Permutations aléatoires et discrétisations d'applications mélangeantes

Alexis Gilles Jean Feydy sous la direction de Serge Cantat

Résumé

Nous nous intéressons ici aux discrétisations d'applications mélangeantes sur le tore, et tentons de comprendre si ces applications sont statistiquement distinguables des applications typiques de $\{1, \dots, n\}$ dans lui même, que l'on notera $\mathrm{End}(n)$. Nous porterons une attention toute particulière à un problème similaire plus simple à implémenter sur ordinateur : la réduction modulo un nombre premier p de polynômes de $\mathbb{Z}[X]$.

Table des matières

1 Introduction

Intéressons nous à à la dynamique sur le long terme d'un système (E, d), où E est un espace et $T : E \to E$ un déplacement de ses éléments, et plus précisément à l'allure d'une trajectoire $(x, T(x), T^2(x), \dots)$ "typique".

Ces trajectoires parcourront-elles tout l'espace, ou bien resteront-elle concentrées dans certaines zones, tournant en rond ? Tout aussi important de nos jours : étant donné un espace et un déplacement continu, un ordinateur sera-t-il capable d'offrir une simulation satisfaisante de la dynamique de (E, T), ou bien les erreurs de calcul dues aux approximations dyadiques des nombres réels auront-elles raison de toute prévision fiable à long terme ?

Une question plus théorique pour finir : la discrétisation (par échantillonnage discret) d'une dynamique $\mathcal{C}^1$ "quelconque" fournit-elle une dynamique discrète "quelconque", ou bien peut-on voir sur la dynamique discrète des traces de sa continuité d'antan ? Si une différence (statistiquement) perceptible existait, cela permettrait d'imaginer un critère permettant de distinguer applications aléatoires et applications résultant d'un processus de discrétisation.

Ce sont les trois questions que nous formulerons précisément, et auxquelles nous donnerons des réponses partielles dans la suite de ce mémoire.

1.1 Propriétés continues d'une dynamique ergodique

Voici un exemple typique d'application que l'on cherche à discrétiser : l'action linéaire usuelle de $\begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$ sur le tore $\mathbb{T}$ par $T : (x, y) \mapsto (y, x + y) \bmod 1$. Notons que l'action est bien définie sur le tore, car le réseau $\mathbb{Z} \times \mathbb{Z}$ est préservé par $(x, y) \mapsto (y, x + y)$.

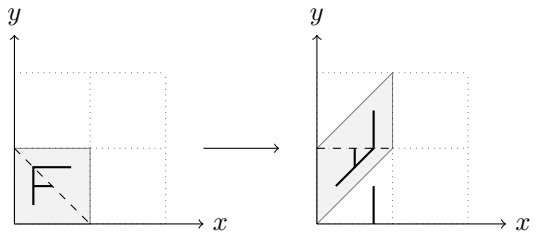

FIGURE 1 – Action de $T : (x, y) \mapsto (y, x + y) \bmod 1$ sur le tore $\mathbb{T}$.

Lemme 1. *Cette action est mélangeante pour la mesure de Lebesgue λ, de masse totale 1 sur $\mathbb{T}$:*

$$\forall A, B \subseteq \mathbb{T} \text{ boréliens }, \lambda(A \cap T^k(B)) \longrightarrow \lambda(A)\lambda(B)$$

Cette propriété, très forte, a de nombreuses conséquences sur la dynamique du système. En voici quelques unes.

1.1.1 Comportement des orbites

Tout d'abord, notons que le comportement d'une orbite est très sensible à de petits changements des coordonnées initiales : avec probabilité 1, une orbite est dense, mais il y a aussi des orbites périodiques. Et de plus, les orbites periodiques sont denses dans les orbites, et de même pour les orbites denses. Voici trois théorèmes qui devraient nous aider à mieux comprendre l'allure des orbites sous l'action de T.

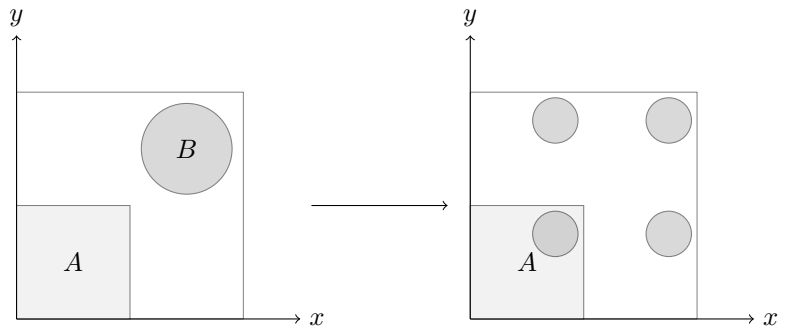

FIGURE 2 – La propriété de mélange sur le tore $\mathbb{T}$.

Les théorèmes qui suivent ne s'appliquent pas qu'à la matrice $\begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$ sur le tore, mais aussi par exemple à $z \in S^1 \mapsto z^2 \in S^1$ où S^1 désigne le cercle, qui correspond à $\theta \mapsto 2\theta$.

Le Théorème de récurrence de Poincaré Le célèbre théorème de récurrence de Poincaré s'applique ici : en disant qu'un point est récurrent s'il est point d'accumulation de son orbite sous T, on obtient que presque tout point est récurrent.

Théorème 1 (Théorème de récurrence de Poincaré)**.** *Si T est un homéomorphisme d'un espace mesurable de mesure finie (X, μ) tel que T préserve la mesure μ, alors l'ensemble des points du tore récurrents pour T est de mesure pleine. Si de plus X est métrique et compact, cet ensemble est aussi dense.*

Presque tout point a donc une orbite qui repasse arbitrairement près de lui une infinité de fois.

Le théorème de Birkhoff La propriété de λ-mélange peut s'affaiblir en une propriété plus générale, la λ-ergodicité : on dit que d est μ-ergodique si toute partie invariante par d est de mesure nulle ou pleine.

Le théorème fondamental suivant est alors vérifié : on peut identifier moyenne spatiale et moyenne temporelle.

Théorème 2 (Théorème ergodique de Birkhoff)**.** *Soit $(X, \mathcal{B}, \mu)$ un espace de probabilité et $T : X \to X$ mesurable μ-ergodique qui préserve la mesure : pour tout $E \in \mathcal{B}, \mu(T^{-1}E) = \mu(E)$. Si $f \in L^1(X, \mathcal{B}, \mu)$, alors pour μ presque tout $x \in E$:*

$$\frac{1}{n} \sum_{k=0}^{n-1} f \circ T^k(x) \longrightarrow \int f \mathrm{d}\mu$$

Démonstration. La preuve est donnée en annexe dans le cas bijectif, au théorème 12. □

Le théorème de Kac Le théorème suivant, qui relie moyennes temporelles et spatiales, reste dans la veine du théorème ergodique de Birkhoff. Il rejoint aussi le théorème de récurrence de Poincaré car il donne un ordre de grandeur du temps de retour d'un point dans un de ses voisinages (temps presque sûrement fini) :

Théorème 3 (Théorème de Kac)**.** *Si T est un homéomorphisme d'un espace métrique compact de mesure finie (X, μ) et que T préserve μ et est μ-ergodique, si A est μ-mesurable de mesure non nulle, alors, en notant pour $a \in A$, $u(a)$ le temps de retour dans A, on a*

$$\frac{1}{\mu(A)} \iint u d\mu = \frac{\mu(\mathbb{T})}{\mu(A)}$$

Démonstration. La preuve, simple, est donnée en annexe. □

La valeur moyenne du temps de retour dans une partie A est donc inversement proportionnelle à sa taille.

1.1.2 Questions subsidiaires

La stabilité structurelle Avant d'aller plus loin, on peut s'interroger sur la généralité de la propriété de mélange, qui est peut-être le symptôme d'une dynamique très particulière. Le théorème de stablité structurelle d'Anosov vient nous rassurer sur ce point.

Théorème 4 (Anosov)**.** *Dès que la matrice $A \in GL_2(\mathbb{Z})$ n'a pas de valeurs propres de module 1 (et c'est bien le cas de $\begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$), alors si B est un difféomorphisme de $\mathbb{T}$ assez C^1-proche de A, il existe H homéomorphisme tel que $B = H^{-1} \circ A \circ H$.*

Un difféomorphisme proche de T aura donc une dynamique semblable.

Quelle mesure naturelle ? On a vu, que notre application particulière préserve la mesure de Lebesgue. Le choix de cette mesure n'est pas restrictif : en effet, si l'on se donne une application f continue qui ne préserve pas la mesure de Lebesgue mais qui possède par exemple une orbite dense $(f^k(x))_k$, alors la suite de mesures $(\frac{1}{n}\sum_{k=0}^{n-1} \delta_{f^k(x)})_k$ (avec δ_y dirac en y) est bornée donc on peut extraire une mesure invariante par f, qui de plus sera sans atome et finie.

1.2 Simulations numériques et discrétisation

Des questions se posent maintenant : toutes ces belles propriétés chaotiques se retrouveront-elles sur l'application discrétisé ? L'application discrétisée est-elle quelconque en tant qu'élément de $\mathrm{End}(n)$?

Nous y répondrons dans la suite du mémoire, mais, pour continuer, nous devons nous entendre sur ce que nous appelons une "discrétisation".

Définition 1. *Soit (X, d) un espace métrique et $T : X \to X$ une application. Une discrétisation du système dynamique (X, T), avec une précision $\varepsilon > 0$, consiste en la donnée d'un ensemble fini de points $E_\varepsilon \subset X$ et d'une projection $P_\varepsilon : M \to E_\varepsilon$ telle que pour tout $x \in X$, $d(P_\varepsilon(x), x) \leqslant \varepsilon$. On impose de plus qu'il existe une constante $C > 0$ telle que $\min\{d(x, y) : x, y \in E_\varepsilon\} > C\varepsilon$ pour que l'ensemble E_ε ne soit pas "trop dense" dans X par rapport à la précision ε.*

Si on se donne une telle discrétisation, on peut définir une application discrète $f_\varepsilon = P_\varepsilon \circ f : E_\varepsilon \to E_\varepsilon$ appelée discrétisation de f.

L'étude des discrétisations est importante du point de vue des simulations numériques. Remarquons tout de même que la manière dont un ordinateur discrétise est particulière et ne correspond pas nécessairement aux discrétisations que nous considérerons.

Le Lemme de Poursuite La dynamique discrète est-elle un tant soit peu fidèle à la dynamique continue ? Cela revient à se demander si un ordinateur, qui fait des approximations successives pour calculer l'orbite d'un point, nous permet de nous rendre compte de la "vraie" dynamique continue du système. Une condition qui semble minimale est celle donnée par le "shadowing lemma", ou lemme de poursuite :

Définition 2 (Lemme de poursuite)**.** *On dit que $f : E \to E$, où E muni d'une distance d, vérifie le lemme de poursuite si :*

$$\forall \delta > 0, \exists \varepsilon > 0, \forall (x_n) \in E^{\mathbb{N}}, (\ \forall n, d(x_{n+1}, f(x_n)) < \varepsilon\ ) \Longrightarrow (\ \exists y \in E, \forall n, d(x_n, f^n(y)) < \delta\ )$$

En d'autres termes, pour tout niveau de précision voulue δ, il existe une discrétisation suffisamment fine de pas ε telle que, dès que l'on prend une pseudo trajectoire approchée à ε près, il existe une véritable trajectoire qui en est proche à δ près. Toute trajectoire calculée par ordinateur est donc δ-proche d'une vraie trajectoire.

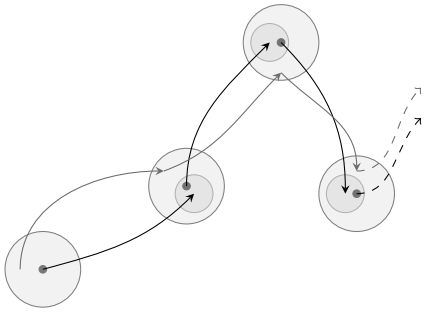

FIGURE 3 – Illustration du lemme de poursuite : les grands disques sont de rayon δ, les petits sont de rayon ε. Les points noirs sont les x_k, les flèches grises représentent l'orbite de y.

À noter que l'on parle de lemme, mais c'est plutôt une propriété, qui doit être vérifiée pour chaque difféomorphisme. Elle l'est en particulier pour $\begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$

Limites du Lemme de poursuite Dans les cas où un lemme de poursuite est vérifié, toute pseudo-trajectoire calculée par ordinateur reste proche d'une “vraie” trajectoire. Mais rien ne nous assure que cette vraie trajectoire est une trajectoire typique : l'ensemble de ces trajectoires peut-être dense tout en étant de mesure négligeable (pour une mesure adaptée). C'est la principale faiblesse du lemme de poursuite vis-à-vis du problème que nous étudions. Nous n'irons pas plus loin à ce sujet.

1.3 Un problème similaire : la réduction modulo p des polynômes

Plutôt que de discrétiser une application $d \in \mathcal{C}^1(E, E)$ sur un espace E continu, comme le tore, nous préférerions nous intéresser à un problème ne faisant intervenir que des entiers, afin d'avoir des simulations numériques exemptes d'erreurs de calculs, bien que cela reste une approximations. Nous éviterions ainsi les problèmes complexes de lemme de poursuite, etc.

Nous avons donc choisi de nous focaliser sur la réduction modulo p premier d'un polynôme $\varphi \in \mathbb{Z}[X]$: on considère l'application réduite $\bar{\varphi} : x \mapsto \varphi(x) \bmod p$, de $\mathbb{Z}/p\mathbb{Z}$ dans lui-même, et l'on étudie ses propriétés.

Obtenir des résultats pour ce problème simple de "discrétisation" (ou du moins, de réduction d'une application) sera, on l'espère, un premier pas vers des travaux plus généraux.

2 Quelques simulations numériques

2.1 Statistiques dans End(n) et $\mathfrak{S}_n$

On cherche à savoir si les applications finies obtenues par nos procédés de discrétisation et de réduction sont particulières ou non. Avant toute chose, il faut donc connaître l'allure d'un élément typique de End(n) et $\mathfrak{S}_n$, afin d'avoir un élément de comparaison : nous allons chercher des propriétés intéressantes, et faire des statistiques dessus.

On définit :

- Pour $f \in \text{End}(n)$, on constate que $[1,n] \supseteq \text{im}f \supseteq \text{im}f^2 \supseteq \dots$. Aussi, avec $R_f := \bigcap_{k\in\mathbb{N}} \text{im}f^k$, f induit une bijection de R_f dans lui-même, notée σ_f. R_f est appelée partie récurrente de f, et on s'intéressera dorénavant aux propriétés statistiques de σ_f.
- Pour $\sigma \in \mathfrak{S}_n$, on note $\mathfrak{D}(\sigma)$ sa décomposition en produit de cycles disjoints.
- Pour $\sigma \in \mathfrak{S}_n$, $U_n(\sigma) := |\mathfrak{D}(\sigma)|$ le nombres de cycles.
- Pour $\sigma \in \mathfrak{S}_n$, $L_n(\sigma)$ la longueur du plus long cycle dans $\mathfrak{D}(\sigma)$.
- Pour $\sigma \in \mathfrak{S}_n$, $\omega_n(\sigma)$ l'ordre de σ dans $(\mathfrak{S}_n, \circ)$.
- Ces notions se généralisent à End(n), via σ_f.

On note aussi $\widetilde{\omega}_n := \max_{\sigma\in\mathfrak{S}_n} \omega(\sigma)$. On munit End$(n)$, $\mathfrak{S}_n$, $[1,n]$ de la mesure de probabilité uniforme $\mathbb{P}$, $\mathbb{P}(\{f\}) = \frac{1}{\text{End}(n)}$, $\mathbb{P}(\{\sigma\}) = \frac{1}{n!}$, $\mathbb{P}(\{x\}) = \frac{1}{n}$. On a alors les statistiques suivantes, qui décrivent $\mathfrak{S}_n$.

Théorème 5. *En voyant U_n, L_n, ω_n comme des variables aléatoires sur $\mathfrak{S}_n$, on a :*

$$\frac{U_n - \log n}{\sqrt{\log n}} \xrightarrow{loi} \mathcal{N}(0,1) \tag{1}$$

$$\frac{\mathbb{E}(L_n)}{n} \underset{n\to\infty}{\downarrow} \int_0^\infty \exp(-x - \int_x^\infty \frac{e^{-y}}{y}\,dy)\,dx \tag{2}$$

$$= \quad 0.62432965\dots \tag{3}$$

$$\frac{\log\omega_n - \frac{1}{2}(\log n)^2}{\sqrt{\frac{1}{3}(\log n)^3}} \xrightarrow{loi} \mathcal{N}(0,1) \tag{4}$$

$$\frac{\log\widetilde{\omega}_n}{\sqrt{n\log n}} \longrightarrow 1 \tag{5}$$

Démonstration. Le lecteur est renvoyé à [Bol01]. On pourra toutefois noter que certaines preuves élémentaires sont données dans [CK10]. □

On s'intéresse maintenant aux applications, éléments de End(n). Pour $f \in \text{End}(n)$, f peut être vue comme un graphe orienté sur $[1,n]$. On note alors G_f ce graphe, mais non-orienté. Le boucles de G_f correspondent alors à la partie récurrente R_f de f : G_f est constitué de "rivières" (en fait, des arbres) se jetant dans des boucles, où l'action de f est cyclique.

On s'intéresse à la forme de G_f. On définit donc :

- $\alpha(f,x) := |\{f^k(x), k \in \mathbb{N}\}|$ le nombre de descendants, i.e. le cardinal de l'orbite
- $\beta(f,x) := |\{y, \exists k \in \mathbb{N}, f^k(y) = x\}|$ le nombre d'ascendants
- $\delta(f,x) :=$ |la boucle associée à x par f|
- $\gamma(f) := U_n(f)$ le nombre de cycles dans G_f
- $\overline{\gamma}(f) :=$ |éléments dans un cycle|, i.e. le cardinal de la partie récurrente
- $\gamma_m :=$ |cycles de taille m dans G_f|

On a alors les statistiques suivantes, qui décrivent End(n) :

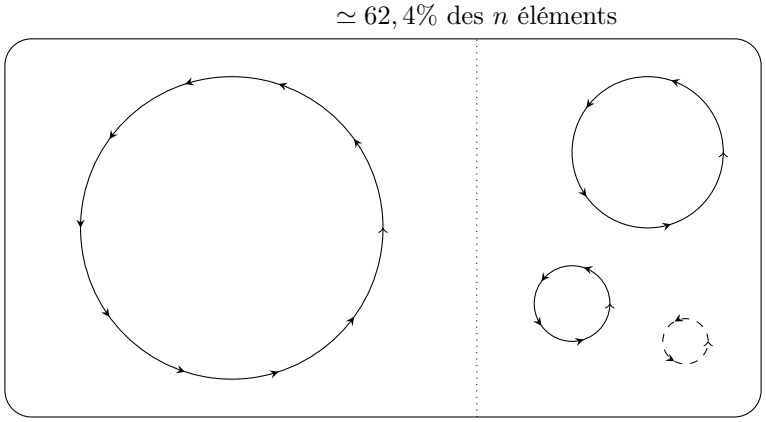

FIGURE 4 – Illustration du théorème 5 : on a environ $\log n$ cycles en moyenne.

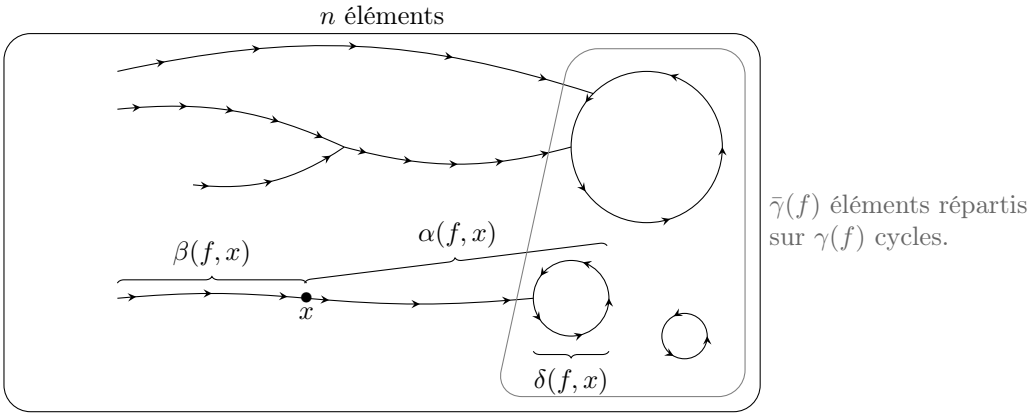

FIGURE 5 – Le graphe G_f, où $f \in \mathrm{End}(n)$.

Théorème 6. *On a, en considérant la mesure produit lorsque c'est approprié :*

$$\mathbb{P}(G_f) \text{ connexe } = \mathbb{P}(\gamma = 1) \tag{6}$$

$$\sim \sqrt{\frac{\pi}{2n}} \tag{7}$$

$$\mathbb{E}(\gamma) \sim \log n \tag{8}$$

$$\mathbb{E}(\alpha) = \mathbb{E}(\beta) = \mathbb{E}(\bar{\gamma}) \sim \sqrt{\frac{\pi n}{2}} \tag{9}$$

$$\mathbb{E}(\delta) \sim \frac{1}{4}\sqrt{2\pi n} \tag{10}$$

Démonstration. Le lecteur est renvoyé à [Bol01]. □

Ce théorème est résumé sur le DESSIN SUIVANT.

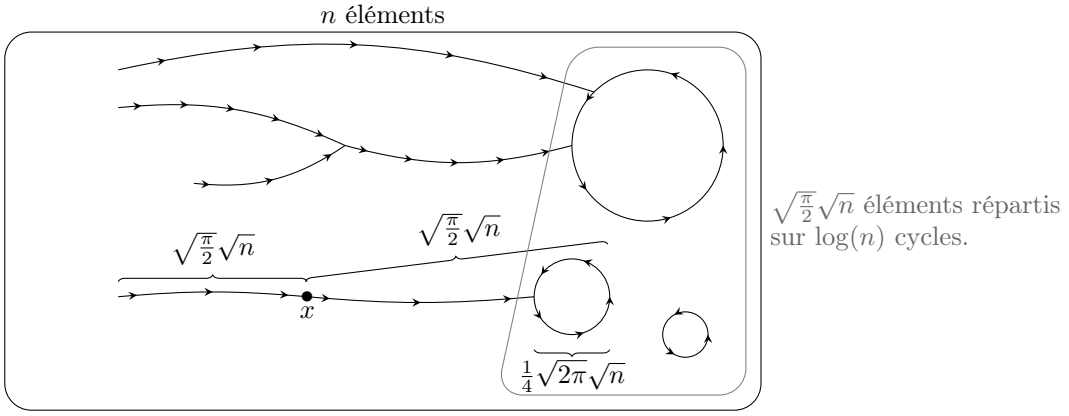

FIGURE 6 – Illustration du Théorème 6.

2.2 Simulations par discrétisations d'un problème continu

On reprend ici l'étude faite dans [LI98]. On prend l'exemple de $f : x \mapsto 2x+0.5\cdot x(1-x) \bmod 1$ sur $[0;1[$. f possède une unique mesure invariante $\mu \ll \lambda$, et (f,μ) ergodique.

On a donc :

$$\forall \varphi \text{ correcte }, \lambda - \text{p.s.}, \frac{1}{m}\sum_{n=0}^{m-1} \varphi(f^n(x)) \longrightarrow \int \varphi \, \mathrm{d}\mu$$

On obtient que presque toute orbite est dense, et uniformément répartie (pour μ).

On peut alors effectuer deux types de statistiques sur les discrétisations :

- Réduire f à une application de $\{k/10^7, k \in [|0, 10^7|]\}$ dans lui-même (i.e. à un ensemble de taille raisonnable), en arrondissant les images, et en faire une étude statistique complète.
- Travailler avec une précision plus grande, typiquement celle des "double", de l'ordre de 10^{-16}, et, au lieu de faire une étude complète (irréalisable), tirer 1000 points au hasard et déterminer l'orbite de ces points.

Mise en garde Avant d'exposer les résultats de ces expérimentations numériques, notons que ceux-ci sont extrêmement sensibles aux détails de l'implémentation choisie. Ainsi, le simple fait de changer d'architecture matérielle, et donc d'algorithme de gestion des flottants, suffit à modifier les résultats. Les données ci-dessous sont donc à prendre à titre indicatif.

Répartition des cycles On constate que les trajectoires, si elles sont finies, et donc en aucun cas denses dans $[0;1[$, sont tout de même "μ-uniformément réparties", en un certain sens. Ainsi, dans notre cas, si l'on prend un grand cycle attracteur $\mathcal{C}$ et que l'on s'intéresse à $x \mapsto \frac{|\mathcal{C}\cap[0;x[|}{|\mathcal{C}|}$, on obtient une très bonne approximation en escalier de $x \mapsto \mu([0;x[)$.

Un autre exemple Un cas où la discrétisation n'a pas donné une bonne approximation du phénomène continu a été celui de l'application $x \mapsto 1 - 2 \cdot x^2$ sur $[-1;1]$.

Les deux points fixes -1 et $1/2$ sont répulsifs. Aussi, à part les trajectoires réduites à un singleton, aucune orbite ne converge.

Il a néanmoins été observé sur un ordinateur un comportement différent : avec une discrétisation très précise (de pas $< 10^{-16}$), en choisissant 1000 points au hasard, on en a :

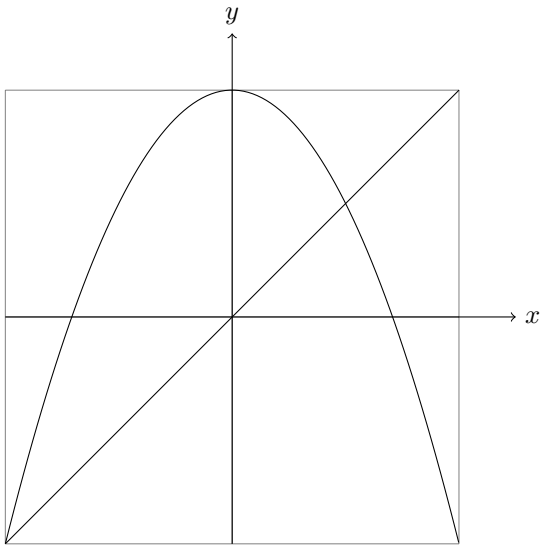

FIGURE 7 – La courbe d'équation $y = 1 - 2 \cdot x^2$. Le point fixe $1/2$ est répulsif.

- 2 qui convergent vers un cycle de taille 1 107 319
- 108 qui convergent vers un cycle de taille 3 490 273
- 890 qui convergent vers -1 ! On peut penser que toutes les trajectoires continues passant trop près de -1 ont été "happées".

Attention donc aux simulations numériques, qui pourraient laisser penser que -1 est un point fixe attracteur, alors qu'il est répulsif.

2.3 Un exemple sur $SL_2(\mathbb{Z})$

On constate que lors du passage du discret au continu, certains théorèmes ne sont plus valables. On s'intéresse ici à la discrétisation de l'action de $\begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$ sur le tore $\mathbb{T}$. On représente le tore discret par un écran composé de $M \times M$ pixels. On part d'une image représentant la tête de Poincaré, et on applique $\begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$ qui agit en tant que bijection de $\{1, \dots, M\} \times \{1, \dots M\}$. C'est bien une bijection car $\begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} \in SL_2(\mathbb{Z})$ préserve le réseau $\mathbb{Z}$ et donc le réseau $\frac{1}{M}\mathbb{Z}$. Ceci déjà est remarquable : en effet pour une application continue quelconque, la discrétisation peut faire perdre le caratère bijectif.

On sait donc que l'image doit revenir au bout d'un certains nombres d'itérations, mais de manière surprenante il suffit de 241 itérations !

Le théorème de Poincaré ne prévoit rien de tel. De plus, le théorème de Kac laisse supposer qu'un pixel retournera à sa place après environ M^2 itérations, et donc tous les pixels retrouveront leur place d'origine après un temps encore plus grand. On constate d'ailleurs par le théorème 5 qu'un retour si rapide est totalement inespéré en prenant aléatoirement une application dans $\mathfrak{S}_n$. Finalement, on voit donc que des miracles arithmétiques détaillés dans [Éti] peuvent court-circuiter les propriétés de la dynamique continue.

2.4 Simulations de réductions modulo p de polynômes à coefficients entiers

Nous nous sommes intéressés à un autre exemple de discrétisation, qui a l'avantage d'être simple à implémenter sur un ordinateur : la réduction modulo p de polynômes de $\mathbb{Z}[X]$.

Ne trouvant pas d'études statistiques suffisamment poussées dans la littérature, et étant intéressés par l'informatique en général, nous avons codé nous-mêmes un programme Caml générant les statistiques recherchées. Voici la manière dont nous avons procédé.

Tout d'abord, nous avons tenu à vérifier "expérimentalement" les propriétés statistiques générales sur les permutations et applications dont nous avons parlé en 2.1, et ce pour de grandes valeurs de n.

Il était bien entendu hors de question de générer $n!$ permutations ou $\mathrm{End}(n)$ applications pour n de l'ordre de 100 000, et d'établir des moyennes sur ces quantités faramineuses de données. Nous avons donc préféré utiliser un modèle probabiliste, et la loi des grands nombres : tirer 10 000 applications et permutations "au hasard", et établir des statistiques sur cet échantillon, que l'on espère représentatif, devrait être suffisant. Notons bien que l'on ne stocke pas les 10 000 applications : on génère les statistiques à la volée.

Un petit problème se pose alors : tirer une application "au hasard", i.e. de sorte que chaque application ait une chance sur $\mathrm{End}(n)$ de tomber, c'est facile : il suffit de tirer n nombres au hasard entre 1 et n. Mais comment tirer une permutation au hasard ? L'algorithme de Fisher-Yates, que l'on peut trouver dans [BB05], répond à ce problème : partant d'une liste ordonnée $[\![1; n]\!]$, pour i allant de n à 1, permuter le i^{e} élément avec un autre tiré au hasard dans $[\![1; i]\!]$.

L'algorithme fonctionne en temps raisonnable : pour des permutations, avec $n = 50\ 000$, le calcul se fait en 42 secondes, et il s'effectue en 90 secondes pour $n = 100\ 000$. L'étape critique étant celle qui consiste à calculer la décomposition en produits de cycles disjoints, le temps de calcul est négligeable lorsque l'on s'intéresse aux statistiques sur $\mathrm{End}(n)$.

Sans surprise, on retrouve bien les résultats théoriques.

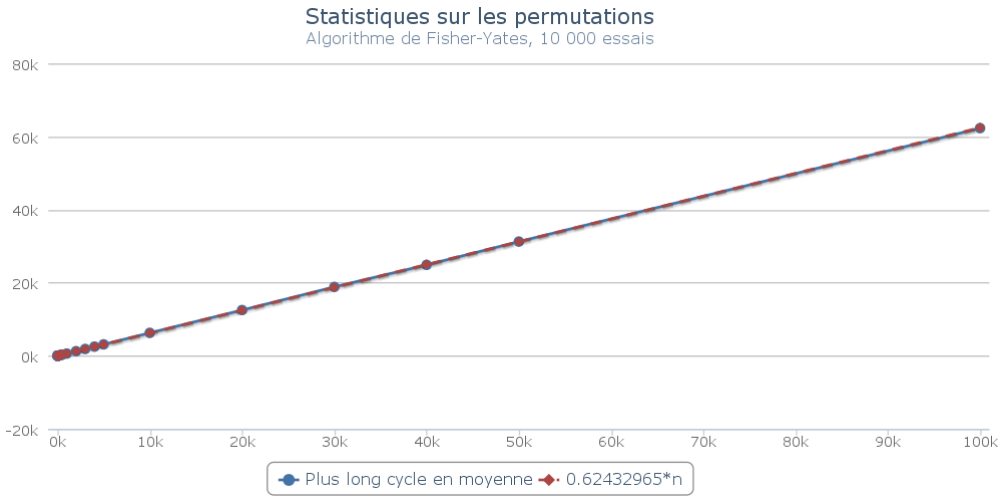

FIGURE 8 – $\frac{\mathbb{E}(L_n)}{n} \to 0.62432965\ldots$

On s'attaque maintenant au cœur de la simulation : la génération de polynômes aléatoires, leur réduction modulo p, et le calcul de statistiques. Nous nous sommes tout d'abord intéressés au cas, supposé plus simple, des polynômes $X^2 + c$, où $c \in [\![0, p-1]\!]$.

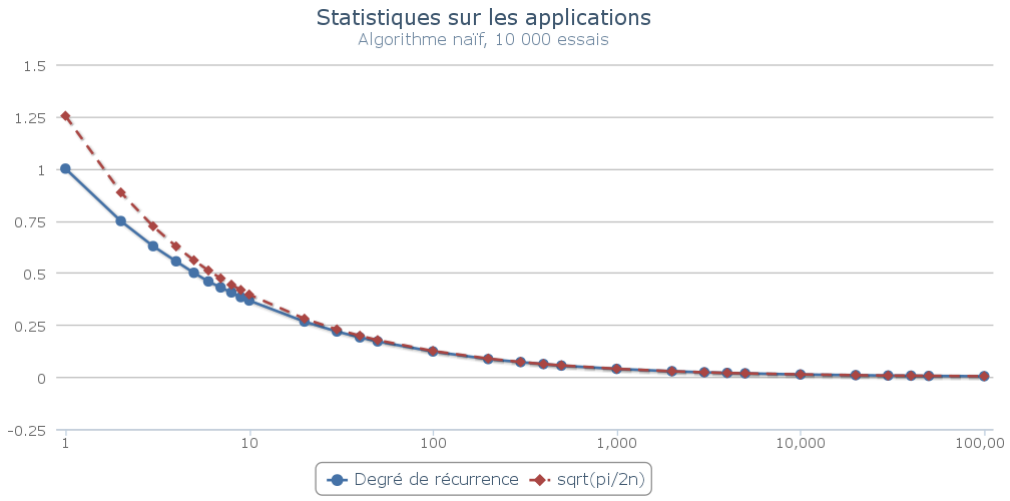

FIGURE 9 – $\frac{\mathbb{E}(\bar{\gamma})}{n} \sim \sqrt{\frac{\pi}{2 \cdot n}}$

Nous avons commencé par une étude probabiliste : pour p premier, nous avons tiré 10 000 constantes c au hasard de manière homogène dans $[\![0; p[\![$, et effectué sur les applications obtenues une étude statistique similaire aux précédentes.

Quelle ne fut pas notre surprise lorsque nous vîmes les courbes ci-dessous.

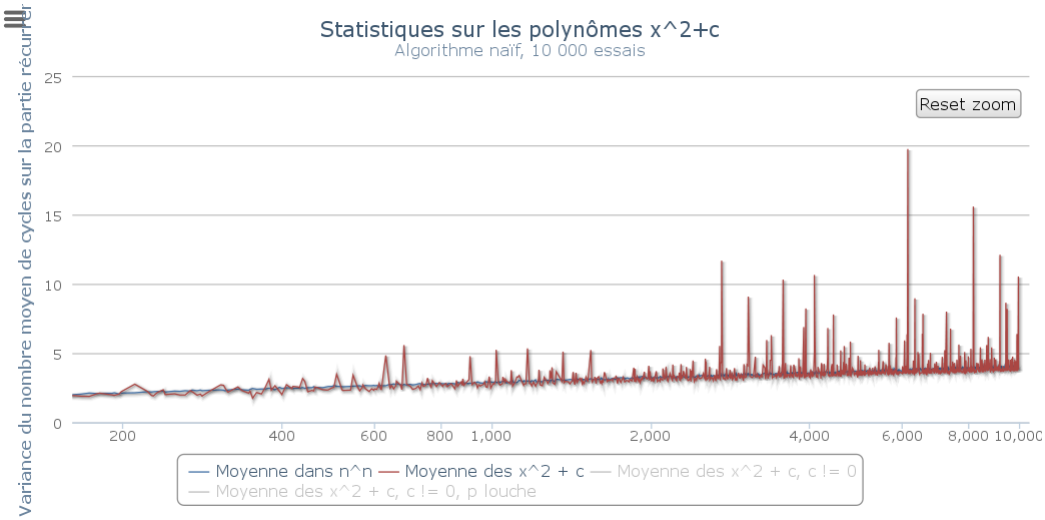

FIGURE 10 – Variance du nombre de cycles sur la partie récurrente

En réitérant les essais, nous nous aperçûmes alors que l'on avait affaire à un phénomène aléatoire d'apparition de pics. Pourquoi donc ? Au vu des données, on pouvait conjecturer que si, pour la plupart des valeurs de c, l'application X^2+c est "standard", il existe néanmoins certaines valeurs pour laquelle elle est très particulière. Les pics de variance correspondraient alors aux cas où l'on a tiré beaucoup de fois cette valeur de p.

Quelles pouvaient être les valeurs remarquables de c ? On pense d'abord à 0. Après tout, si une

valeur doit être remarquable, c'est bien celle-là, d'autant qu'on sait que X^2 est un endomorphisme de $\mathbb{Z}/p\mathbb{Z}^\star$.

Effectivement, si l'on choisit c dans $]\!]0;p[\!]$, on diminue de moitié la hauteur des pics. Qu'en est-il de la moitié restante ?

Pour le savoir, nous avons alors décidé de faire une étude complète, pour p premier allant de 2 à 10 000, en traçant, par exemple, l'histogramme tridimensionnelle du nombre de cycles sur la partie récurrente en fonction de p et $c \in [\![0;p[\!]$.

Le graphe obtenu, en figure 11, nous montre que pour $c = 0$ et $c = p-2$, le degré de récurrence explose, et est bien supérieur à celui obtenu pour les autres valeurs de c, qui stagne autour de $\sqrt{\frac{\pi}{2 \cdot p}}$.

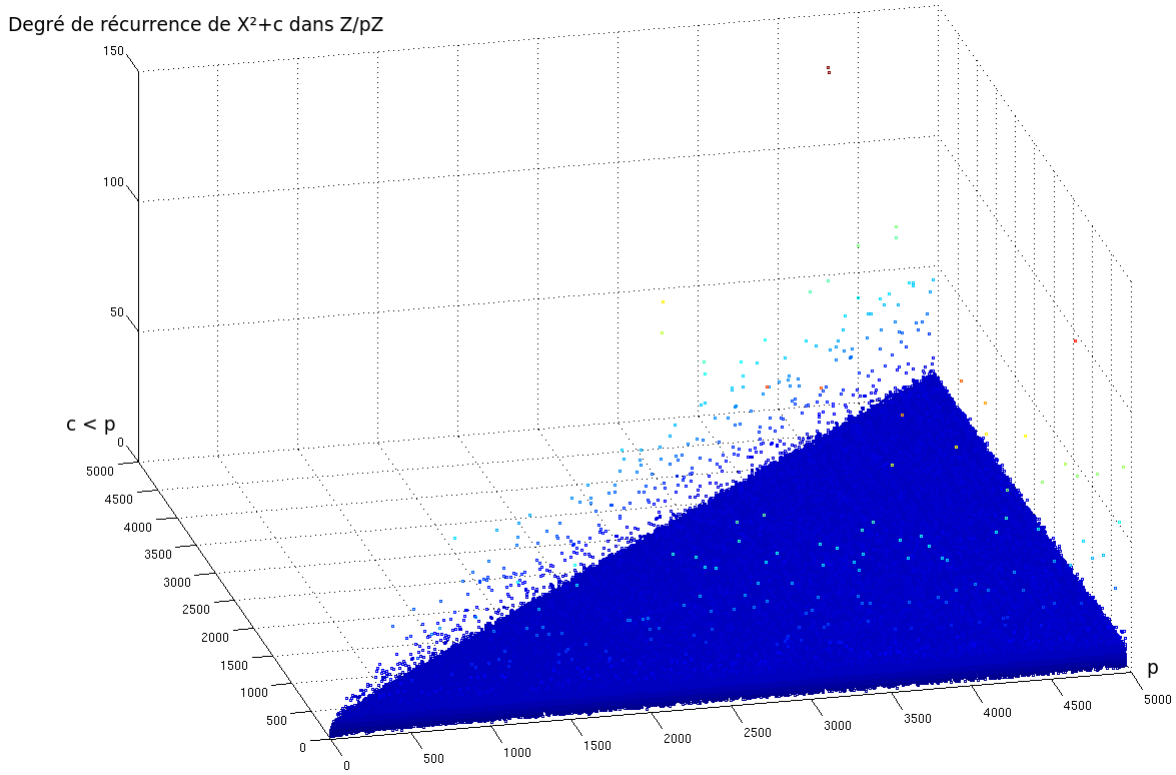

FIGURE 11 – Toutes les valeurs du degré de récurrence de X^2+c sur $\mathbb{Z}/p\mathbb{Z}$, pour $0 \leqslant c < p < 5\,000$

Il appuie donc l'hypothèse que nous avions formulé plus haut, avec deux valeurs de c remarquables : 0 et -2. Reste maintenant à comprendre pourquoi ces valeurs sont remarquables, et surtout, pourquoi les autres ne le sont pas !

La deuxième question n'a toujours pas de réponse satisfaisante à l'heure actuelle. Par contre, on sait quelle est la particularité de $x \mapsto x^2$ et $x \mapsto x^2 - 2$: ce sont des morphismes de groupes.

En effet, X^2 est un endomorphisme de $(\mathbb{U}, \cdot)$, où $\mathbb{U}$ est le cercle unité, tandis que X^2-2 agit sur $[-1;1]$ par projection de l'action précédente. En effet, X^2-2 est à un un changement de variable $z \mapsto z/2$ près le réduit unitaire du polynôme de Tchebytchev $2X^2-1$: $\cos(2\cdot\theta) = 2\cdot(\cos\theta)^2 - 1$.

Cette propriété donne ce caractère particulier à la dynamique de X^2 et X^2-2. Nous n'avons

pas creusé plus loin, mais on peut par exemple conjecturer que des polynômes de Tchebytchev de degré supérieur auront des propriétés similaires.

En dehors de ces cas (très) particuliers, les $X^2 + c$ semblent se comporter comme des applications typiques de End(n). On émet donc la conjecture que "la plupart" de ces réduits sont typiques.

Conjecture 1 (Silverman)**.** *En notant, pour $p \in \mathcal{P}$ et $\varphi \in \mathbb{Z}[X]$, $m_p(\varphi, x)$ l'ordre de x dans $\mathbb{Z}_{/p\mathbb{Z}}$, sous une condition à préciser sur φ,*

$$\exists C > 0, \text{ pour "la plupart des } p\text{"}, m_p \geqslant C\sqrt{p} \tag{11}$$

3 Le théorème de Silverman

L'objet de ce théorème est de présenter une forme (très) affaiblie de la conjecture 1, démontrée dans [Sil08]. En notant, pour $p \in \mathcal{P}$ et $\varphi \in \mathbb{Z}[X]$ vérifiant de bonnes hypothèses, $m_p(\varphi, x)$ l'ordre de x dans $\mathbb{Z}_{/p\mathbb{Z}}$, on va montrer que :

$$\forall \gamma < 1, \text{ pour "la plupart des } p\text{"}, m_p \geqslant (\log p)^\gamma \tag{12}$$

3.1 La densité logarithmique analytique

Il faut bien entendu donner un sens précis à "la plupart des p". On pense en premier lieu à la densité naturelle :

$$\forall \mathcal{Q} \subseteq \mathcal{P}, \Delta(\mathcal{Q}) := \lim \frac{|\mathcal{Q} \cap [1, n]|}{|\mathcal{P} \cap [1, n]|}$$

Δ n'étant bien définie que sur trop peu de parties de $\mathcal{P}$, on l'étend via la densité logarithmique analytique :

$$\forall \mathcal{Q} \subseteq \mathcal{P}, \delta(\mathcal{Q}) := \lim_{s \to 1^+} (s-1) \sum_{p \in \mathcal{Q}} \frac{\log p}{p^s}$$

Et on a :

Théorème 7. *Pour $\mathcal{Q} \subseteq \mathcal{P}$, si $\Delta(\mathcal{Q})$ est bien définie, alors $\delta(\mathcal{Q})$ aussi, et on a égalité entre les deux.*

Démonstration. On pourra trouver la preuve dans [Ten95]. □

Ce qui justifie notre usage de δ. Notons que, pour des raisons pratiques, on définit

$$\overline{\delta}(\mathcal{Q}) := \limsup_{s \to 1^+} (s-1) \sum_{p \in \mathcal{Q}} \frac{\log p}{p^s} \tag{13}$$

$$\underline{\delta}(\mathcal{Q}) := \liminf_{s \to 1^+} (s-1) \sum_{p \in \mathcal{Q}} \frac{\log p}{p^s} \tag{14}$$

Nous allons utiliser deux autres outils pour démontrer notre Théorème. Les voici.

3.2 Le Théorème des nombres premiers

Depuis Hadamard et De La Vallée Poussin, nous savons que, en notant π_n le n^e nombre premier, on a

$$\pi_n \sim n \cdot \log n \tag{15}$$

$$\frac{|\mathcal{P} \cap [2;n]|}{n} \sim \frac{n}{\log n} \tag{16}$$

On pourra trouver une démonstration détaillée de ce théorème dans le cours d'analyse complexe de M. Werner, ou dans [Had96].

3.3 La théorie des hauteurs

On utilisera de manière élémentaire la hauteur logarithmique absolue sur $\mathbb{Q} \cup \{\infty\} = P^1(\mathbb{Q})$, qui peut être vue comme une mesure de la complexité d'un rationnel :

$$h(r) := \log \max(|q|, |p|) \text{ où } r = \frac{p}{q} \text{ irréductible (i.e. } p \wedge q = 1)$$

Ce n'est là que le tout début de la théorie des hauteurs canoniques, développée dans [HS00], et qui a de nombreuses applications en géométrie diophantienne.

3.4 Le Théorème

On peut maintenant s'attaquer au théorème proprement dit. On définit proprement $m_p(\varphi, x)$: pour p premier ne divisant pas un coefficient de φ, on peut réduire φ à une fraction rationnelle de même degré $\tilde{\varphi}_p : P^1(\mathbb{F}_p) \to P^1(\mathbb{F}_p)$. Afin d'éviter de s'encombrer l'esprit avec un trop plein de nouvelles notations, on notera "$\varphi^n(x) \equiv \varphi^s(x) \bmod p$" pour "$\tilde{\varphi}_p^n(x) = \tilde{\varphi}_p^s(x)$". On peut alors poser :

$$\forall p \in \mathcal{P}, m_p := m_p(\varphi, x) := \begin{cases} \infty \text{ si } p \text{ ne divise aucun coefficient de } \varphi \\ \alpha(\varphi, x) = \min\{m, \exists r \geqslant 1, s \geqslant 0, r+s = m \text{ et } \varphi^{r+s}(x) \equiv \varphi^s(x) \bmod p\} \text{ sinon} \end{cases}$$

Le s correspond au temps qu'il faut à x pour tomber dans la partie récurrente de φ réduit modulo p.

On a alors :

Théorème 8 (Théorème de Silverman)**.** *Si $\varphi \in \mathbb{Q}(X)$, $x \in \mathbb{Q} \cup \{\infty\}$ sont tels que $\{\varphi^n(x), n \in \mathbb{N}\}$ soit infini, alors*

1. $\forall \gamma < 1, \delta\{p, m_p \geqslant (\log p)^\gamma\} = 1$
2. $\exists C, \forall \varepsilon > 0, \underline{\delta}\{p, m_p \geqslant \varepsilon \log p\} \geqslant 1 - C\varepsilon$

Démonstration. On fixe $0 < \gamma < 1$ (et $\lambda := \frac{1}{\gamma}$). $\mathcal{P}_\gamma := \{p \in \mathcal{P}, m_p < (\log p)^\gamma\}$. On définit aussi $g(x) := \frac{\log x}{x}$ et $G(y) := \frac{1}{e^{s y^{1/\gamma}}}$.

La preuve se fait en trois étapes. Tout d'abord, par des manipulations techniques (transformée d'Abel, principalement), on ramène le problème à la majoration de $\sum_{p \in \mathcal{P}, m_p \geqslant m} \frac{\log p}{p}$.

Ensuite, grâce au théorème des nombres premiers, on se ramène à la majoration de $\log\log \prod_{p \in \mathcal{P}, m_p \leqslant m} p$.

Enfin, et c'est la troisième étape, la plus importante, on utilise les hypothèses faites sur φ et x pour trouver $\log\log \prod_{m_p \leqslant m} p \leqslant K \cdot m$, ce qu'il nous fallait.

Voici un résumé de la preuve, qui sera détaillé plus bas.

$$\overline{\delta}(\mathcal{P}_\gamma) = \limsup_{s\to 1^+}(s-1)\sum_{p\in\mathcal{P}_\gamma}\frac{\log p}{p^s} \tag{17}$$

$$= \limsup_{s\to 0^+} s\sum_{p\in\mathcal{P}_\gamma}\frac{\log p}{p^{s+1}} \tag{18}$$

$$\leqslant \limsup_{s\to 0^+} s\sum_{p\in\mathcal{P}_\gamma}\frac{\log p}{pe^{s\log p}} \tag{19}$$

$$(\text{Définition de } \mathcal{P}_\gamma) \leqslant \limsup_{s\to 0^+} s\sum_{p\in\mathcal{P}_\gamma}\frac{\log p}{pe^{sm_p^{1/\gamma}}} \tag{20}$$

$$\leqslant \limsup_{s\to 0^+} s\sum_{p\in\mathcal{P}}\frac{\log p}{p}\cdot\frac{1}{e^{sm_p^{1/\gamma}}} \tag{21}$$

$$\leqslant \limsup_{s\to 0^+} s\sum_{p\in\mathcal{P}} g(p)G(m_p) \tag{22}$$

$$(\text{Fubini}) \leqslant \limsup_{s\to 0^+} s\sum_{m\geqslant 1}\sum_{p,m_p=m} g(p)G(m_p) \tag{23}$$

$$(\text{Abel}) \leqslant \limsup_{s\to 0^+} s\sum_{m\geqslant 1}(G(m)-G(m+1))\sum_{m_p\leqslant m} g(p) \tag{24}$$

$$(\text{Théorème des accroissements finis}) \leqslant \limsup_{s\to 0^+} s\sum_{m\geqslant 1} s\lambda(2m)^{\lambda-1}e^{-sm^\lambda}\sum_{m_p\leqslant m}\frac{\log p}{p} \tag{25}$$

$$\leqslant \limsup_{s\to 0^+} s\sum_{m\geqslant 1} s\lambda(2m)^{\lambda-1}e^{-sm^\lambda}\sum_{p|\mathcal{D}(m)}\frac{\log p}{p} \tag{26}$$

$$\leqslant \limsup_{s\to 0^+} s\sum_{m\geqslant 1} s\lambda(2m)^{\lambda-1}e^{-sm^\lambda} \tag{27}$$

$$\left(\sum_{p|\mathcal{D}(m),p>\log n}\frac{\log p}{p}+\sum_{p|\mathcal{D}(m),p\leqslant\log n}\frac{\log p}{p}\right) \tag{28}$$

$$(\text{Théorème des nombres premiers}) \leqslant \limsup_{s\to 0^+} s\sum_{m\geqslant 1} s\lambda(2m)^{\lambda-1}e^{-sm^\lambda}(1+O(\log\log\mathcal{D}(m))) \tag{29}$$

$$\leqslant \limsup_{s\to 0^+} s\sum_{m\geqslant 1} s\lambda(2m)^{\lambda-1}e^{-sm^\lambda}(c_1\log\log\mathcal{D}(m)+c_2) \tag{30}$$

$$(\text{Caractérisation de la taille des orbites}) \leqslant \limsup_{s\to 0^+} s\cdot C\sum_{m\geqslant 1} m^\lambda e^{-sm^\lambda} \tag{31}$$

$$\leqslant \limsup_{s\to 0^+} s\frac{K}{s^{1/\lambda}} \tag{32}$$

$$= 0 \tag{33}$$

Attention : l'étape critique est l'équation 31, qui sera expliquée en détail dans le lemme 3. Maintenant, détaillons un peu.

Transformation d'Abel et théorème des accroissements finis On a :

$$\overline{\delta}(\mathcal{P}_\gamma) = \limsup_{s\to 1^+}(s-1)\sum_{p\in\mathcal{P}_\gamma}\frac{\log p}{p^s} \tag{34}$$

$$= \limsup_{s\to 0^+} s\sum_{p\in\mathcal{P}_\gamma}\frac{\log p}{p^{s+1}} \tag{35}$$

$$\leqslant \limsup_{s\to 0^+} s\sum_{p\in\mathcal{P}_\gamma}\frac{\log p}{pe^{s\log p}} \tag{36}$$

$$\leqslant \limsup_{s\to 0^+} s\sum_{p\in\mathcal{P}_\gamma}\frac{\log p}{pe^{sm_p^{1/\gamma}}} \tag{37}$$

$$\leqslant \limsup_{s\to 0^+} s\sum_{p\in\mathcal{P}}\frac{\log p}{pe^{sm_p^{1/\gamma}}} \tag{38}$$

Or on a :

Théorème 9 (Estimation analytique)**.** $\forall\lambda\geqslant 1, \exists C, \forall s>0, \sum_{p\in\mathcal{P}}\frac{\log p}{pe^{sm_p^\lambda}} \leqslant \frac{C}{s^{1/\lambda}}$

Démonstration. Avec $g(t) := \frac{\log t}{t}$, $G(t) := e^{-st^\lambda}$, et S la somme, on a :

$$S = \sum_{p\in\mathcal{P}} g(p)G(m_p) \tag{39}$$

$$\text{(Fubini)} = \sum_{m\geqslant 1}\sum_{m_p=m} g(p)G(m_p) \tag{40}$$

$$\text{(Abel)} = \sum_{m\geqslant 1}(G(m)-G(m+1))\sum_{p,m_p\leqslant m} g(p) \tag{41}$$

$$\text{(Théorème des accroissements finis)} \leqslant \sum_{m\geqslant 1} s\lambda(2m)^{\lambda-1}e^{-sm^\lambda}\sum_{m_p\leqslant m}\frac{\log p}{p} \tag{42}$$

On veut majorer $\sum_{m_p\leqslant m}\frac{\log p}{p}$.
On pose $\mathcal{D}(m) := \prod_{m_p\leqslant m} p$. On a donc

$$m_p \leqslant m \iff p|\mathcal{D}(m)$$

Et alors :

$$\sum_{m_p\leqslant m}\frac{\log p}{p} = \sum_{p|\mathcal{D}(m)}\frac{\log p}{p} \tag{43}$$

Utilisation du théorème des nombres premiers

Lemme 2 (Variations on a theme of Romanoff, Cor. 2.3)**.**

$$\exists c_1, c_2, \forall\mathcal{D}\geqslant 2, \sum_{p|\mathcal{D}}\frac{\log p}{p} \leqslant c_1\log\log\mathcal{D} + c_2 \tag{44}$$

Démonstration. Une preuve sans Théorème des nombres premiers est donnée dans l'article [Sil08].
Il suffit de montrer que

$$\sum_{p|n} \frac{\log p}{p} = O(\log\log n) \tag{45}$$

On a :

$$\sum_{p|n} \frac{\log p}{p} = \sum_{p|n, p>\log n} \frac{\log p}{p} + \sum_{p|n, p\leqslant \log n} \frac{\log p}{p} \tag{46}$$

Où :

– Avec $f(n) := |\{p \in \mathcal{P}|\ p|n \text{ et } p > \log n\}|$, avec $p \in \mathcal{P}$ implicite dans les sommes,

$$n \geqslant \prod_{p|n, p>\log n} p \geqslant (\log n)^{f(n)}$$

d'où :

$$\sum_{p|n, p>\log n} \frac{\log p}{p} \leqslant f(n)\frac{\log\log n}{\log n} \tag{47}$$

$$\leqslant 1 \tag{48}$$

– De l'autre côté, avec $p \in \mathcal{P}$ implicite dans les sommes,

$$\sum_{p|n, p\leqslant \log n} \frac{\log p}{p} \leqslant \sum_{p\leqslant \log n} \frac{\log p}{p} \tag{49}$$

$$\text{(Théorème des nombres premiers)} = O\left(\sum_{k=1}^{\frac{2\log n}{\log\log n}} \frac{\log \pi_k}{\pi_k}\right) \tag{50}$$

$$\text{(Théorème des nombres premiers)} \sim \sum_{k=1}^{\frac{2\log n}{\log\log n}} \frac{\log(k\cdot\log k)}{k\cdot\log k} \tag{51}$$

$$\sim \sum_{k=1}^{\frac{2\log n}{\log\log n}} \frac{\log(k)}{k\cdot\log k} \tag{52}$$

$$\sim \sum_{k=1}^{\frac{2\log n}{\log\log n}} \frac{1}{k} \tag{53}$$

$$\sim \log\log(\frac{2\log n}{\log\log n}) \tag{54}$$

$$\sim \log\log n \tag{55}$$

D'où le Lemme voulu. □

On a donc :

$$\sum_{m_p\leqslant m} \frac{\log p}{p} = \sum_{p|\mathcal{D}(m)} \frac{\log p}{p} \tag{56}$$

$$\leqslant c_1 \log\log \mathcal{D}(m) + c_2 \tag{57}$$

Hauteur et dynamique On utilise maintenant les hypothèses faites sur φ pour obtenir :

Lemme 3 (Caractérisation de la taille des orbites)**.**

$$\exists C, \forall m \geqslant 1, \log\log \mathcal{D}(m) \leqslant Cm$$

Démonstration. Notons bien que c'est ce lemme qui contient toute l'information, et qui utilise les hypothèses. Le reste de la preuve n'est que détails techniques permettant de passer du lemme à la densité logarithmique analytique.

On a, par définition,

$$m_p = \min\{m, \exists r \geqslant 1, s \geqslant 0, r+s = m \text{ et } \varphi^{r+s}(x) \equiv \varphi^s(x) \bmod p\}$$

Via la mise sous forme canonique des rationnels, on écrit :

$$\varphi^n(x) =: [A_0(n), A_1(n)], \text{ où } \begin{cases} A_i(n) \in \mathbb{Z} \\ A_0 \wedge A_1 = 1 \end{cases}$$

Pour tout $p \in \mathcal{P}$, on a alors :

$$\varphi^{r+s}(x) \equiv \varphi^s(x) \bmod p \Leftrightarrow [A_0(r+s), A_1(r+s)] \equiv [A_0(s), A_1(s)] \bmod p \tag{58}$$

$$\Leftrightarrow A_0(r+s)A_1(s) \equiv A_0(s)A_1(r+s) \bmod p \tag{59}$$

Ainsi, avec

$$\mathcal{B}(r,s) := |A_0(r+s)A_1(s) - A_1(r+s)A_0(s)| \tag{60}$$

$$\neq 0 \text{ car l'orbite de } x \text{ est infinie} \tag{61}$$

$$\mathcal{D}'(m) := \prod_{r+s=m, r\geqslant 1} \mathcal{B}(r,s) \tag{62}$$

On a :

$$\forall p \in \mathcal{P}, m_p \leqslant m \Longleftrightarrow p | \mathcal{D}'(m)$$

d'où $\mathcal{D}(m)|\mathcal{D}'(m)$, donc $\mathcal{D}(m) \leqslant \mathcal{D}'(m)$ (car $\mathcal{D}' \neq 0$; insistons de nouveau : c'est bien ici que l'on utilise l'hypothèse de non-finitude de l'orbite de x dans $\mathbb{Z}$) : il suffit de montrer la majoration pour $\mathcal{D}'$, i.e. montrer que :

$$\exists C, \forall m \geqslant 1, \log\log \mathcal{D}'(m) \leqslant Cm$$

On sait que :

$$\log \mathcal{D}'(m) = \sum_{r+s=m} \log \mathcal{B}(r,s)$$

On a alors la bonne idée d'introduire la hauteur logarithmique absolue,

$$h(r) := \log \max(|q|, |p|) \text{ où } r = \frac{p}{q} \text{ irréductible}$$

On peut voir cette hauteur comme une mesure de la complexité d'un élément de $\mathbb{Q} \cup \{\in \infty\}$. On a alors :

$$\log \mathcal{B}(r,s) \leqslant h(\varphi^{r+s}(x)) + h(\varphi^s(x)) + \log 2 \tag{63}$$

Ce qui revient à dire que $\mathcal{B}(r,s)$ est borné par la "complexité" de $\varphi^{r+s}(x)$ et $\varphi^s(x)$. Or on a

Lemme 4.

$$\exists d \geqslant 2, \exists C \geqslant 0, \forall n, \forall x, h(\varphi^n(x)) \leqslant d^n(h(x) + C)$$

Démonstration. Clair pour φ polynomiale, avec d le degré de φ, un peu plus compliqué pour φ fraction rationnelle. □

Moralité : Une application polynomiale n'augmente que très raisonnablement la complexité. Aussi, on a une borne sur $\mathcal{B}(r, s)$:

$$\forall r, s, \log(\mathcal{B}(r, s)) \leqslant Cd^{r+s} \tag{64}$$
$$\tag{65}$$

Il suffit alors d'utiliser cette borne pour majorer $\mathcal{D}'(m)$:

$$\begin{aligned}
\log \mathcal{D}'(m) &= \sum_{r+s=m} \log \mathcal{B}(r, s) && (66)\\
&\leqslant \sum_{r+s=m} Cd^{r+s} && (67)\\
&\leqslant Cmd^{m+1} && (68)\\
&\leqslant Cd^{2m} && (69)\\
& && (70)
\end{aligned}$$

Et finalement

$$\begin{aligned}
\log\log \mathcal{D}(m) &\leqslant \log\log \mathcal{D}'(m) && (71)\\
&\leqslant Km && (72)
\end{aligned}$$

□

Conclusion Reprenons. On avait :

$$\begin{aligned}
\sum_{p\in\mathcal{P}} \frac{\log p}{pe^{sm_p^\lambda}} &\leqslant \sum_{m\geqslant 1} s\lambda(2m)^{\lambda-1}e^{-sm^\lambda} \sum_{m_p\leqslant m} \frac{\log p}{p} && (73)\\
&\leqslant \sum_{m\geqslant 1} s\lambda(2m)^{\lambda-1}e^{-sm^\lambda} \sum_{p|\mathcal{D}(m)} \frac{\log p}{p} && (74)\\
&\leqslant \sum_{m\geqslant 1} s\lambda(2m)^{\lambda-1}e^{-sm^\lambda}(c_1 \log\log \mathcal{D}(m) + c_2) && (75)
\end{aligned}$$

On a donc désormais montré :

$$\begin{aligned}
\sum_{p\in\mathcal{P}} \frac{\log p}{pe^{sm_p^\lambda}} &\leqslant \sum_{m\geqslant 1} s\lambda(2m)^{\lambda-1}e^{-sm^\lambda}(c_1 \cdot K \cdot m + c_2) && (76)\\
&\leqslant Cs \sum_{m\geqslant 1} m^\lambda e^{-sm^\lambda} && (77)\\
&\leqslant \frac{K}{s^{1/\lambda}} && (78)
\end{aligned}$$

En effet, on a le fait d'analyse suivant :

Lemme 5.

$$\forall \lambda > 0, \forall \mu \geqslant 0, \exists C, \forall s > 0, \sum_{m=1}^{\infty} m^{\mu} e^{-sm^{\lambda}} \leqslant C s^{-(\mu+1)/\lambda}$$

Démonstration. $t \mapsto t^{\mu} e^{-st^{\lambda}}$ a un unique maximum sur $[0; +\infty[$, donc

$$\sum_{m=1}^{\infty} m^{\mu} e^{-sm^{\lambda}} \leqslant 2 \int_0^{+\infty} t^{\mu} e^{-st^{\lambda}} dt \tag{79}$$

$$= 2 s^{-(\mu+1)/\lambda} \int_0^{+\infty} u^{\mu} e^{-u^{\lambda}} du \tag{80}$$

□

On a donc bien démontré que :

$$\forall \lambda \geqslant 1, \exists C, \forall s > 0, \sum_{p \in \mathcal{P}} \frac{\log p}{p e^{s m_p^{\lambda}}} \leqslant \frac{C}{s^{1/\lambda}}$$

□

On peut maintenant terminer la preuve :

$$\overline{\delta}(\mathcal{P}_{\gamma}) \leqslant \limsup_{s \to 0^+} s \sum_{p \in \mathcal{P}} \frac{\log p}{p e^{s m_p^{1/\gamma}}} \tag{81}$$

$$\text{(Estimation analytique)} \leqslant \limsup_{s \to 0^+} C s^{1-\gamma} \tag{82}$$

$$(0 < \gamma < 1) = 0 \tag{83}$$

Aussi, $\delta(\mathcal{P}_{\gamma}) = 0$ et donc $\delta(\mathcal{P}_{\gamma}^C) = 1$, ce qui conclut la preuve du premier point.
Pour le deuxième : posons $\mathcal{P}_{\varepsilon} := \{p \in \mathcal{P}, m_p < \varepsilon \log p\}$. On a :

$$\overline{\delta}(\mathcal{P}_{\varepsilon}) = \limsup_{0^+} s \sum_{\mathcal{P}_{\varepsilon}} \frac{\log p}{p^{1+s}} \tag{84}$$

$$= \limsup_{0^+} u\varepsilon \sum_{\mathcal{P}_{\varepsilon}} \frac{\log p}{p^{1+u\varepsilon}} \tag{85}$$

$$\leqslant \limsup_{0^+} u\varepsilon \sum_{\mathcal{P}_{\varepsilon}} \frac{\log p}{p e^{u\varepsilon \log p}} \tag{86}$$

$$\leqslant \limsup_{0^+} u\varepsilon \sum_{\mathcal{P}_{\varepsilon}} \frac{\log p}{p e^{u m_p}} \tag{87}$$

$$\leqslant \limsup_{0^+} u\varepsilon \sum_{\mathcal{P}} \frac{\log p}{p e^{u m_p}} \tag{88}$$

$$\leqslant u\varepsilon \frac{C}{u} \tag{89}$$

$$\leqslant C\varepsilon \tag{90}$$

□

3.5 Résultats annexes

On constate que, si φ est de degré 1, on a une augmentation extrêmement modeste de la complexité de x par l'action de φ :

$$h(\varphi^n(x)) \leqslant h(x) + Cn$$

En déroulant la preuve, on arrive à conserver cet ordre de grandeur en plus, et on obtient alors un porisme du théorème précédent, contenu dans [MRS96] :

Théorème 10 (Porisme, fort mais restrictif)**.**

$$\underline{\delta}\{p, m_p \geqslant p^{\varepsilon}\} \geqslant 1 - 2\varepsilon$$

4 Annexes

4.1 Le théorème ergodique de Birkhoff

Soit $(X, \mathcal{B}, \mu)$ un espace de probabilité et $T : X \to X$ une bijection mesurable d'inverse mesurable. On dira que T préserve la mesure si pour tout $E \in \mathcal{B}, \mu(T^{-1}E) = \mu(E)$.
Si $f : X \to \mathbb{R}$ est mesurable, on pose $< f >_n = \frac{1}{n}\sum_{k=0}^{n-1} f \circ T^k$ qui est une moyenne de f. Se pose alors la question de la convergence de $< f >_n$.

Pour montrer la convergence ponctuelle des $< f >_n$, on s'intéresse d'abord au maximum des moyennes

$$f^*(x) = \sup_{n>0} < f >_n (x).$$

On a alors :

Théorème 11 (Théorème ergodique maximal)**.** *Si $f \in L^1(X, \mathcal{B}, \mu)$, alors*

$$\int_{\{f^*>0\}} f \mathrm{d}\mu \geqslant 0$$

Démonstration. On en donne deux : une rapide, et une plus explicite.

Première preuve : Soit $M_n f = \sup_{k\leqslant n} < f >_k$ et $K_n f = \sup_{k\leqslant n} \sum_{k=0}^{n-1} f \circ T^k$.

On remarque déjà que $M_n f > 0 \iff K_n f > 0$.

De plus,

$$K_n f \leqslant K_{n+1} f = K_{n+1} f = f + \sup_{k\leqslant n+1} (\sum_{i=0}^{k-2} f \circ T^i) \circ T$$

or, en notant $x^+ = \max(x, 0)$, on a

$$\sup_{k\leqslant n+1} (\sum_{i=0}^{k-2} f \circ T^i) \circ T = (K_n f)^+ \circ T$$

si $K_n f < 0$, alors dans le membre de gauche 0 est atteint pour $k = 1$ (somme vide) et il y a égalité, sinon le membre de gauche est atteint pour $k > 1$ et alors c'est exactement $K_n f$.

En intégrant sur $K_n f > 0$, il vient

$$\int_X (K_n f)^+ \mathrm{d}\mu = \int_{\{K_n f>0\}} K_n f \mathrm{d}\mu \tag{91}$$

$$\leqslant \int_{\{K_n f>0\}} f \mathrm{d}\mu + \int_{\{K_n f>0\}} (K_n f)^+ \circ T \mathrm{d}\mu \tag{92}$$

$$= \int_{\{K_n f>0\}} f \mathrm{d}\mu + \int_X (K_n f)^+ \circ T \mathrm{d}\mu \tag{93}$$

$$\text{(T préserve la mesure)} = \int_{\{K_n f>0\}} f \mathrm{d}\mu + \int_X (K_n f)^+ \mathrm{d}\mu \tag{94}$$

Or $f \in L^1$, donc $(K_n f)^+ \in L^1$, donc $\int_X (K_n f)^+ \mathrm{d}\mu < \infty$. D'où $\int_{\{K_n f>0\}} f \mathrm{d}\mu \geqslant 0$, et finalement par convergence dominée appliqué à $g_n = \chi_{\{K_n f>0\}} f$

$$\int_{\{f^*>0\}} f \mathrm{d}\mu \geqslant 0$$

.

Seconde preuve :

L'idée est que si $x \in X$ est tel que $K_n f(x) > 0$ mais que $f(x) \leqslant 0$, alors $K_{n-1} f(Tx) > 0$, et en continuant ainsi on peut regrouper les termes pour faire apparaitre $K_n f(x)$ dans l'intégrale $\int_{\{f^*>0\}} f \mathrm{d}\mu$, qui sera une contribution positive. Détaillons cela :

Pour $x \in f^*(x) > 0$, soit $n(x) = \inf\{n > 0 : K_n f(x) > 0\} < \infty$ et $B_n = \{x : n(x) = n\}$.

Il suffit de montrer $\int_{B_1 \cup \ldots \cup B_n} f \mathrm{d}\mu \geqslant 0$.

Pour cela, on découpe l'union en regroupant correctement les éléments pour faire apparaitre des termes positifs.

Remarquons déjà que

$$(*)\ k < n \Rightarrow T^k B_n \subset B_1 \cup \ldots \cup B_{n-k}$$

.

En effet, si $x \in B_n$, alors

$$f(x) + f(Tx) + \ldots + f(T^{k-1}x) \leqslant 0$$

tandis que

$$f(x) + f(Tx) + \ldots + f(T^{k-1}x) + f(T^k x) + \ldots + f(T^{n-1}x) > 0$$

donc

$$f(T^k x) + f(T(T^k x)) + \ldots + f(T^{n-k-1}(T^k x)) = f(T^k x) + \ldots + f(T^{n-1}x) > 0$$

De plus, les ensembles $B_n, TB_n, \ldots, T^{n-1}B_n$ sont deux à deux disjoints, car si $i \leqslant j$ et $x \in T^i B_n \cap T^j B_n$, alors $B_n \cap T^{j-i} B_n \neq \emptyset$, et donc $i = j$ par $(*)$.

On pose alors

$$B'_n = B_n, C_n = B_n \cup TB_n \cup \ldots \cup T^{n-1}B_n \tag{95}$$

$$B'_{n-1} = B_{n-1} \setminus C_n, B'_{n-1} \cup TB'_{n-1} \cup \ldots \cup T^{n-2}B'_{n-1} \tag{96}$$

$$\vdots \tag{97}$$

$$B'_1 = B_1 \setminus (C_2 \cup \ldots \cup C_n), C_1 = B'_1 \tag{98}$$

Alors la famille $(C_i)_{1\leqslant i\leqslant n}$ est une partition de $B_1 \cup \ldots \cup B_n$, et on a

$$\int_{B_1\cup\ldots\cup B_n} f\mathrm{d}\mu = \sum_{k=1}^{n} \int_{C_k} f\mathrm{d}\mu \tag{99}$$

$$= \sum_{k=1}^{n} \int_{B'_k\cup TB'_k\cup\ldots\cup T^{k-1}B'_k} f\mathrm{d}\mu \tag{100}$$

$$= \sum_{k=1}^{n} \int_{B'_k} (f + f\circ T + \ldots + f\circ T^{k-1})\mathrm{d}\mu \geqslant 0 \tag{101}$$

□

Corollaire 1. *Pour tout $\alpha \in \mathbb{R}$,*

$$\int_{f^*>\alpha} f\mathrm{d}\mu \geqslant \alpha\mu\{f^* > \alpha\}$$

Démonstration. On pose $g = f - \alpha$ qui est bien L^1 car $\mu(X) < \infty$. Alors

$$0 \leqslant \int_{\{g^*>0\}} g\mathrm{d}\mu = \int_{\{f^*>\alpha\}} (f-\alpha)\mathrm{d}\mu$$

□

On peut désormais démontrer le théorème de Birkhoff

Théorème 12 (Théorème ergodique de Birkhoff)**.** *Soit $(X, \mathcal{B}, \mu)$ un espace de probabilité et $T : X \to X$ une bijection mesurable d'inverse mesurable qui préserve la mesure : pour tout $E \in \mathcal{B}, \mu(T^{-1}E) = \mu(E)$.*
Si $f \in L^1(X, \mathcal{B}, \mu)$, alors :

1. *$\overline{f}(x) = lim_{n\to\infty}(1/n)\sum_{k=0}^{n-1} f(T^k x)$ existe p.s.*
2. *$\overline{f} \in L^1$ et $\|\overline{f}\|_1 \leqslant \|f\|_1$*
3. *$\overline{f}(Tx) = \overline{f}(x)$ p.s., ie $\overline{f}\circ T = \overline{f}$ dans L^1*
4. *$\forall A \in \mathcal{B}, T^{-1}A = A \Rightarrow \int_A f\mathrm{d}\mu = \int_A \overline{f}\mathrm{d}\mu$*
5. *$(1/n)\sum_{k=0}^{n-1} f(T^k x) \longrightarrow \overline{f}$ dans L^1*

Démonstration. Pour tout $\alpha < \beta$ dans $\mathbb{R}$, soit

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

Si on montre que $\mu(E_{\alpha,\beta}) = 0$ pour tout $\alpha < \beta$ dans $\mathbb{R}$, alors

$$\mu\Big(\bigcup_{\substack{\alpha,\beta\in\mathbb{Q}\\ \alpha<\beta}} E_{\alpha,\beta}\Big) = 0$$

ce qui montre l'assertion 1.

On sait que $E_{\alpha,\beta}$ est un sous-ensemble de $\{f^* > \beta\}$ invariant par T, et en appliquant le théorème ergodique maximal à $T : E_{\alpha,\beta} \to E_{\alpha,\beta}$ et $\{f^* > \beta\} \cap E_{\alpha,\beta} = E_{\alpha,\beta}$, on obtient

$$\int_{E_{\alpha,\beta}} f\mathrm{d}\mu \geqslant \beta\mu(E_{\alpha,\beta})$$

On considère maintenant $-f$. On a $E_{\alpha,\beta} \subset \{(-f)^* > -\alpha\}$, et de même que précédemment, on obtient

$$\int_{E_{\alpha,\beta}} -f \mathrm{d}\mu \geqslant -\alpha\mu(E_{\alpha,\beta}).$$

Finalement,

$$\beta\mu(E_{\alpha,\beta}) \leqslant \int_{E_{\alpha,\beta}} f \mathrm{d}\mu \leqslant \alpha\mu(E_{\alpha,\beta})$$

or $\alpha < \beta$, donc $\mu(E_{\alpha,\beta}) = 0$

Pour le point 2, on utilise le lemme de Fatou. On remarque d'abord que comme

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

on a $|\overline{f}| \leqslant \overline{|f|}$ et donc

$$\int \left|\overline{f}\right| \mathrm{d}\mu \leqslant \int \overline{|f|} \mathrm{d}\mu \tag{102}$$

$$\text{(lemme de Fatou)} \quad \leqslant \liminf_{n\to\infty} \int \frac{1}{n} \sum_{k=0}^{n-1} |f(T^k x)| \mathrm{d}\mu \tag{103}$$

$$= \int |f| \mathrm{d}\mu < \infty \tag{104}$$

Le point 4 peut être prouvé de deux manières : avec le théorème ergodique maximal, ou en utilisant le point 5, ce que l'on fera après. On pose $A_{n,k} = \{x \in A : \frac{k}{2^n} \leqslant \overline{f} < \frac{k+1}{2^n}\}$. Les $A_{n,k}$ sont invariants (par T) et pour tout n, $A = \bigcup_k A_{n,k}$ union disjointe.

Soit maintenant $\varepsilon > 0$. Sur $A_{n,k}$, on a $f^* > k/2^n - \varepsilon$, donc par le théorème ergodique maximal appliqué à $T : A_{n,k} \to A_{n,k}$, on obtient

$$\int_{A_{n,k}} f \mathrm{d}\mu \geqslant (\frac{k}{2^n})\mu(A_{n,k}).$$

De même, $(-f)^* > -(k+1)/2^n$ sur $A_{n,k}$, donc

$$\int_{A_{n,k}} -f \mathrm{d}\mu \geqslant -\frac{k+1}{2^n}\mu(A_{n,k}).$$

Et en faisant $\varepsilon \longrightarrow 0$, on obtient

$$\frac{k}{2^n}\mu(A_{n,k}) \leqslant \int_{A_{n,k}} f \mathrm{d}\mu \leqslant \frac{k+1}{2^n}\mu(A_{n,k})$$

De plus par définition de $A_{n,k}$, on a

$$\frac{k}{2^n}\mu(A_{n,k}) \leqslant \int_{A_{n,k}} \overline{f} \mathrm{d}\mu \leqslant \frac{k+1}{2^n}\mu(A_{n,k})$$

et on obtient

$$\left| \int_{A_{n,k}} f \mathrm{d}\mu - \int_{A_{n,k}} \overline{f} \mathrm{d}\mu \right| \leqslant \frac{1}{2^n}\mu(A_{n,k})$$

Et en sommant sur k, les $A_{n,k}$ et $A_{n,k'}$ étant disjoints pour $k \neq k'$, on obtient

$$\left|\int_A f\mathrm{d}\mu - \int_A \overline{f}\mathrm{d}\mu\right| \leqslant \frac{1}{2^n}\mu(A)$$

Il suffit alors de faire $n \longrightarrow \infty$.

On montre désormais le point 5.

Si f est bornée, cela découle du théorème de convergence dominée. On raisonne par approximation pour le cas général, car les fonctions bornée sont denses dans L^1. (On voit qu'on utilise fortement la finitude de la mesure de X)

Soit $\varepsilon > 0$. Quitte à considérer f^+ et f^-, on suppose $f \geqslant 0$. Soit g bornée telle que $\|f-g\|_1 \leqslant \frac{\varepsilon}{3}$. On a alors

$$\left\|\frac{1}{n}\sum_{k=0}^{n-1} f\circ T^k - \overline{f}\right\|_1 \leqslant \left\|\frac{1}{n}\sum_{k=0}^{n-1}(f\circ T^k - g\circ T^k)\right\|_1 + \left\|\frac{1}{n}\sum_{k=0}^{n-1} g\circ T^k - \overline{g}\right\|_1 + \left\|\overline{g}-\overline{f}\right\|_1.$$

Or par 3, $\left\|\overline{g}-\overline{f}\right\|_1 \leqslant \|g-f\|_1$ De plus, comme g est bornée, à partir d'un certain rang,

$$\left\|\frac{1}{n}\sum_{k=0}^{n-1} g\circ T^k - \overline{g}\right\|_1 \leqslant \frac{\varepsilon}{3}$$

Et finalement, à partir d'un certain rang,

$$\left\|\frac{1}{n}\sum_{k=0}^{n-1} f\circ T^k - \overline{f}\right\|_1 \leqslant \varepsilon$$

On donne finalement l'autre preuve du point 4.

$$\left|\int_A f\mathrm{d}\mu - \int_A \overline{f}\mathrm{d}\mu\right| = \left|\int_A \left(\frac{1}{n}\sum_{k=0}^{n-1} f\circ T^k - \int_A \overline{f}\right)\mathrm{d}\mu\right| \tag{105}$$

$$\leqslant \int_A \left|\frac{1}{n}\sum_{k=0}^{n-1} f\circ T^k - \overline{f}\right|\mathrm{d}\mu \tag{106}$$

$$= \left\|\frac{1}{n}\sum_{k=0}^{n-1} f\circ T^k - \overline{f}\right\|_{L^1(A)} \longrightarrow 0 \tag{107}$$

par 5. □

4.2 La stabilité structurelle

On étudie ici certains automorphismes du tore $\mathbb{T} = \mathbb{R}^2/\mathbb{Z}^2$ de dimension deux.

Si $\overline{A} \in GL_2(\mathbb{Z})$, alors l'application $A : \mathbb{T} \to \mathbb{T}$ est bien définie, et est un difféomorphisme. On parlera des valeurs propres, du déterminant, ... de A pour ceux de $\overline{A}$.

Théorème 13 (Anosov)**.** *On suppose que l'automorphisme A précédent n'a pas de valeur propre de module* 1*.*
Alors si B est un difféomorphisme de $\mathbb{T}$ assez C^1-proche de A, il existe H homéomorphisme tel que $B = H^{-1}\circ A\circ H$.

Démonstration. Si on écrit $\overline{H} = Id_{|\mathbb{R}^2} + h$, $\overline{B} = \overline{A} + f$ où h et f sont bornées et telles que si $x \in \mathbb{R}^2, k \in \mathbb{Z}^2$, alors $f(x+k) = f'x)$ et $h(x+k) = h(x)$.
Trouver un $\overline{H}$ convenable revient à résoudre l'équation fonctionnelle

$$h \circ \overline{A} - \overline{A} \circ h = f \circ (Id + h),$$

que l'on résout sur le plan $\mathbb{R}^2$. Comme ici f et h sont petits, on résout d'abord l'équation plus simple

$$h \circ \overline{A} - \overline{A} \circ h = f.$$

Si $h : \mathbb{T} \to \mathbb{T}$ (que l'on peut voir comme le h précédent), on pose $Lh = h \circ \overline{A} - \overline{A} \circ h$. L est alors un opérateur linéaire. On montre d'abord que L est inversible.
Pour cela on remarque que comme $|det(A)| = 1$, les deux valeurs propres λ_1 et λ_2 de A sont distinctes et par exemple $|\lambda_1| > 1 > |\lambda_2|$. A est donc diagonalisable. Soit (e_1, e_2) une base de diagonalisation de A.
On fixe f difféomorphisme de $\mathbb{T}$ et on cherche h tel que $Lh = f$, ie $h(\overline{A}x) - \overline{A}h(x) = f(x)$. On écrit $f = f_1 e_1 + f_2 e_2$ et $h = h_1 e_1 + h_2 e_2$.
En projetant sur la base (e_1, e_2), on obtient le sytème

$$h_1(\overline{A}x) - \lambda_1 h_1(x) = f_1(x) \tag{108}$$
$$h_2(\overline{A}x) - \lambda_2 h_2(x) = f_2(x) \tag{109}$$

Soit S opérateur défini par : si $g : \mathbb{T} \to \mathbb{T}$ continue, $S(g) = g \circ A$. S est un opérateur linéaire de $C(\mathbb{T}, \mathbb{T})$ espace des fonctions continues de $\mathbb{T}$ dans $\mathbb{T}$. On munit $\mathcal{L}(C(\mathbb{T}, \mathbb{T}))$ de la norme opérateur. Comme A inversible, S inversible et $\|S\| = \|S^{-1}\| = 1$, et le système précédent devient

$$(S - \lambda_i Id)h_i = f_i, \quad i = 1, 2.$$

Pour $i = 1$, on a $(S - \lambda_1 Id)$ inversible car $|1/\lambda_1| < 1$ et $\|S\| = 1$, et,

$$(S - \lambda_1 Id)^{-1} = -\frac{1}{\lambda_1} \sum_{k \geqslant 0} (\frac{1}{\lambda_1} S)^k$$

et comme $|\lambda_1 \lambda_2| = |detA| = 1$, en particulier,

$$\|(S - \lambda_1 Id)^{-1}\| \leqslant \frac{|\lambda_2|}{1 - |\lambda_2|}.$$

Et de même,

$$(S - \lambda_2 Id)^{-1} = S^{-1}(Id - \lambda_2 S^{-1} = S^{-1} \sum_{k \geqslant 0} (\lambda_2 S)^k$$

et

$$\|(S - \lambda_1 Id)^{-1}\| \leqslant \frac{1}{1 - |\lambda_2|}.$$

Donc finalement L est inversible, et de plus $\|L^{-1}\| \leqslant \frac{1}{1-|\lambda_2|}$.

On revient maintenant à l'équation de départ et on pose

$$\Phi(h)(x) = f(x + h(x)) - f(x).$$

Et l'équation fonctionnelle s'écrit

$$Lh = \Phi(h) + f$$

c'est-à-dire

$$h = L^{-1} \circ \Phi(h) + L^{-1} f.$$

Pour la résoudre, montrons que si la norme C^1 de $f = B - A$ est assez petite alors $L^{-1} \circ \Phi$ est contractante.
Si h^1 h^2 sont des difféomorphismes, on a

$$\begin{aligned} L^{-1}\Phi h^1 - L^{-1}\Phi h^2\| &\leqslant \frac{\|\Phi h^1 - \Phi h^2\|}{1 - |\lambda_2|} && (110)\\ &= \frac{\max_{x\in\mathbb{T}} \|f(x + h^1(x)) - f(x + h^2(x))|}{1 - |\lambda_2|} && (111)\\ &\leqslant \frac{\|f\|_{C^1}\|h^1 - h^2\|}{1 - |\lambda_2|} && (112) \end{aligned}$$

Donc si $\|f\|_{C^1} < 1 - |\lambda_2|$, $L^{-1} \circ \Phi$ est contractante et donc $h \mapsto L^{-1} \circ \Phi(h) + L^{-1} f$ aussi.
On a construit $\overline{H} : \mathbb{R}^2 \to \mathbb{R}^2$ tel que $\overline{B}\,\overline{H} = \overline{H}\,\overline{A}$ et H continue. On s'intéresse maintenant à $H : \mathbb{T} \to \mathbb{T}$.

Il reste à montrer que le H est un homéomorphisme.

Montrons que H est injective. Soit $x, y \in \mathbb{T}$ tels que $H(x) = H(y)$. Comme $\overline{B}\,\overline{H} = \overline{H}\,\overline{A}$, on a $\overline{H}\,\overline{A}\overline{x} = \overline{H}\,\overline{A}\overline{y}$, et par suite $\overline{H}\,\overline{A}^n\overline{x} = \overline{H}\,\overline{A}^n\overline{y}$.
Or si $x \neq y$, la distance entre $\overline{A}^n\overline{y}$ et $\overline{A}^n\overline{x}$ diverge soit en ∞, soit en $-\infty$, car si on considère la norme (elles sont toutes équivalentes) $|ae_1 + be_2| = \sqrt{a^2 + b^2}$, alors

$$|\overline{A}^n\overline{x} - \overline{A}^n\overline{y}|^2 = (x_1^2 - y_1^2)\lambda_1^{2n} + (x_2^2 - y_2^2)\lambda_2^{2n}.$$

Mais h est bornée, et on a

$$\overline{A}^n\overline{x} - \overline{A}^n\overline{y} = h(\overline{A}^n\overline{y}) - h(\overline{A}^n\overline{x})$$

qui conduit a une contradiction. Donc $x = y$.

Montrons la surjectivité. Comme h est bornée, il existe un rayon R assez grand tel que l'image par H du cercle de rayon R et de centre 0 contienne le cercle de rayon 2 et de centre 0, donc en particulier le carré $[0, 1[\times[0, 1[$. Comme tout élément de $\mathbb{T}$ peut-être représenté par un élément de ce carré, H est bien surjective. Enfin, comme T est compact, H^{-1} est un homéomorphisme. □

4.3 Le théorème de Kac

Théorème 14 (Théorème de Kac)**.** *Si T est un homéomorphisme d'un espace métrique compact de mesure finie (X, μ) et que T préserve μ et est μ-ergodique, si A est μ-mesurable de mesure non nulle, alors, en notant pour $a \in A$, $u(a)$ le temps de retour dans A, on a*

$$\frac{1}{\mu(A)} \iint u\, d\mu = \frac{\mu(\mathbb{T})}{\mu(A)}$$

Démonstration. On écrit $A_n := u^{-1}(\{n\})$. $\mu(A) > 0$, et T bijective préserve μ, donc $\mu(A_\infty) = 0$

On note C_n l'orbite de A_n. Notons que T est bijective, donc les C_n sont disjoints. On note de plus que $\bigsqcup_{n=1}^{\infty} C_n$ est T-invariant et contient A, et est donc de mesure pleine.

On a alors :

$$\int_A u \mathrm{d}\mu = \sum_{n=1}^{\infty} n\mu(A_n) \tag{113}$$

$$(\text{T préserve } \mu) = \sum_{n=1}^{\infty} \mu(C_n) \tag{114}$$

$$(\text{T est } \mu\text{-ergodique}) = \mu(X) \tag{115}$$

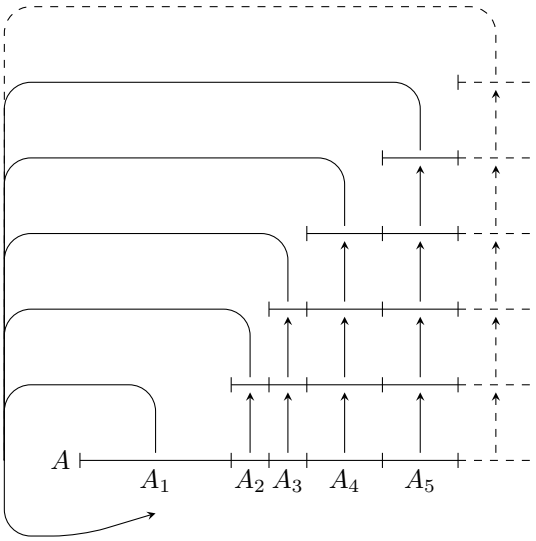

FIGURE 12 – Illustration de la preuve du théorème de Kac.

□

Remerciements

Nous tenons à remercier Serge Cantat de nous avoir fait découvrir ces belles mathématiques en s'adaptant à nos goûts.

Nous voudrions aussi remercier Paul Galvan pour sa relecture attentive.

Références

[AE80] Vladimir Igorevich Arnold and Djilali Embarek. *Chapitres supplémentaires de la théorie des équations différentielles ordinaires*. Mir Moscow, 1980.

[BB05] Vladimir Batagelj and Ulrik Brandes. Efficient generation of large random networks. *Physical Review E*, 71(3) :036113, 2005.

[Bol01] Béla Bollobás. *Random graphs*, volume 73. Cambridge university press, 2001.

[CK10] Xavier Caruso and Igor Kortchemski. Statistiques du nombre de cycles d'une permutation. 2010.

[Éti] GHYS Étienne. Variations autour du théoreme de récurrence de Poincaré.

[Had96] J. Hadamard. Sur la distribution des zéros de la fonction $\zeta(s)$ et ses conséquences arithmétiques. *Bulletin de la Société Mathématique de France*, 24 :199–220, 1896.

[HS00] Marc Hindry and Joseph H Silverman. *Diophantine geometry : an introduction*, volume 201. Springer, 2000.

[LI98] Oscar E Lanford III. Informal remarks on the orbit structure of discrete approximations to chaotic maps. *Experimental Mathematics*, 7(4) :317–324, 1998.

[MRS96] M Ram Murty, Michael Rosen, and Joseph H Silverman. Variations on a theme of Romanoff. *International Journal of Mathematics*, 7(03) :373–391, 1996.

[Pet89] Karl E Petersen. *Ergodic theory*, volume 2. Cambridge University Press, 1989.

[Sil08] Joseph H Silverman. Variation of periods modulo p in arithmetic dynamics. *New York J. Math*, 14 :601–616, 2008.

[Ten95] Gérald Tenenbaum. Introduction à la théorie analytique et probabiliste des nombres. *Cours spécialisés*, 1 :473, 1995.