare en une partie strictement dilatée et une partie strictement contractée.

2.3.4 Flots

Définition 2.8 (Flots)**.** Un *flot* sur X est une famille $(\phi_t)_{t\in\mathbb{R}_+}$ d'applications de X dans X indexée par $\mathbb{R}_+$ telle que ϕ_0 soit l'identité et $\phi_{t+s} = \phi_t \circ \phi_s$.

On peut souvent ramener l'étude d'un flot à celle d'un système dynamique, soit en le regardant à intervalles de temps donnés, soit en fixant une "section transverse" et en étudiant l'application de premier retour. Nous ne rentrerons pas plus dans les détails techniques. Mentionnons simplement pour guider l'intuition qu'un *flot sur une variété* $\mathcal{M}$ est en fait un flot défini sur le fibré unitaire tangent. Par exemple,

Exemple 2.16. *Le flot géodésique au temps t sur une variété $\mathcal{M}$ munie d'une métrique raisonnable associe au couple $(x, v) \in \mathcal{M} \times \mathcal{TM}$ le couple (y, w) formé du point y où serait un point parti de x en 0 à vitesse 1 dans la direction v et du vecteur unitaire w représentant sa vitesse à cet instant. Formellement, on peut voir le flot géodésique sur le plan hyperbolique comme un flot sur $SL(2, \mathbb{R})$ défini par*

$$\phi_t(x) = x.g_t, \text{ où } g_t := \begin{pmatrix} e^{\frac{t}{2}} & 0 \\ 0 & e^{-\frac{t}{2}} \end{pmatrix}.$$

Signalons qu'on obtient des informations sur les géodésiques d'une surface de Riemann munie d'une métrique hyperbolique en analysant comment le flot géodésique sur le plan hyperbolique (comme revêtement universel) rencontre des domaines fondamentaux de son groupe fondamental. Ce type de codage de géodésique est à l'origine de la dynamique symbolique (Morse [**?**]).

Exemple 2.17 (Billards)**.** *Soit D un domaine polygonal du plan. On peut définir un flot à l'intérieur du domaine en fixant $\phi_t(x, v) = x + tv$, tant que $x + tv \in D$. On décrit qui se passe lorsque le flot atteint le bord en un point s avec une vitesse v en posant $\phi_{0^+}(s, v) = (s, v - 2(n.v)n)$ quand $(n.v) > 0$, où n est la normale sortante au bord en s. Cela permet de définir un flot pour tous les points qui n'atteignent pas les singularités du bord. Un tel flot est appelé un billard.*

Un codage naturel d'un tel système est obtenu en associant à chaque orbite la suite symbolique des segments du bord qu'elle rencontre. Lorque les angles formés par les segments du bord sont rationnels, on sait transformer ce flot en un flot sur une surface plate sans bord (mais avec des singularités). Dans ce cas, si on se restreint à une direction, on peut analyser le flot à l'aide d'un échange d'intervalle. Ces systèmes sont d'entropie nulle. On peut aussi étendre ces définitions au cas où les segments de la frontière sont remplacés par des courbes lisses. Si les bords sont concaves, les propriétés du système qu'on obtient sont radicalement différentes : on entre dans le cadre de l'hyperbolicité.

Exemple 2.18 (Casse-briques)**.** *A la main.*

Exemple 2.19 (Flots Hamiltoniens)**.** *Sans rentrer dans les détails. On se place sur un espace de dimension $2n$, avec des variables conjuguées et une fonction H. Les équations différentielles décrivant la dynamique s'écrivent :*

$$\begin{cases} \dot{x}_i = \frac{\partial H}{\partial y_i} \\ \dot{y}_i = -\frac{\partial H}{\partial x_i} \end{cases}.$$

Evoquer le théorème KAM.

2.3.5 Actions de groupes

Translation sur un pavage de $\mathbb{Z}^d$. Pavages de $\mathbb{R}^2$. Actions qui ne commutent pas. Action linéaire de $\mathrm{SL}(2, \mathbb{Z})$. Action de sous groupes de groupes d'isométries (sur le plan, le plan hyperbolique, ...). Action par isométries sur un arbre réel. Automates cellulaires.

2.3.6 Processus stochastiques

Soit $(X_n)_{n\in\mathbb{Z}}$ un processus stochastique stationaire à valeurs dans I défini sur son espace de probabilités canonique $(\Omega, \mathcal{A}, \mathbb{P}) = (I^{\mathbb{Z}}, \mathcal{J}^{\otimes\mathbb{Z}}, \mathbb{P})$, i.e. $X_n(\omega) = \omega_n$. Considérons le décalage sur Ω défini par $(T\omega)_n = w_{n+1}$. Le triplet $(\Omega, T, \mathbb{P})$ est un système dynamique mesuré. L'exemple le plus simple est celui d'une suite de va iid à valeurs dans un ensemble fini : on parle de schéma (ou de shift) de Benoulli.

Exemple 2.20 (Marche aléatoire). *Soit* $(\Gamma, +)$ *un groupe et* μ *une mesure sur ce groupe. On considère une suite de va iid à valeurs dans le groupe* Γ *et de loi* μ*,* (Y_n)*. On appelle marche aléatoire de loi de sauts* μ *partant de* $g \in \Gamma$ *le processus* $(X_n)_{n \geq 0}$ *défini par* $X_0 = g$ *et pour tout* $n \geq 1$*,* $X_n = X_{n-1} + Y_n$*.*

Ces processus sont des chaînes de Markov dont les transitions sont homogènes dans l'espace. Leur étude a été initiée par Kesten [**?**]. Un des premiers résultats important fut la caractérisation des groupes sur lesquels une telle marche peut être récurrente (groupes finis et extensions finies de $\mathbb{Z}$ ou $\mathbb{Z}^2$), question liée à la classification des groupes à croissance polynomiale (Gromov [**?**]).

Exemple 2.21 (L'exemple de Kalikow). *Soit* $(Y_n)_{n \in \mathbb{Z}}$ *une suite de va iid uniformes sur* $\{0, 1\}$ *et* $(Z_n)_{n \in \mathbb{N}}$ *une marche aléatoire sur* $\mathbb{Z}$*. On appelle processus de Kalikow le processus* $(X_n)_{n \in \mathbb{N}}$ *défini par :*

$$X_n = Y_{Z_n}.$$

Cet exemple, a été introduit par Kalikow [**?**] pour montrer qu'un processus peut avoir la propriété K sans être Bernoulli. Ainsi c'est un processus "très" mélangeant qui n'est pas Bernoulli.

2.4 Exemples symboliques

Automates cellulaires

Définition 2.9. Soit A un ensemble fini. Un automate cellulaire est une application continue de $A^{\mathbb{Z}^d}$ qui commute avec le shift.

Nota : l'odomètre (Exemple 2.5) n'est pas un automate cellulaire.

Exemple 2.22. *Homeomorphisme d'un Cantor. Homeomorphismes du cercle.*

3 Mesures invariantes

3.1 Espaces de Lebesgue

Définition 3.1. Deux espaces probabilisés $(X, \mathcal{A}, m)$ et $(X', \mathcal{A}', m')$ sont isomorphes (mod 0) s'il existe $X_0 \in \mathcal{A}$ et $X'_0 \in \mathcal{A}$ de mesures 1 et une bijection φ bimesurable entre X_0 et X'_0 satisfaisant $\varphi(\mu) = \mu'$.

Définition 3.2. Un espace probabilisé isomorphe à $([0, 1), \mathcal{B}$, Lebesgue) est appelé espace de Lebesgue. On peut éventuellement lui adjoindre un nombre dénombrable d'atomes.

Un théorème un peu technique dit qu'un espace X muni d'une probabilité diffuse m est un espace de lebesgue si, d'une part, il existe une famille dénombrable $\mathcal{C}$ d'ensembles (mesurables) qui séparent les points de X, et, d'aute part, il existe une classe compacte jouissant de la propriété d'approximation (i.e. classe compacte=pour toute famille d'intersection vide, il existe une sous famille d'intersection vide ; approximation=la mesure de tout mesurable A s'écrit comme supremum des mesures des éléments de la classe inclus dans A). Ce cadre est le bon pour définir proprement des partitions mesurables associées à l'espérance conditionnelle par rapport à une sous tribu. Nous serons peut-être obligé de rentrer un peu plus dans les détails plus loin.

Exemple 3.1. *Les espaces polonais (métrique, complet, séparable), munis de leur tribu borélienne (complétée pour la mesure) et d'une mesure de probabilité sont des espaces de Lebesgue. Passage au produit sans pb.*

Un point de vue sur ce qu'on va faire : étude des automorphismes d'espaces de Lebesgue. On y reviendra. Ce n'est pas mon préféré parcequ'il met la mesure avant de mettre le système dynamique topologique ...

Plus de détails dans les notes de T. Delarue [4]

3.2 Mesure invariante

Définition 3.3 (Mesure invariante)**.** Une mesure m sur $(X, \mathcal{B})$ est dite T-invariante si pour tout borelien $A \in \mathcal{B}$, $m(T^{-1}A) = m(A)$.

Equivalence de la definition sur les boréliens, $\mu(T^{-1}A) = \mu(A)$, et de μ invariante ssi pour toute fonction f mesurable, $f \circ T = f$, μ-ps. Classes de fonctions. Le sens moins trivial : approximation par des sommes d'indicatrices.

Définition 3.4. Le triplet (X, T, m) est appelé *système dynamique (mesuré).*

Un premier résultat garantit l'existence d'une mesure invariante. La question de leur nombre et de leur classification pour un système donné est souvent intéressante. Par exemple si mesure de reference, problematique SRB ou recherche d'une absolument continue (par rapport à une mesure de reference).

Théorème 3.1. *Soit X compact et T continue. Il existe une mesure T-invariante sur $(X, \mathcal{B})$.*

Démonstration. On fixe une valeur d'adhérence de la suite $\mu_n = \frac{1}{n}\sum_{k=0}^{n-1} \delta_{T^k x}$. On utilise la continuité de T pour montrer qu'elle est invariante. □

Exercices 3.5. *Lebesgue invariante pour une rotation, un échange d'intervalle. Mesure de Haar invariante pour les translations. Mesure invariante pour un β-shift raisonnable.*

Exemple 3.2. *Chaine de Markov. Bernoulli. Processus stationnaire. Lebesgue pour rotation. Mesure de Haar pour translation / groupe. Doublement de l'angle. Peut-être faire voir ce que peuvent être les mesures orbitales pour le doublement de l'angle (ou le full shift).*

3.3 Isomorphisme

Définition 3.6. Deux systèmes (X, T, μ) et (Y, S, ν) sont *isomorphes* (ou métriquement conjugués) s'il existe une application $h : (X, \mu) \to (Y, \nu)$ établissant une bijection bi-mesurable entre deux sous ensembles de mesures pleines, telle que $\nu(B) = \mu(T^{-1}B)$ pour tout borélien B de X et $h \circ T = S \circ h$.

C'est à dire que les deux espace mesurés sont isomorphes comme espaces de Lebesgue et les applications T et S conjuguées par l'isomorphisme.

Exercices 3.7. *Montrer que le doublement de l'angle est isomorphe au full shift unilatère. Montrer que la transformation du boulanger est isomorphe au full shift bilatère. Verifier qu'une mesure invariante reste une mesure invariante.*

3.4 Théorie spectrale

Espace $L^2(m)$. Opérateur de composition comme isométrie ; $U : L^2(m) \to L^2(m)$, $f \mapsto f \circ T$. Action sur les fonctions. Sur les indicatrices. Indications sur la théorie spectrale. En particulier valeurs propres.

3.5 Tribu des invariants

Un ensemble mesurable I est dit T-invariant (ou simplement invariant) mod 0, si $\mu(I\Delta T^{-1}I) = 0$, i.e. si $T^{-1}I = I$ *modulo* un ensemble de mesure nulle. Les ensembles invariants (mod 0) forment une tribu notée $\mathcal{I}$. Exemples. Liens avec la reductibilité des chaines de Markov. Mais attention, ici on parle en mesure.

Première définition de l'ergodicité : un système dynamique sera dit *ergodique* si la tribu des invariants est triviale.

Ensemble des mesures invariantes (simplexe, compact convexe). Points extremaux et mesures ergodiques. Unique ergodicité : un système dynamique topologique est réputé *uniquement ergodique* s'il n'admet qu'une seule mesure invariante. [Vérifier que le système dynamique (sous cette mesure) doit alors être ergodique]. Convergence des moyennes ergodiques.

Remarque. Comparaison avec la tribu asymptotique. Et plus tard la tribu de Pinsker.

Exercices 3.8. *Construire un point x du shift pour lequel les moyennes ergodiques ne convergent pas.*

3.6 Existence d'un générateur dénombrable

Une partition de X peut être vue comme une application mesurable de X dans un ensemble fini (ou dénombrable). Elle est génératrice si $\bigvee_{k\in\mathbb{Z}} T^{-k}\mathcal{P} = \mathcal{A}$.

Théorème 3.2. *Tout système dynamique apériodique admet une partition génératrice dénombrable.*

Démonstration. Preuve dans [D], Lemme de Rohlin, page 8. En gros on construit une suite de Tours de Rohlin de plus en plus fines. Et on raffine. .

3.7 Extension naturelle

T mesurable mais non inversible, mais avec une mesure μ invariante. $Y = \{(x_n)_{n\in\mathbb{N}}, Tx_n = x_{n-1}, n \geq 1\} \subset X^{\mathbb{N}}$ se projette sur X par $(x_n)_{n\in\mathbb{N}} \mapsto x_0$. On définit $\tilde{T}$ de Y dans lui-même par, si $y = (x_n)_{n\in\mathbb{N}}$, $\widetilde{(Ty)}_0 = Tx_0$, et $\widetilde{(Ty)}_n = x_{n-1}$. Inversible et facile d'étendre μ en une mesure $\tilde{\mu}$ invariante pour $\tilde{T}$. On pose $\tilde{\mu}_{|\sigma(X_0)} = \mu$; puis pour définir $\tilde{\mu}_{|\sigma(X_0,X_1)}$, on pose $\tilde{\mu}(A) = \tilde{\mu}(T^{-1}A)$ pour $T^{-1}A$ mesurable par rapport à $\sigma(X_0)$; on continue par récurence.

3.8 Mesure physique

Pour un système différentiable (ou plus généralement lorsqu'on a une mesure non singulère ($T^*\mu << \mu$) ou du moins une mesure "naturelle"), on peut s'interroger sur l'existence d'une mesure invariante absolument continue par rapport à la mesure de Lebesgue (ou par rapport à la mesure de référence).

3.9 Mesure infinie

On travaille en général avec des probabilités invariantes. Mais une partie des notions introduites s'étendent au cas de mesures invariantes non nécessairement finies. (remplacer "mesure 1" par "complémentaire de mesure 0"). C'est intéressant par exemple lorsque la seule mesure invariante absolument continue par rapport à une mesure de référence donnée est infinie puisque cette mesure décrit le comportement asymptotique d'orbites typiques pour la mesure de référence.

4 Récurrence

Il va s'agir de comparer la mesure d'un ensemble (disons A) avec le temps passé par une orbite dans cet ensemble.

4.1 Théorème de récurence de Poincarré

Un premier résultat simple mais important valable dans ce cadre est le théorème de récurrence de Poincarré qui dit essentiellement que :

Théorème 4.1. *Si $\mu(A) > 0$, alors $\sharp\{n > 0 : T^n x \in A\} = \infty$ μ-p.s sur A.*

Démonstration. Soit A tel que $\mu(A \cap T^{-n}A) = 0$ pour tout $n > 0$. L'invariance de la mesure entraîne alors que $\mu(A \cup T^{-1}A \cup \cdots \cup T^{-n}A) = n\mu(A)$ pour tout $n > 0$. Comme $\mu(X) = 1$ il ressort que $\mu(A) = 0$. Ainsi $\mu(A) > 0$ entraîne que il existe $n > 0$ tel que $\mu(A \cap T^{-n}A) > 0$ et donc $T^n A$ rencontre A. Bon, pour faire vite et bien la preuve il faut dégager la construction du fameux schéma du temps de retour : $A = A_1 \cup A_2 \cup \cdots \cup A_\infty detellemanierequesur A_n \subset A$, $\tau_A(x) = n$. Il apparaît que Xest l'union disjointe

$$X = \left(\bigcup_{n=1}^{\infty} \bigcup_{k=0}^{n-1} T^k A_n \right) \cup \left(\bigcup_{k=0}^{\infty} T^k A_\infty \right)$$

Clairement, l'argument précédent entraîne que $\mu(A_\infty) = 0$ et donc que le premier temps de retour est fini. Les suivants le sont aussi par invariance, et donc, m-ps, le nombre de passages est infini. □

Il est alors naturel de se demander comment se comporte asymptotiquement le nombre de passages $\sharp\{n \in \mathbb{N} : T^n x \in A\}$ dans le borelien A. Lien entre les sommes ergodiques et les temps de retour (classique). Point récurent.

4.2 Induction

Soit (X, T) un système dynamique et Y une partie de X. Pour tout $x \in X$, on appelle *temps d'entrée dans* Y l'entier

$$\tau(x) = \tau_Y(x) = \inf\{n > 0 : T^n(x) \in Y\}.$$

Si $x \in Y$, on parle aussi de *temps de retour*. On appelle application *induite* de T sur Y, ou application de premier retour, l'application $T_Y(x) = T^{\tau(x)}(x)$. Autrement dit, on associe à un point $x \in Y$ le premier de ses itérés qui est dans Y. Cela revient en quelque sorte à "accélérer" l'application pour ne retenir que les passages du système dans l'ensemble Y. On parle aussi d'induite si τ n'est pas un temps d'entrée, mais une application localement constante à valeurs entières. On définit récursivement les temps de retour successifs dans l'ensemble Y par $\tau^{(1)} = \tau_Y$ et

$$\tau^{(k+1)} = \inf\{n > \tau^{(k)} : T^n(x) \in Y\}.$$

Si μ est une mesure de probabilité invariante, et $\mu(Y) > 0$, le théorème de récurrence de Poincaré garantit que le temps de retour à Y est fini ps. La formule de Kac donne l'espérance du temps de retour : $E(\tau.1_Y) = 1$, c'est-à-dire que si on est sur Y, sa moyenne est $\frac{1}{\mu(Y)}$. La restriction μ_Y de μ à Y est T_Y-invariante. Réciproquement, on peut construire une mesure invariante par T à partir d'une mesure invariante par T_Y en posant

$$\mu(B) = \sum_k \mu_Y(T^k(B \cap \{\tau = k\})). \tag{5}$$

Remarquons que cette mesure est finie si et seulement si le temps de retour, comme variable aléatoire sous μ_Y est intégrable, c'est-à-dire si $\sum_{k\geq 1} k\mu_Y(\{\tau = k\}) < \infty$. Notons que les propriétés de mélange des mesures passent mal entre l'application et son induite.

Considérons la partition de Y par les ensembles $Y_k = Y \cap \{\tau = k\}$. On appelle *tour de Kakutani de base* Y la partition $\mathcal{P}$ de X

$$\mathcal{P} = \{T^j Y_k : k \in \mathbb{N}, 0 \leq j < k\}.$$

Au dessus de Y on construit des "étages" (j =cste) de plus en plus petits. La dynamique de T consiste à monter d'un étage quand c'est possible et à revenir dans Y sinon.

Remarque. Pour des raisons techniques, on peut être amenés à construire les tours à l'envers. Cela donne une partition de la forme, $\mathcal{P} = \{T^{-j}\tilde{Y}_k : k \in \mathbb{N}, 0 \leq j < k\}$, où les $\tilde{Y}_k$ forment une partition de Y et vérifient $T^{-j}\tilde{Y}_k \cap Y = \emptyset$ pour $0 < j < k$.

A priori, ces définitions sont valables pour des systèmes inversibles. Mais elles s'adaptent souvent facilement à des cadres un peu plus généraux.

Exercices 4.1. *Montrer que l'application induite d'une rotation d'angle $\alpha < 1/2$ sur l'intervalle $[0, 1-\alpha]$ est isomorphe à une rotation (d'angle à déterminer). Idem si $\alpha > 1/2$,sur l'intervalle $[0, \alpha]$. Relier à l'application de Farey. Maintenant induire systématiquement sur l'intervalle $[0, \alpha]$. relier à l'application de Gauss. Même genre de chose pour les échanges d'intervalles.*

Exercices 4.2. *On peut dans certaines situations itérer cette procédure d'induction, en induisant sur des sous-ensembles de plus en plus petits. Pour illustrer cette assertion, considérons la rotation T_α d'angle α irrationnel. Supposons $\alpha < 1/2$. Si on induit T_α sur $[0, 1-\alpha]$, on obtient l'application*

$$T_{[0,1-\alpha]}(x) = \begin{array}{ll} x+\alpha & \textit{si } 0 \leq x < 1-2\alpha \\ x+\alpha-1 & \textit{si } 1-2\alpha \leq x < 1-\alpha. \end{array}$$

Si on renormalise l'intervalle $[0, 1-\alpha]$ par une homotétie de rapport $(1-\alpha)^{-1}$, on obtient de nouveau une rotation, mais d'angle $\alpha/(1-\alpha)$. Supposons $\alpha > 1/2$. Une induction sur l'intervalle $[0, \alpha]$ donne

$$T_{[0,\alpha]}(x) = \begin{array}{ll} x+2\alpha-1 & \textit{si } 0 \leq x < 1-\alpha \\ x+\alpha-1 & \textit{si } 1-\alpha \leq x < \alpha. \end{array}$$

Si on renormalise l'intervalle $[0, \alpha]$ par une homotétie de rapport $-\alpha^{-1}$, on obtient de nouveau une rotation, mais d'angle $(1-\alpha)/\alpha$. Ainsi, cette opération d'induction/renormalisation de la rotation d'angle α transforme la rotation d'angle α en une rotation d'angle $F(\alpha)$ où F est l'application de Farey (Exemple 2.11). On peut itérer cette opération. Remarquons que si α est petit, on sera amené à utiliser un grand nombre de fois la première opération. Une autre stratégie consiste à induire systématiquement sur l'intervalle $[0, \alpha]$. On obtient alors, si n désigne la partie entière de $1/\alpha$,

$$T_{[0,\alpha]}(x) = \begin{array}{ll} x+(n+1)\alpha-1 & \textit{si } 0 \leq x < 1-n\alpha \\ x+n\alpha-1 & \textit{si } 1-n\alpha \leq x < \alpha. \end{array}$$

Soit, après renormalisation par $-\alpha^{-1}$, une rotation d'angle $(1-n\alpha)/\alpha$ sur $[0, 1]$. On retrouve l'application de Gauss (Exemple 2.10). Cela revient à répéter l'induction/renormalisation de Farey jusqu'à ce que α soit supérieur à $1/2$.

4.3 Lemme de Kacs

Théorème 4.2. *Soit* (X, T, μ) *un système dynamique ergodique. Soit* $A \in \mathcal{A}$*, avec* $\mu(A) > 0$*, alors*

$$\frac{1}{\mu(A)} \int_X \tau_A d\mu = \frac{1}{\mu(A)}.$$

Démonstration. L'ensemble $\cup_{n\geq 0} T^n A$ est invariant. L'ergodicité entraine donc que $\mu(\cup_{n\geq 0} T^n A) = 1$. Or, $\cup_{n\geq 0} T^n A = \cup_{n\geq 0, k\geq 0} T^n A_k$, réunion disjointe (notation du théorème de récurrence) ; ainsi

$$\sum_{k=0}^{\infty} \sum_{n=0}^{\infty} \mu(T^n A_k) = 1.$$

Autrement écrit : $\sum_{k=0}^{\infty} k\mu(A_k) = 1$. On conclut en écrivant $\int_A \tau_A d\mu = \sum_{k=0}^{\infty} k\mu(A_k)$. □

4.4 Tours de Rokhlin

On appelle tour de Rokhlin (de base B et de hauteur h) une famille finie de parties mesurables de la forme $(B, \mathrm{TB}, \ldots, T^{h-1}B)$, où les $T^j B$ sont deux à deux disjoints. La mesure de la tour est $h\mu(B)$. L'action de T se réduit à monter d'un étage, sauf à partir du dernier !

Théorème 4.3. *(Lemme de Rokhlin). Soit* (X, T, μ) *un système dynamique ergodique (et* μ *non atomique). Pour tout* $\epsilon > 0$*, et tout entier* $h \geq 1$*, il existe une tour de Rokhlin de hauteur* h *et de mesure supérieure à* $1 - \epsilon$*.*

Démonstration. On choisit un ensemble mesurable A de mesure $0 < \mu(A) < \epsilon/(h-1)$. Puisque le système est ergodique, la décomposition $\cup_{n\geq 0, k\geq 0} T^n A_k$ recouvre un ensemble de mesure 1. C'est une réunion disjointe de tours de Rokhlin. On peut décomposer verticalement chaque tour de cette construction en tours de hauteur h (sauf peut-être les r derniers étages de la tour de hauteur k où r est le reste de k/h). Considérons la réunion B des bases de ces tours. Clairement $B, \mathrm{TB}, \ldots, T^{h-1}B$ est une tour de Rokhlin de hauteur h. La contrainte imposée sur la mesure de A entraine que

$$\mu\left(X \setminus (\cup_{n\geq 0} T^n B)\right) \leq \sum_{k\geq 0} \mathrm{reste}(k \div h)\mu(A_k) \leq (h-1)\mu(a) \leq \epsilon.$$

□

Remarque. Le résultat reste valable sous l'hypothèse plus faible d'apériodicité, i.e. nullité de l'ensemble des points periodiques (qui entraîne la non atomicité). Nous ne le prouverons pas. Quoiqu'il nous soit peut-être utile par la suite.

Remarque. Ce type de construction à l'aide de tours est intéressant aussi du point de vue topologique. Et aussi : notion de rang ?

Interprétations de ces résultats (physique, irréversibilité, billards, ...)

Application induite. Mesure invariante pour l'induite. Extension (et suspension).

4.5 Sinon

Intervalles errants, ensemble non errant, ensemble ω-limite, attracteurs, ...

5 Ergodicité

5.1 Définitions

Tribu des invariants et fonctions invariantes. Les fonctions invariantes sont $\mathcal{I}$-mesurables. Les constantes sont invariantes.

Un système (X, T, μ) est *ergodique* si et seulement si $\forall B \in \mathcal{B}, T^{-1}B = B, \mathrm{mod}\ \mu \Rightarrow \mu(B) =$ 0ou $\mu(B) = 1$. Rappelons que cela est équivalent à demander que $\forall B, C \in \mathcal{B}, \mu(B) > 0, \mu(C) > 0$, $\exists n \geq 0$, $\mu(T^{-n}B \cap C) > 0$. Ou encore que $\forall f : X \to \mathbb{R}$, mesurable $f \circ T = f, \mu - \text{ps} \Rightarrow f = \text{cte}, \mu - \text{ps}$ sur X. (f mesurable, f dans L^1, f dans L^2). (Proposition 4.1 p. 28 dans [PB]).

Remarque. Dans de bonnes conditions topologiques l'ergodicité entraîne la densité de μ presque toute orbite. (Proposition 4.2, p. 29 dans [PB]).

Lemme 5.1. *Le sous espace $\mathfrak{C}$ des fonctions de la forme $f = g - g \circ T$ est dense dans l'orthogonal de l'ensemble $\mathfrak{I}$ des fonctions $\mathcal{I}$-mesurables.*

Démonstration. On montrera qu'une fonction f de L^2_μ orthogonale à la fois à $\mathfrak{I}$ et $\mathfrak{C}$ est nulle. $\square$

5.2 Le théorème ergodique : énnoncé

Le *théorème ergodique* (de Birkhoff) assure pour toute fonction $f \in L^1(\mu)$ l'existence d'une fonction f^* invariante de même intégrale telle que

$$\frac{1}{n}\sum_{i=0}^{n-1} f \circ T^i \longrightarrow_{n\to\infty} f^*,\ \mu - \text{ps}.$$

Ainsi si le système est ergodique, les sommes ergodiques convergent vers une constante qui ne peut être que l'intégrale de f :

$$\frac{1}{n}\sum_{i=0}^{n-1} f \circ T^i \longrightarrow_{n\to\infty} \int_X f d\mu,\ \mu - \text{ps}.$$

Ce théorème fournit aussi une caractérisation de l'ergodicité sous la forme,

$$\forall B, C \in \mathcal{B},\ \frac{1}{n}\sum_{i=0}^{n-1} \mu(T^{-i}B \cap C) \longrightarrow_{n\to\infty} \mu(B)\mu(C).$$

$$S_n f = \frac{1}{n}\sum_{i=0}^{n-1} f \circ T^i$$

5.3 Le théorème ergodique : preuve courte

Une première preuve classique consiste à montrer d'abord le dit lemme ergodique maximal. Pour $f \in \mathbb{L}^1(X, \mu, \mathbb{R})$, on définit

$$f^*(x) = \sup_{n\geq 1} \frac{1}{n}\sum_{i=0}^{n-1} f(T^i(x))$$

Lemme 5.2. *Pour toute $f \in \mathbb{L}^1(X, \mu, \mathbb{R})$ on a*

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

Démonstration. On pose $\psi_n = \max_{k=0,\dots,n}\{kS_kf\}$ et $E_n = \{x \in X|\psi_n(x) > 0\}$. Sur E_n, $\psi_n = f + \psi_{n-1} \circ T$. Sur son complémentaire ψ_n=0. Sur X, $\psi_{n-1} \circ T \geq 0$. Donc

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

Or $\{f^*(x) > 0\} = \cup_{n \geq 1} E_n$. Le résultat suit.

$\square$

Nous allons dans un premier temps montrer l'existence de la limite. Il suffit pour cela de montrer que l'ensemble

$$E_{\alpha,\beta} = \left\{x \in X : \liminf_{n\to\infty} S_nf(x) < \alpha < \beta < \limsup_{n\to\infty} S_nf(x)\right\}$$

est de mesure nulle pour tous $\alpha < \beta$. Remarquons que $E_{\alpha,\beta}$ est invariant ($T^{-1}E_{\alpha,\beta} = E_{\alpha,\beta}$). Le fait que $\limsup_{n\to\infty} S_nf \leq f^*$ entraîne que $E_{\alpha,\beta} \cap \{(f-\beta)^* > 0\} = E_{\alpha,\beta}$ ($\{(f-\beta)^* > 0\} \subset E_{\alpha,\beta}$?). Même chose dans l'autre sens. Le lemme maximal appliqué sur $E_{\alpha,\beta}$ aux fonctions $f - \beta$ et $\alpha - f$ donne

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

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

La limite est invariante car la suite $\frac{n+1}{n}S_{n+1}f(x) - S_nf(x) = \frac{1}{n}f(x)$ tend vers 0.

5.4 Le théorème ergodique : une autre preuve

Démonstration. (Ledrappier) Une autre vision du lemme maximal : pour $f \geq 0$, $M_Nf(x) = \sup_{n\leq N} \frac{1}{n}\sum_{i=0}^{n-1} f(T^i(x))$ et $M_*f = \sup_{N\geq 1} M_N$ Le lemme dit que $\mu(M_*f > \lambda) \leq \frac{\int_X f d\mu}{\lambda}$. La preuve utilise $A_N = \{M_Nf > \lambda\}$. pour $J \gg N$, $\sum_{i=0}^{J-1} f(T^i(x)) \geq \lambda \ \geq \lambda\sharp\{i \leq J - N : T^ix \in A_N\}$ et donc en integrant $J \int f \geq \lambda(J-N)\mu(A_N)$. On fait tendre $J \to \infty$ pour voir que $\mu(A_N) \leq \int f/\lambda$, puis $N \to \infty$ pour obtenir le résultat.

Maintenant, $\{f : \text{La somme converge}\}$est un ssev fermé de L^1. On écrit

$$\begin{aligned}\frac{1}{n}\sum_{i=0}^{n-1} f \circ T^i - \mathbb{E}^{\mathcal{I}} f \quad &\leq M_*(f - f_j) + \mathbb{E}^{\mathcal{I}}|f - f_j| + \tfrac{1}{n}\textstyle\sum_{i=0}^{n-1} |f_j \circ T^i - \mathbb{E}^{\mathcal{I}} \quad f_j| \\ \limsup_{n\to\infty} &\leq \frac{\epsilon_j}{\lambda} + \end{aligned}$$

Le point est qu'on peut approcher f (d'integrale nulle ?) par une suite de f_j de la forme $g - Tg$, i.e. pour lesquels le résultat est clairement vrai, de manière à ce que $\|f - f_j\|_1 < \epsilon_j$ tende vers 0. Il faudrait arriver à préciser un peu pour que ce soit sortable. $\square$

5.5 Le théorème ergodique : preuve à la main

La preuve esquissée ici est faite proprement dans les notes de T. Delarue [4].

On définit $f^* = \limsup_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} f(T^ix)$ et f_*. Elles sont invariantes et ordonnées. Il faut montrer l'égalité. On va montrer que si N est suffisament grand, la moyenne ergodique est aussi grande que f^*. Plus précisement :

Lemme 5.3. *$E_b = \{X^* > b\}$. Pour tout $\epsilon > 0$ et tout N assez grand, $\mathbb{P}(E_b \setminus \frac{1}{N}\sum_{k=0}^{N-1} X_k > b - \epsilon) < \epsilon$.*

Démonstration. Pour le dire vite. On décompose $f = f_1 + f_2$, avec f_1 bornée et f_2 d'intégrale petite. Sur E_b, le plus petit entier l tel que $S_l > b - \delta$ est fini ps. On se donne un entier n tel que $\mu(l > n)$ petit. La proba de $M = E_b \cap \{l > n\}$ est petite. On prend N grand. La proba que la proportion des entiers entre 0 et N avec $T^k x \in M$ soit grande est petite. De même, celle que la moyenne des $f_2 \circ T^k$ soit grande est petite. On se place mainetant sur l'ensemble (de mesure grande) des x pour lesquels ces deux quantités sont assez petites. On décompose alors $\{0, \ldots, N\}$ en sous intervalles $\{0, \ldots, j_1 - 1\}, \{j_1, \ldots, j_1 + l_1\}, \ldots, \{j_i, \ldots, j_i + l_i\}, \{j_i + l_i, \ldots, j_{i+1} - 1\}$, alternativement mauvais/bon, de telle manière que j_{i+1} soit le plus petit entier $> j_i + l_i$ tel que $T^{j_i} x \notin M$ et $l_i = l(T^{j_i} x) \leq n$. Les points de M étant peu nombreux les mauvais intervalles ne peuvent être longs et la proportion recouverte par les autres est grande. Or, sur les bons intervalles, la moyenne partielle est grande ($> b - \epsilon/2$). Sur les mauvais, peu nombreux, la contribution est faible (en module) parceque d'une part f_1 est bornée, et d'autre part, les moyennes de f_2 sont petites ($< \epsilon/2$). Au total, on y arrive en sommant ces contributions. □

5.6 Caractérisations de l'ergodicité

Théorème 5.4. *Un système dynamique (X, T, μ) est ergodique ssi*
- $\forall f \in L^2_\mu, Uf = f \Rightarrow f$ *est constante.*
- *La valeur propre 1 de U est simple.*
- $\forall f \in D$, *avec* $\overline{D} = L^2_\mu$, $\frac{1}{n}\sum_{k=0}^{n-1} \int f \circ T^k.\overline{f} d\mu \to |\int f d\mu|^2$.
- $\forall f, g \in L^2_\mu$, $\frac{1}{n}\sum_{k=0}^{n-1} \int f \circ T^k.\overline{g} d\mu \to \int f d\mu. \int \overline{g} d\mu$.
- $\forall A \in \mathcal{D}$, *avec* $\sigma(\mathcal{D}) = \mathcal{A}$, $\frac{1}{n}\sum_{k=0}^{n-1} \mu(A \Delta T^{-k} A) \rightsquigarrow \mu(A)^2$.
- $\forall A, B \in \mathcal{A}$, $\frac{1}{n}\sum_{k=0}^{n-1} \mu(A \Delta T^{-k} B) \rightsquigarrow \mu(A)\mu(B)$.

5.7 Autres versions

Version L^2.

Théorème 5.5 (Théorème ergodique de Von Neumann)**.** *Soit (X, T, μ) un système dynamique ;*

$$\forall f \in L^2(\mu),\ \frac{1}{n}\sum_{k=0}^{n-1} f \circ T^k \to_{n\to\infty} \mathbb{E}[f|\mathcal{I}],\ \text{dans } L^2.$$

Démonstration. La convergence vers 0 est immédiate pour les cocycles. Mais l'ensemble des cocycles est dense dans l'orthogonal de $L^2(\mathcal{I})$ (Lemme 5.1) □

La *Loi forte des grands nombres* est une conséquence immédiate du théorème ergodique. La prouver à partir du théorème, à titre d'exercice.

Il n'est pas évident de dire quoi que ce soit sur la vitesse de la convergence ; en général, on peut choisir les fonctions f de manière à rendre cette convergence arbitrairement lente. Eventuellement on peut dire qqchose en se restreignant à certaines classes de fonctions.

Et aussi ergodicité sur une sous-tribu ; ergodicité de l'induite sur un ensemble de mesure positive (réciproque : trouver un contre-exemple).

5.8 Décomposition ergodique

Sous conditions topo (X compact metrisable, T continue), les probas T-invariantes ergodiques sont les points extrémaux de l'ensemble des probas T-invariantes. (qui est lui-même un compact). (Proposition 4.12 p. 34 de [PB]). Idée de preuves à la main (mais incomplètes) dans [K] vers la page 23. A base d'espérance conditionelle sachant l'orbite. Preuve du fait que μ_x est ergodique dans [D], Theoremes ergodiques, page 10 (et aussi dans [L]).

5.9 Récurrence multiple

Furstenberg a montré que les deux résultats suivants sont équivalents.

Théorème 5.6 (Szemeredi). *Pour tout $\delta > 0$ et tout k, il existe un entier N tel que dès que $E \subset [1, N]$ est tel que $|E|/N \geq \delta$, E contienne une progression arithmétique de longueur k.*

Théorème 5.7 (Récurence multiple de Furstenberg). *Si (X, T, μ) est un système dynamique et A un ensemble mesurable $\mu(A) > 0$, alors pour tout k entier, il existe n entier tel que*

$$\mu(A \cap T^n A \cap \cdots \cap T^{(k-1)n} A) > 0.$$

Plus généralement on peut s'intéresser à des quantités comme

$$\frac{1}{n} \sum_{k=0}^{n-1} f(T^k x) g(T^{2k} x).$$

6 Melange

6.1 Présentation

L'ergodicité dit que partant de (presque) n'importe où, le système va essentiellement partout ; en tous cas, l'espace ne peut pas se décomposer en deux parties (de mesures positives) qui seraient "disjointes". Cependant, par nature, ce qui se passe dans le futur est determiné par le passé, dans une certaine mesure ; je veux dire : si on connait exactement le point de départ, on connait toute l'orbite. Mais si on n'a qu'une information partielle sur le point de départ, qu'elle information cela nous donne sur le comportement à venir ? Cette connaissance peut-elle avoir une influence asymptotiquement. C'est à ce type de question que s'intéresse la propriété de *mélange*. La situation la "moins" mélangeante serait celle d'un système réduit à une orbite périodique. La plus mélangeante celle d'une suite de variables aléatoires indépendantes.

Définition 6.1 (Mélange, l'idée). Les propriétés de mélange d'un système dynamique (X, T, μ) concernent les modes de convergence des suites $\mu(T^{-n} B \cap C)$ vers $\mu(B)\mu(C)$ pour des boreliens $B, C \in \mathcal{B}$.

L'ergodicité garantit la convergence en moyenne. Cette covergence peut avoir lieu sur un ensemble de densité 1 (mélange faible), partout (mélange ou mélange fort), voire de manière relativement uniforme (mélange uniforme ou propriété K).

Exemples. Interprétations. Hiérarchie.

Du point de vue technique, on peut, comme d'habitude, travailler avec les boréliens ou avec des fonctions dans L^2_μ..

6.2 Théorie spectrale

Je ne veux pas rentrer dans les subtilités de la théorie spectrale, mais quand même essayer de donner une idée du fait que les propriétés de mélange se lisent sur le spectre de l'opérateur unitaire Ude L^2.

6.2.1 Théorème spectral

Soit (X, T, μ) un système dynamique ; on considère l'opérateurunitaire U de $L^2(\mu)$ tel que $Uf(x) = f \circ T(x)$.

Théorème 6.1. *Pour toute fonction $f \in L^2(\mu)$, il existe une mesure σ_{ff} sur $[0,1)$ et une isométrie W : vect($U^n f; n \in \mathbb{Z}$) $\to L^2([0,1), \mathcal{B}, \sigma_{ff})$telle que $W(f) = 1$, $W(Ug) = M_{e^{2i\pi x}}.W(g)$, où $M_{\lambda(x)}g(x) = \lambda(x)g(x)$.*

La propriété d'isométrie s'écrit :

$$\forall n \in \mathbb{N}, \int_X U^n f.\overline{f} d\mu = \int_0^1 e^{2i\pi nx} d\sigma_{ff}$$

On déduit de ce théorème qu'il existe une famille (denombrable) de mesures $(\sigma_i)_{i\in\mathbb{N}}$telle que $(L^2(\mu), U)$ soit unitairement équivalent à $(\oplus L^2(\sigma_i), M_{e^{2i\pi x}})$. La classe de la mesure $\sum 2^{-i}\sigma_i$ est appelée *type spectral maximal* du système. (+multiplicité, ...)

Remarque. Une fonction f est invariante ssi $\sigma_{ff} = a\delta_0$. Le système est ergodique ssi pour toute fonction f d'intégrale nulle, $\sigma_{ff}(\{0\}) = 0$. Couplé avec le théorème ergodique, ce résultat fournit une preuve de la caractérisation de l'ergodicité en terme des corrélations (fait avant ou pas ?).

6.2.2 Spectre discret

On dit qu'un système est à spectre discret si $L^2(X, \mathcal{A}, \mu) = \oplus \mathfrak{e}_n$, où $U\mathfrak{e}_n = \lambda_n \mathfrak{e}_n$, i.e. s'il existe une base de fonctions propres.

Un système à spectre discret est isomorphe à un produit de rotations sur un groupe compact (voir le dual topologique). En tous cas les rotations sur des groupes compacts sont à spectre discret. C'est grosso modo le point de départ de l'analyse de Fourrier.

6.3 Mélange faible

Définition 6.2. Un système dynamique (X, T, μ) a la propriété de *mélange faible* si,

$$\forall B, C \in \mathcal{B},\ \frac{1}{n}\sum_{i=0}^{n-1} |\mu(T^{-i}B \cap C) - \mu(B)\mu(C)| \longrightarrow_{n\to\infty} 0.$$

L'ergodicité et le mélange faible ont une interprétation spectrale sur l'opérateur $U : L^2(\mu) \to L^2(\mu), f \mapsto f \circ T$. L'ergodicité correspond à la simplicité de la valeur propre 1, i.e. toutes les fonctions propres sont constantes. Le mélange faible à l'absence de spectre discret, i.e. pas d'autres valeurs propres que 1, et à la simplicité de cette dernière. L'interprétation spectrale de la propriété de mélange fort (voir plus loin) est moins directe.

Caractérisations Le théorème suivant fournit diverses caractérisations du mélange faible. Avant de l'énoncer, observons que la convergence *sur un ensemble de densité* 1 a les propriétés suivantes :

$$a_n \rightsquigarrow a \Leftrightarrow \frac{1}{n}\sum_{i=0}^{n-1}|a_i - a| \to 0 \Leftrightarrow \exists \mathfrak{N} \subset \mathbb{N}, \frac{\sharp\mathfrak{N} \cap \{1, \ldots, n\}}{n} \longrightarrow_{n\to\infty} 1, \text{et}, a_n \longrightarrow_{\substack{n\to\infty \\ n\in\mathfrak{N}}} a.$$

Dit encore autrement : $\exists \varphi : \mathbb{N} \nearrow \mathbb{N}, \frac{\sharp\{i\in\mathbb{N}, \varphi(i)\leq n\}}{n} \longrightarrow_{n\to\infty} 1, \text{et}, a_{\varphi(n)} \longrightarrow_{n\to\infty} a$

Théorème 6.2. *Un système dynamique (X, T, μ) a la propriété de mélange faible ssi*

- $\forall f \in L^2_\mu, \forall \lambda \in \mathbb{U}, Uf = \lambda f \Rightarrow \lambda = 1$ *et f est constante.*
- $\forall f \in L^2_\mu$, *non constante, la mesure σ_{ff} est diffuse.*
- $\forall f \in D$, *avec $\overline{D} = L^2_\mu, \int f \circ T^i.\overline{f} d\mu \rightsquigarrow |\int f d\mu|^2$.*
- $\forall f, g \in L^2_\mu, \int f \circ T^i.\overline{g} d\mu \rightsquigarrow \int f d\mu. \int \overline{g} d\mu$.

- $\forall A \in \mathcal{D}$, *avec* $\sigma(\mathcal{D}) = \mathcal{A}$, $\mu(A \Delta T^{-n} A) \rightsquigarrow \mu(A)^2$.
- $\forall A, B \in \mathcal{A}$, $\mu(A \Delta T^{-n} B) \rightsquigarrow \mu(A)\mu(B)$.
- $\forall (Y, S, \nu)$ *ergodique,* $(X \times Y, T \otimes S, \mu \otimes \nu)$ *est ergodique .*
- *Le système dynamique* $(X \times X, T \otimes T, \mu \otimes \mu)$ *est ergodique.*

6.4 Mélange fort

Définition 6.3. On dit que le système a la propriété de *mélange* (fort) si

$$\forall B, C \in \mathcal{B},\ \mu(T^{-n}B \cap C) \longrightarrow_{n\to\infty} \mu(B)\mu(C).$$

Caractérisations : proposer et démontrer.
Exemples : SFT, Chaînes de Markov.

Exercices 6.4. *Trouver une exemple de système dynamique satisfaisant la propriété de mélange faible mais pas celle de mélange fort.*

6.5 Mélange uniforme

Définition 6.5. Il a la propriété de *mélange uniforme*, dite aussi *propriété K de Kolmogorov* si pour tout $B \in \mathcal{B}$ et toute partition $\mathcal{P}$,

$$\sup_{C \in \bigvee_{k=n}^{\infty} T^k \mathcal{P}} |\mu(B \cap C) - \mu(B)\mu(C)| \longrightarrow_{n\to\infty} 0$$

6.6 Bernoulli

Définition 6.6. On dit que le système est *Bernoulli* si il est isomorphe à un shift muni de la mesure produit.

Notons les implications,

Théorème 6.3.

Bernoulli $\Rightarrow$ *K* $\Rightarrow$ *mélange fort* $\Rightarrow$ *mélange faible* $\Rightarrow$ *ergodicité.*

Discussion. Exemple de Kalikow. Autres propriétés spectrales ?

6.7 Vitesse de mélange

Ces propriétés donnent des renseignements qualitatifs sur l'indépendance asymptotique du système qui ne permettent pas d'obtenir de théorèmes limites plus précis que le théorème ergodique. Lorsque le système a la propriété de mélange, on peut se demander à quelle vitesse a lieu la décroissance des corrélations

$$C_n(f, g) = \int_X f \circ T^n g d\mu - \int_X f d\mu \int_X g d\mu.$$

Cela fournit des indications sur le comportement statistique asymptotique du système. On ne peut espérer contrôler cette vitesse que pour certaines familles d'observables. Typiquement, on peut se demander si la vitesse de la convergence est exponentielle pour des observables Hölder. Méthodes : quasi-compacité de l'opérateur de Peron-Frobenius (dual L^2 de l'opérateur de composition) et technique à base de couplages. Ce type de résultat sur les corrélations permet d'obtenir des théorèmes limites pour des sommes ergodiques $\sum_{k=0}^{n-1} f \circ T^n$.

6.8 Vers des théorèmes limites

On peut espérer, pour certains systèmes dynamiques ET pour certaines fonctions f obtenir des convergences en loi du type, sous μ,

$$\frac{1}{\sqrt{n}}\left(\sum_{k=0}^{n-1} f\circ T^k - n\int_X f d\mu\right) \stackrel{\text{(loi)}}{\to} N(0,\sigma^2).$$

Un ingrédient crucial pour obtenir ce type de résultats est d'avoir un contrôle de la vitesse de décroissance des corrélations (genre sommable) ; cela permet d'utiliser des résultats connus de TLC pour des martingales dont la norme L^2 est contrôlée par les corrélations. Une difficulté classique est de s'assurer que la variance n'est pas nulle (cas par exemple où f est un cobord)

On peut aussi bien sûr espérer montrer d'autres théorèmes limites ; par exemple des principes de grandes deviations, des loi limites de temps d'entrées dans des suites d'événements asymptotiquement rares, ... Le point central est de s'assurer que les corrélations décroissent suffisamment vite pour que le comportement asymptotique des v.a. considérées soient essentiellement le même que si les v.a. étaient indépendantes.

7 Entropie

7.1 Entropie et information

Entropie et complexité. Point de vue compression et théorie de l'information. Point de vue thermodynamique.

7.2 Entropie d'une partition

Soit $(X,\mathcal{A},\mu)$ un espace mesuré. Une partition $\mathcal{P}$ mesurable de X est une famille (finie) de parties mesurables non vide de X formant une partition.

Information d'une partition :

$$I_{\mathcal{P}}(x) = \sum_{A\in\mathcal{P}} (-\log\mu(A))1_A(x).$$

L'entropie de $\mathcal{P}$ est la moyenne de l'information :

$$H_\mu(\mathcal{P}) = \int_X I_{\mathcal{P}}(x)d\mu(x) = -\sum_{A\in\mathcal{P}} \mu(A)\log\mu(A).$$

Notons $\varphi(t) = -t\log t$.

On note $\mathcal{P}\vee\mathcal{Q}$ le sup de deux partitions, i.e. la partitions formée des intersections des atomes : $\mathcal{P}\vee\mathcal{Q} = \{A\cap B; A\in\mathcal{P}, B\in\mathcal{Q}, A\cap B\neq\varnothing\}$. Associatif. Relation d'ordre $\mathcal{P}\preccurlyeq\mathcal{Q}$ pour dire que $\mathcal{Q}$ est plus fine que $\mathcal{P}$ si tout élément de $\mathcal{P}$ est réunion d'éléments de $\mathcal{Q}$. Observons que $\mathcal{P}\preccurlyeq\mathcal{Q}$ ssi $\mathcal{P}\vee\mathcal{Q} = \mathcal{Q}$.

L'information relative de $\mathcal{P}$ par rapport à $\mathcal{Q}$ est la fonction :

$$I_{\mathcal{P}|\mathcal{Q}}(x) = \sum_{A\in\mathcal{P},B\in\mathcal{Q}} -\log\left(\frac{\mu(A\cap B)}{\mu(B)}\right)1_{A\cap B}(x)$$

L'entropie relative $H_\mu(\mathcal{P}|\mathcal{Q})$ est son integrale $= \sum_{A\in\mathcal{P},B\in\mathcal{Q}} \varphi\left(\frac{\mu(A\cap B)}{\mu(B)}\right)\mu(B)$. L'information est positive. Son integrale aussi. Observons que $I_{\mathcal{P}}(x) = I_{\mathcal{P}|\{X\}}(x)$. Aussi $H(\mathcal{P})\leq\log|\mathcal{P}|$avec égalité seulement pour la mesure uniforme. Cela vient de la concavité stricte de φ. Esssentiellement : $1/p\sum_A \varphi(\mu(A)) \le \varphi(1/p\sum_A \mu(A))$. Enfin controle par les -log des plus gros et plus petits atomes.

Additivité : $H_\mu(\mathcal{P}\vee\mathcal{Q}|\mathcal{R}) = H_\mu(\mathcal{P}|\mathcal{Q}\vee\mathcal{R}) + H_\mu(\mathcal{Q}|\mathcal{R})$

[Facile sur l'information ; puis integrer].

En prenant $\mathcal{R} = \{X\}$, on déduit : $H_\mu(\mathcal{P}\vee\mathcal{Q}) = H_\mu(\mathcal{P}|\mathcal{Q}) + H_\mu(\mathcal{Q})$

Monotonies : si $\mathcal{P} \preccurlyeq \mathcal{Q}$, alors $H_\mu(\mathcal{P}|\mathcal{R}) \le H_\mu(\mathcal{Q}|\mathcal{R})$, et, si $\mathcal{R} \preccurlyeq \mathcal{Q}$, alors $H_\mu(\mathcal{P}|\mathcal{Q}) \le H_\mu(\mathcal{P}|\mathcal{R})$

Pour une preuve voir p87.

On déduit : si $\mathcal{P} \preccurlyeq \mathcal{Q}$, alors $H_\mu(\mathcal{P}) \le H_\mu(\mathcal{Q})$. mais aussi $H_\mu(\mathcal{P}|\mathcal{Q}) \le H_\mu(\mathcal{Q})$ et $H_\mu(\mathcal{P}) - H(\mathcal{Q}) \le H_\mu(\mathcal{P}|\mathcal{Q})$.

Sous additivité : $H_\mu(\mathcal{P}\vee\mathcal{Q}|\mathcal{R}) \le H_\mu(\mathcal{P}|\mathcal{R}) + H_\mu(\mathcal{Q}|\mathcal{R})$ et $H_\mu(\mathcal{P}\vee\mathcal{Q}) \le H_\mu(\mathcal{P}) + H_\mu(\mathcal{Q})$.

Comprendre que $H(P|\mathcal{Q}_n) \to 0$ si $(\mathcal{Q}_n)$ est une suite de partitions de plus en plus fines dont la réunion engendre la tribu $\mathcal{A}$. Assez fin ; voir p88 et 89.

7.3 Entropie d'une transformation

Définition et définitions équivalentes. Continuité / partitions. Sup sur les partitions. Partitions génératrice.

Soit μ une mesure T-invariante. On définit l'entropie d'une partition $\mathcal{P} = (P_1, \ldots, P_n)$ mesurable de X et celle de la transformation T pour cette partition par,

$$H_\mu(\mathcal{P}) = -\sum_{i=1}^{n} \mu(P_i)\log\mu(P_i), \text{et } H_\mu(T,\mathcal{P}) = \lim_{n\to\infty} \frac{1}{n} H_\mu(\bigvee_{i=0}^{n-1} T^{-i}\mathcal{P}).$$

On a aussi

$$H_\mu(T,\mathcal{P}) = \lim_{n\to\infty} H_\mu(\mathcal{P}|\bigvee_{i=1}^{n} T^{-i}\mathcal{P}).$$

Si on note $\mathcal{P}_+ = \bigvee_{i=1}^{\infty} T^{-i}\mathcal{P}$ et $\mathcal{P}_- = \bigvee_{i=1}^{\infty} T^{-i}\mathcal{P}$, on remarque que $H_\mu(T,\mathcal{P}) = 0$ est équivalent à dire que $\mathcal{P}$ est $\mathcal{P}_-$-mesurable (tout aussi bien $\mathcal{P}_+$-mesurable, d'ailleurs). En outre, $H_\mu(T,\mathcal{P}) = H_\mu(\mathcal{P})$ équivaut à dire que les partitions $T^{-i}\mathcal{P}$ sont indépendantes.

Définition 7.1 (Entropie d'un système dynamique)**.** L'entropie *métrique* $h_\mu(T)$ du système dynamique (X,T,μ) est obtenue en maximisant $H_\mu(T,\mathcal{P})$ sur l'ensemble des partitions d'entropie finie.

Ce maximum est atteint par $H_\mu(T,\mathcal{P})$ si la partition $\mathcal{P}$ est génératrice. (Rappel : existence d'un générateur).

L'entropie est un invariant pour l'isomorphisme.

Exercices 7.2. *Autres propriétés : sous additivité, continuité par rapport aux partitions, remplacement de $\mathcal{P}$ par $\bigvee_{i=0}^{n-1} T^{-i}\mathcal{P}$. Iteration : $h_\mu(T^k) = kh_\mu(T)$. Systeme produit. Entropie de l'inverse dans le cas bijectif.*

Exemple 7.1. *Autour des Bernoulli. Isomorphisme. Et aussi Markov ? Entropie d'une rotation (irrationnelle). Entropie nulle pour les isométries par morceaux (Buzzi) : ennoncé précis. Entropie des automorphismes du tore. Entropie d'une application dilatante de l'intervalle.*

7.4 Entropie topologique

On peut aussi définir l'*entropie topologique* d'un système (X, T). Pour un recouvrement $\mathcal{U}$ de X par des ouverts, notons $H(\mathcal{U})$ le logarithme du cardinal du plus petit sous-recouvrement de $\mathcal{U}$ recouvrant X. L'*entropie topologique* $H(T)$ du système dynamique est obtenue en maximisant, sur l'ensemble des recouvrements par un nombre fini d'ouverts, les limites

$$H(T,\mathcal{U}) = \lim_{n\to\infty} \frac{1}{n} H(\bigvee_{i=0}^{n-1} T^{-i}\mathcal{U}).$$

Cesdeuxnotionssontreliesparunprincipevariationnel, stipulant(voirWalters, [**?**])que

$$H(T) = \sup\{H_\mu(T) : \mu \text{ mesure } T - \text{invariante}\}.$$

7.5 Shannon-McMillan

Ennoncé (et preuve) du théorème. Interprétation.

Théorème 7.1. *Soit (X, T, μ) un système dynamique et $\mathcal{P} = (P_n)_{n\in I}$ une partition génératrice.*

$$\frac{1}{n}\log(\mu\left(\bigcap_{k=0}^{n-1} P_k(x)\right)) \longrightarrow_{n\to\infty} H(T), \mu - ps.$$

7.6 Décalages de Bernoulli

L'entropie d'un shift de Bernoulli de paramètres $(p_i)_{i\in A}$ est $h = \sum_{i\in A} p_i \log p_i$. Deux shifts de Bernoulli isomorphes ont évidement la même entropie. Mais, en fait, et c'est nettement plus subtil, deux shifts de Bernoulli sont isomorphes ssi ils ont la même entropie. Ornstein.

7.7 Pinsker, K-systèmes etc...

Définition 7.3. La famille $\Pi(T) = \{A \in \mathcal{A} : H_\mu(T, \{A, A^c\}) = 0\}$ est une tribu appellée *tribu de Pinsker* de (X, T, μ).

Théorème 7.2. *Le système dynamique (X, T, μ) est un K-système si $\Pi(T)$ est grossière.*

C'est-à-dire si pour toute partition $\mathcal{P}$ non triviale d'entropie finie, $H_\mu(T, \mathcal{P}) > 0$. Lorsque la tribu de Pinsker n'est pas triviale, elle permet de definir le plus grand facteur d'entropie nulle du système.

8 Exercices

Exercices 8.1. *Nombre de 7 comme premier chiffre du développement décimal de* 2^N*. A relier à la rotation d'angle* $\log_{10} 2$*.*

Exercices 8.2. *Trouver (chercher) des orbites periodiques dans un billard triangulaire.*

Exercices 8.3. *Montrer que le doublement de l'angle est isomorphe au full shift unilatère. Montrer que la transformation du boulanger est isomorphe au full shift bilatère. Verifier qu'une mesure invariante reste une mesure invariante.*

Exercices 8.4. *Entropie d'un SFT.*

Exercices 8.5. *On considère l'ensemble des mots infinis bilatéraaux ne contenant que des blocs de* 1 *de longueur paire. Montrer que c'est un sous-shift. Montrer que ce n'est pas un sous-shift de type fini.*

Exercices 8.6. *Lebesgue invariante pour une rotation, un échange d'intervalle. Mesure de Haar invariante pour les translations. Mesure invariante* μ *(absolument continue par rapport à Lebesgue) pour le* $\beta - shift$ $x \mapsto \beta x (mod 1)$*, pour* $\beta^2 = \beta + 1$*.*

Exercices 8.7. *Soit* (X, T, μ) *un système dynamique et* A *une partie mesurable de* X*. On note* T_A *l'application induite de* T *sur* A*. Montrer que* $\mu_{A?} \frac{\mu_{|A}}{\mu(A)}$ *est* T_A*-invariante.*

Exercices 8.8. *On reprend le* β*-shift sous* μ*. Montrer qu'il est conjué à un SFT muni d'une mesure* $\tilde{\mu}$*. Montrer qu'il satisfait la propriété de mélange fort.*

Exercices 8.9. *Construire un point* x *du shift pour lequel les moyennes ergodiques ne convergent pas.*

Exercices 8.10. *Montrer que pour une rotation irrationnelle, une partition du cercle en deux intervalles (d'interieur non vide) est toujours génératrice.*

Exercices 8.11. *Montrer que* (X, T, μ) *est ergodique si et seulement si pour tous* A, B *mesurables avec* $\mu(A) > 0, \mu(B) > 0$*, il existe* $n > 0$ *tel que* $\mu(T^{-n} A \cap B) > 0$*.*

Exercices 8.12. *Soit* μ *et* ν *deux mesures invariantes ergodiques pour* (X, T)*. Montrer qu'il existe* A *mesurable tel que* $\mu(A) = 1$ *et* $\nu(A) = 0$ *(*μ *et* ν *sont étrangères).*

Exercices 8.13. *Soit* (X, T, μ) *un système dynamique avec* X *metrique compact,* T *continue et* μ *de support* X *(le plus petit fermé de mesure* 1 *est* X*). Montrer que, dans ce cadre, le mélange entraine le mélange topologique.*

Exercices 8.14. *Calculer l'entropie d'une chaine de Markov.*

Exercices 8.15. *Soit* (X, T, μ) *un système dynamique et* A *une partie mesurable de* X*. On note* T_A *l'application induite de* T *sur* A*. Calculer* $h_{\mu_A}(T_A)$*.*

Exercices 8.16. *Pour un système* (X, T, μ) *inversible, montrer que* $h_\mu(T) = h_\mu(T^{-1})$*.*

9 Chaînes de Markov, Mesure de Gibbs, chaînes d'ordre infini

Chaînes de Markov On dit que le processus discret $(X_n)_{n\in\mathbb{N}}$ est *markovien* si

$$\mathbb{P}(X_{n+1} = b \mid X_0 = a_0, \dots, X_{n-1} = a_{n-1}, X_n = a_n) = \mathbb{P}(X_{n+1} = b \mid X_n = a).$$

Soit $P = (p_{i,j})_{i,j\in A}$ une matrice à entrées positives. C'est une matrice *stochastique* si pour tout $i \in A$, $\sum_{a\in A} p_{i,a} = 1$.

Définition 9.1 (Chaîne de Markov)**.** Étant donnée une matrice stochastique P et une mesure ν sur A, on appelle chaîne de Markov de transition P et de mesure initiale ν le processus markovien $(X_n^\nu)_{n\geq 0}$ tel que $\mathbb{P}(X_0^\nu = a) = \nu(a)$ et

$$\mathbb{P}(X_{n+1}^\nu = b \mid X_n^\nu = a) = p_{a,b}.$$

À une matrice de transition on peut associer le graphe de sommets A dont les arêtes sont les transitions de probabilités strictement positives. On peut ainsi transposer naturellement les notions d'irréductibilité, de période et d'apériodicité pour des chaînes de Markov. Une chaîne irréductible admet une unique mesure invariante (i.e. une mesure initiale telle que la chaîne soit un processus stationnaire). C'est un vecteur propre (associé à la valeur propre 1) à gauche de la matrice de transition. L'apériodicité garantit l'absence d'autres valeurs propres de module 1 de cette matrice. On ne spécifiera pas la mesure initiale, lorsque c'est l'unique mesure invariante.

La notion de chaîne de Markov s'étend facilement au cas d'un alphabet dénombrable. La matrice est remplacée par un tableau infini. Soit a un état d'une chaîne irréductible sur un alphabet infini. Considérons le temps de retour à l'état a (partant de a). La chaîne est dite récurrente si ce temps est fini p.s. et transiente sinon. Elle est dite récurrente positive si l'espérance de ce temps est finie et récurrente nulle sinon. Une chaîne irréductible récurrente admet une unique mesure de probabilité invariante. Pour plus de détails, voir Chung [**?**].

On peut aussi définir des chaînes de Markov non *homogènes* en remplaçant la matrice de transition par une suite de matrices de transitions. La notion de mesure invariante n'a alors plus de sens, mais on peut définir le processus pour chaque mesure initiale et, le cas échéant s'intéresser à ses propriétés de mélange.

Modèle d'Ising et états d'équilibre On note A l'ensemble $\{-1, 1\}$ et X_N l'ensemble $\{-1, 1\}^N$, appelé dans ce contexte espace des configurations. On choisit sur $\{0, \dots, N\}$ une famille de voisinages i.e. un ensemble de parties $\mathcal{V}$ comprenant les singletons. À chaque voisinage $V \in \mathcal{V}$ on associe une fonction H_V (énergie d'interaction) de $\{-1, 1\}^V$ dans $\mathbb{R}$. L'énergie d'une configuration est alors définie comme $H(x) = \sum_{V\in\mathcal{V}} H_V((x_i)_{i\in V})$. On appelle fonction de partition la fonction $Z(\beta) = \sum_{x\in X_N} e^{-\beta H(x)}$. Pour toute "température" T, réel > 0, on définit une mesure de probabilité μ_T sur X_N par

$$\mu_T(\{x\}) = \frac{e^{-\frac{1}{T}H(x)}}{Z(\frac{1}{T})}.$$

Cette construction est la même si on remplace $\{0, \dots, N\}$ par une partie finie Λ de $\mathbb{Z}^d$, pour $d \geq 1$ entier. L'un des objets de la mécanique statistique est de comprendre comment ce type de mesures s'étend à des mesures sur $\{-1, 1\}^{\mathbb{Z}^d}$ tout entier. On parle de passage à la *limite thermodynamique*.

Soyons plus précis. On se donne une famille de voisinage sur $\mathbb{Z}^d$ et d'énergies d'interaction, qu'on suppose invariants par translation. On ne peut à priori pas définir l'énergie d'une configuration, mais on peut (peut-être) définir la différence entre les énergies de deux configurations distinctes seulement en un nombre fini de points. Ainsi si on fixe une configuration y de référence (condition au bord) on peut, pour toute “boite” finie Λ, effectuer la construction ci dessus en remplaçant $H(x)$ par $\Delta H(x,y) = H(x) - H(y)$ pour obtenir une mesure sur A^Λ. Pour $x_\Lambda \in A^\Lambda$,

$$\mu_T^{\Lambda,y}(\{x_\Lambda\}) = \frac{e^{-\frac{1}{T}(\Delta H(x,y))}}{Z^{\Lambda,y}(\frac{1}{T})}$$

où x est la configuration égale à x_Λ dans Λ et à y en dehors. On interprète cette mesure comme mesure conditionnelle dans Λ, sachant la configuration hors de Λ fixée égale à y. Une problématique importante, est de déterminer s'il existe une (seule) mesure sur $A^{\mathbb{Z}^d}$ invariante par translation dont les conditionnelles sont exactement ces mesures, alors appelée *mesure de Gibbs*, ou *mesure d'équilibre*. Cela revient essentiellement à dire que, pour une boite assez grande, la mesure obtenue ne dépend pas de la condition au bord choisie. En dimension $d = 1$ et si les voisinages sont de tailles uniformément bornées, il n'existe qu'une telle mesure. En revanche, en dimensions supérieures, y compris pour des systèmes de voisinages bornés, on peut avoir, sous une température critique, plusieurs mesures d'équilibre distinctes ; on parle de *transitions de phases*. En dimension 1, ce phénomène ne peut se produire qu'en présence d'interactions à longue portée, i.e. de voisinages de tailles non bornées, avec des interactions ne décroissant pas trop vite.

Pour exploiter les idées de la mécanique statistique dans le cadre qui nous intéressera, on formule traditionellement les choses un peu différemment. C'est le cadre du formalisme thermodynamique proposé par Ruelle [**?**] pour décrire certains systèmes dynamiques. On considère le shift $\mathcal{X}$ (on pourrait travailler aussi avec un sous-shift de type fini). C'est-à-dire que l'espace est $\mathbb{N}$ et les états 0 ou 1. On note $x \overset{m}{=} y$ si $x_j = y_j$ pour tous $0 \leq j \leq m$. Soit $\phi : \mathcal{X} \to \mathbb{R}$. On définit l'oscillation de ϕ sur les cylindres,

$$var_n(\phi) = sup_{x \overset{n}{=} y} |\phi(x) - \phi(y)|.$$

Supposons $\sum_n var_n(\phi) < \infty$. On appelle *pression* de ϕ et on note $P(\phi)$ la quantité,

$$P(\phi) = \lim_{n\to\infty} \frac{1}{n} \log \sum_{a_1,\ldots,a_n \in A^n} \exp\left(\sup_{x \in c(n,a_1\cdots a_n)} \sum_{k=0}^{n-1} \phi(T^k x) \right).$$

Remarquons que la condition de sommabilité des oscillations garantit que l'on pourrait remplacer le supremum par n'importe quelle valeur prise dans le cylindre sans changer $P(\phi)$.

Définition 9.2 (Mesure d'équilibre)**.** On appelle mesure d'équilibre associée à un potentiel ϕ satisfaisant $\sum_n var_n(\phi) < \infty$ l'unique mesure de probabilité μ_ϕ sur $\mathcal{X}$ satisfaisant, pour une constante C,

$$\frac{1}{C} \leq \frac{\mu_\phi(c(n, a_1 \cdots a_n))}{e^{\sum_{k=1}^n \phi(a_k \cdots a_n x) - nP(\phi)}} \leq C.$$

La mesure μ_ϕ satisfait un principe variationnel, au sens où elle réalise le maximum

$$P(\phi) = \sup \left\{ H_\mu(T) + \int_X \phi d\mu : \mu \text{ mesure } T - \text{invariante} \right\}.$$

La mesure d'entropie maximale est la mesure d'équilibre associée au potentiel uniformément nul. Pour plus de détails, renvoyons à Walters [**?**] et Bowen [**?**]. Notons qu'on peut voir ϕ comme la contribution du site 0 à l'énergie du système, $\phi(x) \simeq \sum_{\substack{V \in \mathcal{V} \\ 0 \in V}} H_V(x)$. La condition $\sum_n var_n(\phi) < \infty$ implique que l'interaction décroît suffisamment vite quand les voisinages grandissent et permet d'éviter les transitions de phases.

Chaînes d'ordre infini Soit P une famille de probabilités de transitions sur $A \times A^{-\mathbb{N}}$, c'est-à-dire une famille de probabilités sur A dépendant mesurablement d'un paramètre dans $A^{-\mathbb{N}}$,

$$\begin{array}{rccc} P: & A \times A^{-\mathbb{N}} & \longrightarrow & [0;1] \\ & (a,z) & \longmapsto & P(a|z) \quad . \end{array} \tag{6}$$

La régularité de cette probabilité de transition est décrite par la suite

$$\gamma_m = 1 - \inf_{a \in A, x, y \in A^{-\mathbb{N}} : x \overset{m}{=} y} \frac{P(a|x)}{P(a|y)}, \tag{7}$$

où $x \overset{m}{=} y$ signifie ici $x_j = y_j$ pour tous $-m \leq j \leq 0$.

Définition 9.3 (Chaîne d'ordre infini)**.** Nous dirons qu'un processus stationnaire $(X_n)_{n\in\mathbb{Z}}$ à valeurs dans A est une chaîne d'ordre infini de transition P si

- pour tous $a_1, \ldots, a_n \in A$,
$$\mathbb{P}(X_1 = a_1, \ldots, X_n = a_n) > 0 \tag{8}$$
- la limite
$$\lim_{m\to\infty} \mathbb{P}(X_0 = a | X_j = x_j, -m \leq j \leq -1) = \mathbb{P}(X_0 = a | X_j = x_j, j \leq -1) =: P(a|x) \tag{9}$$
existe pour tout $a \in A$ et tout $x \in A^{-\mathbb{N}}$,
- la suite $(\gamma_m)_{m\geq 1}$ décrivant la régularité de P satisfait $\lim_{m\to\infty} \gamma_m \quad = \quad 0$,

Dans la littérature, ces processus sont parfois appelés aussi chaînes à liaisons complètes. Étant donné un passé $x \in A^{-\mathbb{N}}$, la *chaîne de passé x et de transitions P*, est le processus $(Z_n^x)_{n\in\mathbb{Z}}$ dont les probabilités conditionnelles satisfont

$$P(Z_n^x = a \mid Z_{n+j}^x = z_j, j \leq -1) \quad = \quad P(a \mid z) \text{pour } n \geq 0 \quad , \tag{10}$$

pour tout $a \in A$ et tout passé z avec $z_{j-n} = x_j, j \leq -1$, et tel que $Z_n^x = x_n \quad , \quad$ pour $n \leq -1$.

Les propriétés de décroissance de la suite (γ_m) influencent notoirement les propriétés du processus. En particulier, si (γ_m) atteint 0, on dit que le processus est une *chaîne de Markov de portée $k = min\{m > 0, \gamma_m = 0\}$*. Si (γ_m) décroît exponentiellement, on dit que la chaîne est à décroissance exponentielle, ou que la probabilité de transition est de régularité Hölder (en fait, le log est Hölder pour la métrique produit). Si $\sum_{m\geq 0} \gamma_m < \infty$, on dit que la chaîne est sommable (à décroissance sommable, ou à mémoire sommable). Enfin, on s'intéressera à la condition suivante, légèrement plus faible que la sommabilité.

$$\sum_{m\geq 1} \prod_{k=0}^{m} (1 - \gamma_k) \quad = \quad +\infty \quad . \tag{11}$$

Remarquons que cette formulation implique que les probabilités de transition sont toutes strictement positives. Il est possible de définir le même type de processus en se restreignant à des SFT.

La propriété (11) est suffisante pour garantir l'ergodicité de la mesure (ainsi que —c'est équivalent ici— l'unicité d'une mesure invariante ayant les lois conditionnelles spécifiées. En outre il est possible de calculer la vitesse de mélange en termes de la suite (γ_n).

Plus précisement :

Théorème 9.1. *Pour tous x, y dans $A^{\mathbb{N}}$ et tout $a \in A$,*

$$|P(X_n^x = a) - P(X_n^y = a)| \leq \gamma_n^*. \tag{12}$$

La suite (γ_n^*) est définie de la manière suivante. On considère la chaîne de Markov $(S_n^{(\gamma)})_{n\in\mathbb{N}}$ à valeurs dans $\mathbb{N}$, commencant à l'origine et de matrice de transition donnée par,

$$\begin{cases} p_{i,i+1} & = & 1 - \gamma_i \\ p_{i,0} & = & \gamma_i \ , \end{cases} \tag{13}$$

pour tout $i \in \mathbb{N}$. Pour tout $n \geq 1$ on pose

$$\gamma_n^* = P(S_n^{(\gamma)} = 0) \ . \tag{14}$$

Pour exploiter ce résultat, il faut relier les propriétés de γ^* à celles de γ. Le premier point est que sous la condition (11), la chaîne de Markov $S^{(\gamma)}$ est récurrente nulle (ou transiente), entraînant $\gamma_n^* \to 0$. La sommabilité de γ_n entraîne celle de γ_n^*. Une décroissance polynomiale pour γ_n entraîne une décroissance au pire du même ordre pour γ_n^*. Une décroissance exponentielle de γ_n entraîne une décroissance exponentielle de γ_n^*, quoiqu'éventuellement avec un taux moindre.

Ergodicité Une conséquence importante (déjà remarquée par Berbee [?]) est que la condition la plus faible (11) entraîne l'unicité de la mesure compatible avec ce système de probabilités conditionnelles. Cette problématique est directement reliée à celle de l'existence de transition de phases pour des mesures de Gibbs avec des potentiels à longues interactions. Dans le cadre des g-mesures, mentionnons le travail de Bramson et Kalikow [?] qui construisent un exemple de fonction g avec 2 mesures invariantes distinctes, celui de Quas [?] qui adapte cette construction au cadre des applications de l'intervalle, celui de Quas et Coelho [?] sur le couplage et la distance $\overline{d}$ dans ce contexte et celui de Hulse [?] qui, en imposant une condition de monotonicité sur les transitions, obtient l'unicité sous une condition de régularité plus faible.

Décroissance des corrélations Ce résultat sur la vitesse de relaxation se traduit directement en un résultat comparable concernant la vitesse de décroissance des corrélations pour le processus stationnaire. Ce résultat se transpose naturellement au cadre du formalisme thermodynamique, permettant d'estimer la vitesse de décroissance des corrélations pour des mesures d'équilibre associées à des potentiels de régularité faible.

Remarque. Pour ces systèmes, les mesures sur les shifts que l'on obtient ont une propriété très forte exprimée sur les probabilités conditionnelles, qui, entraîne des propriétés de mélange (plus ou moins rapide). Il n'est en général pas possible d'aller dans l'autre sens. On peut avoir du mélange, même rapide, sans avoir ce type de propriété de continuité des probabilités conditionnelles.

10 Substitutions

Définition Soit A un ensemble fini.

Définition 10.1. Une substitution σ est une application de A dans $A^* \setminus \{\epsilon\}$.

Elle s'étend à A^* en un morphisme de monoïde par $\sigma(WW') = \sigma(W)\sigma(W')$. Elle s'étend à $A^{\mathbb{N}}$ (et à $A^{\mathbb{Z}}$) par continuité. (C'est (en général) une contraction).

Exemple 10.1. *Quelques exemples classiques :*
- *Morse : $a \to ab, b \to ba$.*

- *Fibonacci :* $a \to ab, b \to a$.
- *Tribonacci :* $a \to ab, b \to ac, c \to a$.
- *Mirror Tribonacci :* $a \to ba, b \to ca, c \to a$.
- *Flip Tribonacci :* $a \to ab, b \to ca, c \to a$.

Points fixes Un point fixe est un mot $u \in A^{\mathbb{N}}$ tel que $\sigma(u) = u$. Point périodique. Si $|\sigma^n(a)| \underset{n\to\infty}{\longrightarrow} \infty$ pour $a \in A$, alors σ a un point périodique. Si en outre il existe $a \in A$ et $W \in A^*$ tels que $\sigma(a) = aW$, alors $\sigma^n(a)$ converge vers un point fixe. Quitte à itérer, on peut ne parler que de points fixes.

10.1 Système Dynamique Substitutif

Sous-shift Une substitution σ est *primitive* s'il existe un entier n tel que pour toutes lettres a et b de A, b apparaisse dans $\sigma^n(a)$. Une substitution primitive admet un point périodique. Elle peut en avoir plusieurs ; cependant : l'adhérence Ω_σ de l'orbite d'un point fixe d'une substitution primitive σ ne dépend pas du point périodique choisi.

Remarque. Attention : les points σ-periodiques peuvent être T-périodiques. A proscrire.

Proposition 10.1. *Si σ est primitive, (Ω_σ, T) est minimal.*

Démonstration. L'argument repose sur la primitivité. Tout mot qui W apparait dans un mot infini de Ω_σ apparait dans un point periodique et donc dans un $\sigma^n(a)$ pour n assez grand. Mais pour tout n, il existe N tel que $\sigma^n(a)$ apparait dans tous les $\sigma^N(b)$ pour tout b. Ainsi, si on appelle L_N la longueur maximale de $\sigma^N(b)$, notre mot W apparait dans tout mot de longueur $2L_N$ (qui contient necessairement une occurence de $\sigma^N(b)$ pour un b. □

Complexité Pour mémoire :

Théorème 10.2 (Pansiot, 70's)**.** *La complexité d'un point fixe de substitution est de l'un des types suivants :*
- $O(1)$ *(ex : $a \to a$),*
- $O(n)$ *(ex : $a \to ab, b \to a$),*
- $O(n \log\log n)$ *(ex : $a \to abc, b \to ac \to cdd \to c$),*
- $O(n \log n)$ *(ex : $a \to aabc, b \to a, c \to cd, d \to c$),*
- $O(n^2)$ *(ex : $a \to abc, b \to bc, c \to c$).*

Mais dans le cas primitif non périodique, la complexité est linéaire. Voir Exercice 10.3 pour l'étude d'un cas particulier.

10.2 Abelianisation

Matrice d'une substitution On définit la matrice $M_\sigma = (m^\sigma_{a,b})_{a,b\in A}$ de la substitution σ par $m^\sigma_{a,b} = |\sigma(b)|_a$ nombre de a dans $\sigma(b)$. Cela permet de définir un morphisme e de A^* dans $\mathbb{Z}^{|A|}$, par $W \mapsto e_W = e(W) = (|W|_a)_{a\in\mathcal{A}}$. On a $e_{\sigma(a)} = M_\sigma e_a$ soit donc,

$$\begin{array}{ccc} A^* & \overset{\sigma}{\to} & A^* \\ \downarrow e & & \downarrow e \\ \mathbb{Z}^{|A|} & \overset{M_\sigma}{\to} & \mathbb{Z}^{|A|} \end{array}$$

Unique ergodicité On observe que M_σ est une matrice entière positive. La substitution σ est *primitive* ssi il existe n, $M_\sigma^n > 0$. Ainsi, le théorème de Perron-Frobenius garantit l'existence d'une valeur propre λ de module maximal associée à un vecteur propre (normalisé) à coordonnées positives $\rho = (\rho_a)_{a\in\mathcal{A}}$, $\sum_a \rho_a = 1$. la valeur propre λ peut-être interprétée comme taux de croissance exponentielle de $|\sigma^n(a)|$; pour tous $a \in A$, $\frac{1}{n}\log|\sigma^n(a)| \underset{n\to\infty}{\longrightarrow} \lambda$. On en déduit que la fréquence de chaque lettre $b \in \mathcal{A}$ dans u existe et vaut ρ_b. Schématiquement, l'idée est que le nombre de b dans $\sigma^n(a)$est donné par $(M_\sigma^n)_{a,b}$. Donc pour le vecteur de fréquences

$$(|\sigma^n(b)|_a)_{a\in\mathcal{A}} = \frac{M_\sigma^n e_b}{||M_\sigma^n e_b||_1} \underset{n\to\infty}{\longrightarrow} \rho.$$

On en conclu (après un peu de travail) que le système (Ω_σ, T) est uniquement ergodique.

10.3 Automates

Automate des prefixes-suffixes Soit σ une substitution primitive. On construit un graphe étiqueté $G_\sigma = (V, L, E)$ où $E \subset V \times V \times L$ sont des arêtes étiquetées par L :

- Sommets : $V = A$
- Etiquettes : $L = A^* \times A^*$, (un préfixe et un suffixe).
- Arête étiquetée $a \xrightarrow{p,s} b$ ssi $\sigma(b) = pas$.

Automate des prefixes-suffixes Pour tout préfixe d'un point fixe u, il existe un chemin fini dans G_σ étiqueté par des prefixes $(p_i)_{i=1\cdots N}$ tels que

$$u_1\cdots u_n = \sigma^N(p_N)\sigma^{N-1}(p_{N-1})\cdots\sigma(p_1)p_0.$$

Soit $(s_N)_{N\geq 0}$ la suite d'étiquettes d'un chemin infini dans G_σ ; alors

$$x = s_0\sigma(s_1)\sigma^2(s_2)\cdots\sigma^N(s_N)\cdots$$

est dans Ω_σ sauf si $s_N = \epsilon$ à partir d'un certain rang N_0. Dans ce dernier cas, le mot $x = s_0\sigma(s_1)\sigma^2(s_2)\cdots\sigma^{N_0}(s_{N_0-1})u$ est une préimage d'un point fixe. Essentiellement on a construit une bijection entre les chemins infinis dans le graphe G_σ et Ω_σ.

10.4 Fractal de Rauzy

Conditions sur la matrice. La substitution σ est *unimodulaire* si $det(M_\sigma) = 1$. Soit P_σ le polynôme caractéristique de M_σ. Un nombre algébrique est un nombre de Pisot si il est de module plus grand que 1 et tous ses conjugués sont de module strictement plus petit que 1. La substitution σ est *Pisot* si P_σ est irréductible et sa valeur propre principale est un nombre de Pisot. (Type Pisot).

Remarque. σ Pisot $\Rightarrow$ σ Primitive.

Ligne brisée On considère maintenant $\mathbb{R}^{|\mathcal{A}|}$ muni de la base canonique $\{e_a, a \in \mathcal{A}\}$. A un point u de $\mathcal{A}^{\mathbb{Z}^+}$, on associe une *ligne brisée* L^u de sommets

$$L_n^u = \sum_{k=0}^{n-1} e_{u_k}.$$

Si $u \in \Omega_\sigma$, L_n^u tend vers l'infini dans la direction principale ρ. Si u est un point fixe de σ on peut utiliser la décomposition en préfixes pour écrire

$$L_n^u = \sum_{k=1}^{N} e_{\sigma^k(p_k)} = \sum_{k=1}^{N} M_\sigma^k e_{p_k}.$$

Fractal de Rauzy Supposons M_σ *unimodulaire* et *Pisot.* La ligne brisée L_n^u reste à une distance bornée de la direction principale. La fermeture des projections de L_n^u pour $n \geq 0$ parallelement à la direction principale (par exemple sur l'hyperplan contractant) est un compact $\mathcal{R}$ — appelé *fractal de Rauzy.* Le système dynamique (Ω_σ, T) est (presque) conjugué à un échange de morceaux sur $\mathcal{R}$.

Exercices 10.2. *Reconnaissablitié. Lien avec l'induction. Autres sommes ergodiques pour des fonctions dépendant d'un nombre fini de coordonnées.*

Exercices 10.3. *Mot de Fibonacci. Soit $A = \{a, b\}$ et $\varphi : A \to A^*$, $a \mapsto ab$, $b \mapsto a$.*

1 *Montrer que $\varphi^{n+1}(a) = \varphi^n(a)\varphi^{n-1}(a)$. En déduire que $|\varphi^n(a)| = F_{n+2}$ où (F_n) est la suite de Fibonacci, définie par recurence par , $F_0 = 0$, $F_1 = 1$ et $F_{n+1} = F_n + F_{n-1}$.*

2 *Montrer que* $\lim_{n\to+\infty} \phi(a)$ *existe dans $A^{\mathbb{N}}$. On notera $\underline{f}$ ce mot infini. Determiner $F_n(\underline{f})$ pour $n \leq 5$. Tracer la pertie correspondante de l'arbre des facteurs à droite. Qu'observe-t-on sur $p(n)$?*

3 *Montrer que $\underline{f}$ est uniforment recurent.*

4 *Montrer que $F(\underline{f})$ est stable par image miroir.*

5 *Montrer que $\underline{f}$ n'est pas ultimement periodique.*

6 *Montrer que tout prefixe de $\underline{f}$ est special a gauche.*

7 *Trouver un mot w tel que $w \not\in F(\underline{f})$ mais $\varphi(w) \in F(\underline{f})$.*

8 *Demontrer le lemme de décomposition : tout facteur $w \in F(\underline{f})$ s'ecrit de maniere unique sous la forme $w = r\phi(v)s$ avec $v \in F(\underline{f})$, $r \in \{\epsilon, b\}$, $s = a$ si w finit par a et $s = \epsilon$ sinon.*

9 *Facteurs spéciaux. Montrer que si w est un facteur special à gauche, non vide, il existe un unique facteur spécial à gauche v tel que $w = \varphi(v)s$ avec $s = a$ si w finit par a et $s = \epsilon$ sinon. En déduire la forme générale des facteurs spéciaux à gauche de $\underline{f}$, puis celle des facteurs spéciaux a droite.*

10 *Calculer la fonction de complexité de $\underline{f}$.*

11 Automorphismes du Tore

Le tore $\mathbb{T}^n = \mathbb{R}^n/\mathbb{Z}^n \cong [0, 1)^n$. M une matrice carrée $n \times n$ à coefficients entiers agit sur $\mathbb{R}^n$; l'action passe au quotient. $T_M : x = (x_1, \ldots, x_n) \longmapsto Mx = \left(\sum_{j=1,\ldots n} m_{i,j}x_j(\text{mod}1)\right)_{i=1,\ldots,n}$.

T_M est une surjection de $\mathbb{T}^n$ sur lui-même ssi $\det(M) \neq 0$. Si —$\det(M)| = 1$, c'est une bijection. Dans les 2 cas, T_M préserve la mesure de Lebesgue. Exemples.

On trouvera une justification du résultat suivant dans $[D]$:

Théorème 11.1. *Soit M une matrice carrée $n \times n$ à coefficients entiers de determinant ± 1 et T_M l'automorphisme du tore associé. Les conditions suivantes sont équivalentes : M n'a pas de valeurs propres racine de l'unité $\Leftrightarrow$ T_M est ergodique $\Leftrightarrow$ T_M est mélangeante ($\Leftrightarrow$ T_M a spectre de Lebesgue).*

L'entropie de T_M est donnée par $h_\lambda(T_M) = \sum_{i=1}^n \log^+ |\lambda_i|$. Pour faire la minoration, on découpe en petits cubes de coté ϵ paralleles à la base de Jordan, avec ϵ assez petit pour que $T^{-1}P \cap P$ soit dans un parallelepipède (genre $\epsilon|\sup \lambda_i|$¡1). Maintenant on regarde les $P_{i_1} \cap T^{-1}P_{i_2} \cap \cdots \cap T^{-k}P_{i_k}$ et on evalue leur mesure (a un polynome près). On minore en utilisant la forme de l'entropie. Pour une majoration, on évalue le nombre d'atomes rencontrés par $T^{-m}P_i$ et on déduit une majoration de $H(\mathcal{P}|T^{-m}\mathcal{P})$ (comme somme sur les atomes de $T^{-m}\mathcal{P}$ de $m(P)H_\star(P)$) qui assure une majoration de $H(T^m)$et donc de $H(T)$. et on utilise l'autre definition de l'entropie. Bon, pour dire l'idée, le mieux est probablement de regarder ce qui se passe dans une base vraiment propre. Le faire.

Bon, on découpe comme préconisé en ϵ-cubes. Quand on itère on voit des lammelles de hauteur λ_+^{-k} (petite) et de largeur λ_+^{-k} (grande) qui traversent nos cubes. La mesure de leur intersection avec les cubes est $\epsilon^{d-} \times \lambda_+^k$. Essentiellement cette mesure est la même partout. Donc l'integration ne coute pas cher et le ϵ^{d-} disparait avec la limite $\frac{1}{n}\log$. Ce raisonnement suffit pour la minoration.

Pour aller plus loin, on pourrait définir les difféomorphismes Anosov, voire les systèmes hyperboliques. Et évoquer l'existence de partitions de Markov.

12 Hyperbolicité (Heuristique)

Hyperbolicité non uniforme, faible, relative, partielle, quasi-hyperbolicité, semi, presque, etc... les qualificatifs sont variés lorsqu'on s'intéresse à des systèmes qui pour certaines raisons ne sont pas loin d'être hyperboliques, mais de fait, ne le sont pas. Le plus souvent, on s'inspire de ce qu'on sait faire dans le cas hyperbolique, par exemple pour montrer que finalement pour cette notion affaiblie, les conclusions restent les mêmes. Il arrive aussi qu'aux frontières du cadre hyperbolique le paysage change plus radicalement. Mais commençons par dire ce qui est *vraiment* hyperbolique.

Point de vue *géométrique* D'un point de vue géométrique l'hyperbolicité évoque la courbure négative. Dans le triptyque des géométries — sphérique, euclidienne, hyperbolique — le modèle standard pour la dernière est $\mathbb{H}$, le plan hyperbolique : demi plan de Poincaré ou disque de Poincaré muni de la métrique hyperbolique. Les surfaces de genre supérieur ou égal à deux peuvent être munies d'une métrique qui les rend hyperboliques. C'est une propriété locale. L'idée est que localement l'espace tangent se décompose en deux sous-espaces (disons deux directions), formant point selle.

Mais, cette propriété "locale" influence les propriétés "globales" des espaces hyperboliques. Il est possible de formaliser ces propriétés plus globales en s'affranchissant de la notion de courbure (Gromov, [**?**] ; Ghys, [**?**]). Par exemple, dans un espace hyperbolique non compact, on doit pouvoir partir vers l'infini de multiples façons. De ce point de vue, un arbre est un espace discret dont les propriétés globales rappellent celle de $\mathbb{H}$ (pensons à un plongement du graphe de Cayley du groupe libre F_2 dans $\mathbb{H}$).

La notion d'espace (métrique) hyperbolique introduite par Gromov permet de rendre compte de ces propriétés globales. Un espace métrique (géodésique) (H, d) est dit *δ-hyperbolique* si pour tout triangle (A, B, C), chaque coté est inclut dans un δ voisinage de la réunion des deux autres côtés. Il est *hyperbolique* s'il existe δ tel qu'il soit δ-hyperbolique. Cette notion est invariante par quasi-isométrie : deux espaces métriques (X, d) et (Y, d') sont quasi-isométriques s'il existe deux applications $h_1 : X \to Y$ et $h_2 : Y \to X$ ultimement lipschitziennes —$d(h(x), h(y)) \leq d(x, y) + \text{cte}$— dont les composées $h_1 \circ h_2$ et $h_2 \circ h_1$ sont bornées. Un arbre est 0-hyperbolique.

Cette notion s'applique en particulier à des espaces "discrets". En particulier, on peut munir un graphe d'une métrique en donnant longueur 1 à chaque arête et dès lors se demander s'il est ou pas hyperbolique. Dans le même esprit, on peut se demander si un groupe discret est hyperbolique, en se demandant si son graphe de Cayley (pour un système fini de générateurs) l'est. Comme la notion est invariante par quasi-isométrie, cela se trouve finalement être une propriété du groupe. $\mathbb{Z}$ est hyperbolique, mais est un cas un peu à part. Les groupes libres F_n et les groupes fondamentaux de surfaces fermées de genre supérieur ou égal à 2 sont hyperboliques. $\mathbb{Z}^d$ et $\mathbb{R}^d$, pour $d \geq 2$, ne sont pas hyperboliques. B_n, le groupe des tresses à n brins n'est pas hyperbolique. Pour qu'un groupe puisse être hyperbolique, il ne faut pas qu'il ait "trop" de relations.

Ces considérations sont importantes en particulier lorsqu'on s'intéresse au "bord" d'un groupe engendré par un nombre fini de générateurs. Mentionnons le fait que les groupes hyperboliques entre dans le cadre des groupes "automatiques", pour lesquels il existe une forme normale régulière (voir Epstein et al, [?]). Par ailleurs, le caractère hyperbolique permet de définir systématiquement une notion de bord topologique. Pour les groupes hyperboliques, cet objet coïncide avec sa frontière de Poisson (Ancona [?]).

Point de vue *matrices et cônes* Un système est hyperbolique si quelque part il y a un cône qui est envoyé strictement dans lui même, assez vite (et assez uniformément). Un cadre naturel pour comprendre ce point de vue est celui des matrices positives, incluant les matrices à entrées binaires comme matrices d'incidences de graphes, les matrices à entrées entières comme matrices de substitutions, les matrices stochastiques, les matrices jacobiennes d'applications dilatantes.

Le théorème de Perron-Frobenius est un résultat d'algèbre élémentaire qui assure que toute matrice positive M, primitive, admet une unique valeur propre de module maximal r, que cette valeur propre est réelle et simple, et que le vecteur propre associé a toutes ses coordonnées strictement positives. La simplicité de la valeur propre et la stricte positivité des coordonnées proviennent de l'irréductibilité, l'unicité (parmi les valeurs propres de module maximal) est garantie par l'apériodicité. Une matrice irréductible peut avoir un nombre fini de valeurs propres sur le cercle de rayon r.

Ce résultat permet de calculer l'entropie d'un sous-shift de type fini (taux de croissance du nombre de chemins dans le graphe associé). Il permet de calculer la fréquence des mots dans des points fixes de substitutions. Il permet d'assurer l'existence (et l'unicité) de mesures invariantes pour des chaînes de Markov. Il garantit aussi un taux de convergence exponentiel de la mesure, partant d'une mesure initiale quelconque, vers la mesure d'équilibre (vitesse de relaxation).

Le point crucial est que le cône positif est envoyé strictement dans lui même. Toutes les directions sont contractées sur la direction principale. La vitesse de cette contraction est donnée par le ratio —*trou spectral*— entre la valeur propre principale et le module de la "seconde" valeur propre.

Cette idée garde toute sa pertinence si l'on regarde non pas les puissances d'une matrice mais un produit de matrices distinctes — provenant par exemple d'une chaîne de Markov non stationnaire, de produit de matrices jacobiennes le long d'une orbite pour calculer des exposants de Lyapunov (théorème d'Oseledec) ou, plus généralement, un cocycle au dessus d'un système dynamique.

Le cadre hyperbolique est celui dans lequel il est possible d'utiliser directement cette idée clef pour l'analyse du système.

Point de vue *dynamique* La notion d'hyperbolicité des systèmes dynamiques s'est révélée fructueuse. L'analogie peut paraître formelle, mais le nom est au moins est justifié par le fait que le flot géodésique sur le plan hyperbolique est hyperbolique.

Essentiellement, un difféomorphisme T régulier d'une varieté riemannienne $\mathcal{M}$ est hyperbolique s'il admet un sous-ensemble Λ, sur lequel il existe une décomposition invariante de l'espace tangent $T_xM = E^s(x) \bigoplus E^u(x)$ telle que D_xT^n soit uniformément dilatant dans la "direction" instable $E^u(x)$ et uniformément contractant dans la direction stable $E^s(x)$. Cette définition se généralise aux flots (en utilisant une section). Le flot géodésique sur le plan hyperbolique est, heureusement, hyperbolique ; c'est plus généralement le cas des flots géodésiques sur des surfaces munies de métriques à courbure négative.

Ces systèmes géométriques ont un codage particulièrement simple. Une manière de faire le lien est de réaliser que la transformation du boulanger (Exemple 2.14) —essentiellement hyperbolique— est isomorphe au décalage bilatère (comme le 2-shift (Exemple 2.7) est isomorphe au décalage unilatère). Mais la connection est en fait beaucoup plus forte, comme le montre l'existence de partitions de Markov pour les systèmes hyperboliques. Ainsi, en effet, pour tout système hyperbolique, il existe une partition qui conjugue le système à un sous-shift de type fini. Cela fait du décalage l'archétype du sytème hyperbolique. Remarquons qu'on retrouve les propriétés de dilatation et de contraction dans le fait que le décalage (y compris restreint à un SFT) est dilatant vers le futur et vers le passé. Plus précisement, le décalage sur le shift unilatère est une application localement dilatante. Pour le décalage sur le shift bilatère $A^{\mathbb{Z}}$, le rôle des variétés stables est joué par les suites coïncidant à partir d'un certain rang. La distance entre deux suites suffisament proches qui ne coïncident pas à partir d'un certain rang est uniformément dilatée. On peut faire la même analyse pour son inverse. Il reste à bien comprendre la manière dont la mesure naturelle (sur la variété) est transportée sur le SFT. Ce sont en fait des mesures d'équilibres dont les potentiels (qui décrivent comment la mesure conforme est transformée par le décalage) sont reliés à la différentielle du difféomorphisme (qui décrit comment le volume est transformé localement par le difféomorphisme).

Mentionnons une autre passerelle entre la géométrie des espaces hyperboliques et la dynamique symbolique, liée à la définition de dynamiques sur le bord de groupes d'isométries de ces espaces, et à l'étude de leurs propriétés pour différentes mesures, que nous ne développerons pas. Citons simplement Bowen et Series [**?**].

Point de vue *statistique* La philosophie est que les systèmes hyperboliques, si on ne regarde pas de trop près, se comportent d'un point de vue statistique pour des points choisis au hasard (volume, mesure naturelle) comme des systèmes (très) aléatoires, i.e., si on prend une fonction f assez régulière (observable), la suite $(f \circ T^n)$ ressemble à une suite de va iid.

Cette idée est directement liee à l'idée que si l'on choisit deux points très proches dans le système (mais pas dans la même varieté stable) leurs orbites vont s'éloigner rapidement (exponentiellement), jusqu'à ce qu'on "oublie" en les regardant évoluer qu'ils ont un jour été proches. Cette indépendance asymptotique se lit concrètement dans la vitesse de mélange du système. Plus généralement, dans le cadre hyperbolique, on regroupe les points dont le futur sera essentiellement le même (variétés stables) et ceux qui au contraire s'éloigneront (et donc se rapprochent si on inverse le temps).

Pas d'hyperbolicité On sort du cadre hyperbolique dès que dans une région de l'espace les propriétés requises ne sont pas satisfaites (présence d'un point fixe (ou de directions) neutres ou d'une singularité de la différentielle).

Mais la situation extrême de système sans hyperbolicité est celle des isométries, comme, notamment, les rotations ou les échanges d'intervalles qui correspondent à des codages de flots sur des surfaces "plates", i.e. munies de métriques euclidiennes. Nous verrons que l'hyperbolicité contribue aussi à l'étude de ces dynamiques, par le biais de la renormalisation.

Références

[1] Bowen, Rufus, *Equilibrium states and the ergodic theory of Anosov diffeomorphisms*, Lecture Notes in Mathematics **470**, Springer-Verlag, Berlin, réédition (2008).

[2] Benoist, Yves. et Paulin, Frédéric.. *Systèmes dynamiques élémentaires*, Notes de cours. http ://www.dma.ens.fr/edition/NotesCours/index.html. (2002)

[3] Coudene, Yves, Systèmes dynamiques et théorie ergodique, Cours de DEA. http ://www.math.univ-brest.fr/perso/yves.coudene/dea-cours-v2.pdf

[4] Delarue, Thierry, *Théorie Ergodique*, Notes de cours. http ://www.univ-rouen.fr/LMRS/Persopage/Delarue/te.html. (2003).

[5] Kalikow, Steven, *Outline of Ergodic Theory*, Cambridge Studies in Advanced Mathematics (No. 122) (2010). (et aussi : http ://www.math.unc.edu/Faculty/petersen/erg3.doc)

[6] Petersen, Karl, *Ergodic theory*, Cambridge Studies in Advanced Mathematics, Cambridge University Press (1989).

[7] Sinai, Ya. G., *Introduction to ergodic theory*, Mathematical Notes **18**, Princeton University Press, Princeton, N.J. (1976).

[8] Walters, Peter, *An introduction to ergodic theory*, Graduate Texts in Mathematics **79**, Springer-Verlag, New York (1982).