Abstract.

1. Entropie et information

1.1. Introduction. On cherche un élément x dans un ensemble fini E de cardinal N. Supposons que l'on puisse déterminer si x possède ou non une certaine caractéristique. Cela revient à déterminer si x appartient ou non à la partie A des éléments de E possèdant cette caractéristique. Quelle "quantité" d'information obtient-on ainsi ?
Si $x \in A$ et que la partie A est petite par rapport à E alors la quantité d'information est importante, au contraire si la partie A est grande E, l'information est faible. De même, si $x \in A^C$ la quantité d'information obtenue est une fonction décroissante de la taille de A^C. Pour définir la quantité d'information, on peut utiliser une fonction décroissante $f : [0,1] \to \mathbf{R}^+$ et poser pour chaque x, $I_A(x) = f(\frac{\operatorname{card} A}{\operatorname{card} E})$ si $x \in A$ et $I_A(x) = f(\frac{\operatorname{card} A^C}{\operatorname{card} E})$ si $x \in A^C$. On obtient ainsi une fonction définie sur E :

$$I_A(x) = f(\frac{\operatorname{card} A}{\operatorname{card} E})\, 1_A(x) + f(\frac{\operatorname{card} A^C}{\operatorname{card} E})\, 1_{A^c}(x).$$

Si $A = E$, l'information est nulle on peut donc imposer à f la condition $f(1) = 0$. La caractéristique utilisée pour classer les éléments de E peut avoir plus de 2 modalités : l'ensemble E est partagé en catégories $C_1, ..., C_k$, et pour chaque élément on peut déterminer la catégorie C_i à laquelle il appartient. De la même manière, la fonction f permet de quantifier l'information apportée par la connaissance de la catégorie à laquelle appartient un élément :

$$I(x) = \sum_{i=1}^{k} f(\frac{\operatorname{card} C_i}{\operatorname{card} E})\, 1_{C_i}(x).$$

Les catégories $C_1, ..., C_k$ définissent une partition $\mathcal{C} = \{C_1, ..., C_k\}$ de E et on note $I_{\mathcal{C}}$ l'application que l'on vient de définir. Les conditions imposées à f sont $f(1) = 0$ et sa décroissance. Ajoutons deux autres contraintes naturelles. D'une part, on souhaite choisir une fonction f indépendante de E. D'autre part, on voudrait que la quantité d'information soit additive. Plus précisément, supposons que l'on ait deux renseignements sur les éléments de E. Cela revient à disposer de deux partitions de E, $\mathcal{C} = \{C_1, ..., C_k\}$ et $\mathcal{D} = \{D_1, ..., D_l\}$. Ces deux partitions nous en donne une nouvelle

$$\mathcal{C} \vee \mathcal{D} = \{C_i \cap D_j : i \in \{1, ..., k\},\ j \in \{1, ..., l\}\}$$

qui correspond à la connaissances simultanée des deux renseignements. Lorsque ces renseignements sont indépendants, on aimerait avoir la formule

$$I_{\mathcal{C} \vee \mathcal{D}} = I_{\mathcal{C}} + I_{\mathcal{D}}.$$

Que signifie l'indépendance de $\mathcal{C}$ et $\mathcal{D}$? Les deux caractéristiques $\mathcal{C}$ et $\mathcal{D}$ sont indépendantes si pour chaque j, la proportion des éléments possédant le caractère D_j parmi ceux possédant le caractère C_i ne dépend pas de i. Par conséquent, l'indépendance signifie que pour chaque j, on a

$$\forall i,\ \frac{\operatorname{card} D_j \cap C_i}{\operatorname{card} C_i} = \frac{\operatorname{card} D_j}{\operatorname{card} E}$$

ou encore
$$\frac{\operatorname{card} D_j \cap C_i}{\operatorname{card} E} = \frac{\operatorname{card} D_j}{\operatorname{card} E} \times \frac{\operatorname{card} C_i}{\operatorname{card} E}.$$
La relation souhaitée est
$$\forall x \in E,\ \sum_{i,j} f(\frac{\operatorname{card} C_i \cap D_j}{\operatorname{card} E})\, 1_{C_i \cap D_j}(x) = \sum_{i=1}^{k} f(\frac{\operatorname{card} C_i}{\operatorname{card} E})\, 1_{C_i}(x) + \sum_{j=1}^{j} f(\frac{\operatorname{card} D_j}{\operatorname{card} E})\, 1_{D_j}(x).$$
Pour les x de $C_i \cap D_j$, cela donne
$$f(\frac{\operatorname{card} C_i \cap D_j}{\operatorname{card} E}) = f(\frac{\operatorname{card} C_i}{\operatorname{card} E}) + f(\frac{\operatorname{card} D_j}{\operatorname{card} E})$$
ou encore
$$f(\frac{\operatorname{card} D_j}{\operatorname{card} E} \times \frac{\operatorname{card} C_i}{\operatorname{card} E}) = f(\frac{\operatorname{card} C_i}{\operatorname{card} E}) + f(\frac{\operatorname{card} D_j}{\operatorname{card} E})$$
Cette dernière relation est vraie avec la fonction $f(t) = -\ln t$. N'importe quel multiple positif f convient encore. Grâce à la décroissance, on peut montrer que ce sont les seuls choix possibles. Pour la suite nous fixons le choix de f : $f(t) = -\ln t$. Le fait que $\ln 0$ ne soit pas définie n'est pas gênant : si $\operatorname{card} C_i = 0$ alors la fonction 1_{C_i} est nulle.

Supposons, maintenant que l'on apprenne d'abord $\mathcal{C}$ puis $\mathcal{D}$, on aimerait mesurer le supplément d'information apporté $I_{\mathcal{D}|\mathcal{C}}$ par $\mathcal{D}$ et on devrait avoir
$$I_{\mathcal{C}\vee\mathcal{D}} = I_{\mathcal{C}} + I_{\mathcal{D}|\mathcal{C}}$$
ce qui donne comme définition de l'information apportée par $\mathcal{D}$ sachant $\mathcal{C}$:
$$I_{\mathcal{D}|\mathcal{C}} = I_{\mathcal{C}\vee\mathcal{D}} - I_{\mathcal{C}}.$$
La fonction $I_{\mathcal{D}|\mathcal{C}}$ obtenue est-elle positive ? Pour cela il faut vérifier que pour chaque $x \in C_i \cap D_j$,
$$(-\ln \frac{\operatorname{card} C_i \cap D_j}{\operatorname{card} E}) - (-\ln \frac{\operatorname{card} C_i}{\operatorname{card} E}) \geq 0$$
mais cela résulte de la décroissance.

Finalement, l'information moyenne apportée par une partition $\mathcal{C}$ est
$$\frac{1}{\operatorname{card} E} \sum_{x\in E} I_{\mathcal{C}}(x) = \sum_i \frac{\operatorname{card} C_i}{\operatorname{card} E}(-\ln \frac{\operatorname{card} C_i}{\operatorname{card} E})$$
et ce nombre s'appelle l'entropie de la partition $\mathcal{C}$.

1.2. Définitions dans le cas des espaces probabilisés. Un ensemble fini E peut être vu comme un espace probabilisé où chaque partie A a la probabilité $P(A) = \frac{\operatorname{card} A}{\operatorname{card} E}$. Les définitions précédentes se transposent alors aux espaces probabilisés.

Définition 1. *Soit $(\Omega, \mathcal{A})$ un espaces mesurable et $\mathcal{C}$ une partition de Ω. On dit que la partition $\mathcal{C}$ est mesurable si tous les éléments de $\mathcal{C}$ appartiennent à $\mathcal{A}$.*

Définition 2. *Soit* $(\Omega, \mathcal{A}, \mu)$ *un espaces probabilisé et* $\mathcal{C} = \{C_1, ..., C_n\}$ *une partition mesurable finie de* Ω*. La fonction information* $I_{\mathcal{C}}$ *de* $\Omega \to \mathbf{R}^+$ *associée à la partition* $\mathcal{C}$ *est définie presque partout sur* Ω *par*

$$I_{\mathcal{C}}(x) = -\sum_{i=1}^{n} \ln \mu(C_i)\, 1_{C_i}(x).$$

L'entropie de la partition $H_\mu(\mathcal{C})$ *est définie par*

$$H_\mu(\mathcal{C}) = E(I_{\mathcal{C}}).$$

Lorsqu'il n'y pas d'ambiguïté sur la probabilité μ, on note l'entropie de la partition $\mathcal{C}$ par $H(\mathcal{C})$.

En calculant l'espérance, on obtient immédiatement :

Proposition 1. *Soit* $(\Omega, \mathcal{A}, \mu)$ *un espaces probabilisé et* $\mathcal{C} = \{C_1, ..., C_n\}$ *une partition mesurable finie de* Ω*. Avec la convention* $0 \ln 0 = 0$*, on a*

$$H(\mathcal{C}) = -\sum_{i=1}^{n} \mu(C_i) \ln \mu(C_i).$$

Notation. Dans la suite ϕ désigne l'application $\phi : [0,1] \to \mathbf{R}^+$ définie par $\phi(0) = 0$ et $\phi(x) = x \ln x$ si $x \in]0,1]$. Remarquons que ϕ est strictement concave.

Proposition 2. *Soit* $(\Omega, \mathcal{A}, \mu)$ *un espace probabilisé et* $\mathcal{C} = \{C_1, ..., C_n\}$ *une partition mesurable finie de* Ω*. On a* $H(\mathcal{C}) \leq \ln n$ *et* $H(\mathcal{C}) = \ln n$ *ssi* $\mu(C_1) = ... = \mu(C_n) = \frac{1}{n}$.

Dem : exercice.

Exercice. Soit $(\Omega, \mathcal{A}, \mu)$ un espaces probabilisé et $\mathcal{C}$ une partition mesurable finie de Ω. Si pour chaque $C \in \mathcal{C}$ on a $\mu(C) \leq \varepsilon$ alors $H(\mathcal{C}) \geq -\ln \varepsilon$ (intégrer $I_{\mathcal{C}}$)

Définition 3. *Soit* $\mathcal{C} = \{C_1, ..., C_n\}$ *et* $\mathcal{D} = \{D_1, ..., D_m\}$ *deux partitions finies d'un ensemble* E*.*
1. La partition engendrée par $\mathcal{C}$ *et* $\mathcal{D}$ *est la partition* $\mathcal{C} \vee \mathcal{D} = C_i \cap D_j : i = 1, ..., n,\ j = 1, ..., m\}$*.*
2. Lorsque tout les éléments de $\mathcal{C}$ *sont des réunions d'éléments de* $\mathcal{D}$ *on dit que* $\mathcal{D}$ *est plus fine que* $\mathcal{C}$ *et on écrit* $\mathcal{C} \prec \mathcal{D}$*.*

Définition 4. *Soit* $(\Omega, \mathcal{A}, \mu)$ *un espaces probabilisé et* $\mathcal{C} = \{C_1, ..., C_n\}$, $\mathcal{D} = \{D_1, ..., D_m\}$ *deux partitions mesurables finies de* Ω*. La fonction information conditionnelle de* $\mathcal{D}$ *sachant* $\mathcal{C}$*,* $I_{\mathcal{D}|\mathcal{C}}$ *de* $\Omega \to \mathbf{R}$ *est définie presque partout sur* Ω *par*

$$I_{\mathcal{D}|\mathcal{C}} = I_{\mathcal{D} \vee \mathcal{C}} - I_{\mathcal{C}}.$$

Comme dans l'introduction, on obtient immédiatement la proposition suivante.

Proposition 3. *Soit* $(\Omega, \mathcal{A}, \mu)$ *un espaces probabilisé et* $\mathcal{C} = \{C_1, ..., C_n\}$, $\mathcal{D} = \{D_1, ..., D_m\}$ *deux partitions mesurables finies de* Ω*.*
1. Pour presque tout x*,*

$$I_{\mathcal{D}|\mathcal{C}}(x) = -\sum_{i=1}^{n} \sum_{j=1}^{m} \ln \frac{\mu(C_i \cap D_j)}{\mu(C_i)}\, 1_{C_i \cap D_j}(x)$$

2. La fonction information conditionnelle de $\mathcal{D}$ *sachant* $\mathcal{C}$*,* $I_{\mathcal{D}|\mathcal{C}}$ *est positive.*

Définition 5. *Soit* $(\Omega, \mathcal{A}, \mu)$ *un espaces probabilisé et* $\mathcal{C} = \{C_1, ..., C_n\}$, $\mathcal{D} = \{D_1, ..., D_m\}$ *deux partitions mesurables finies de* Ω*. L'entropie de la partition* $\mathcal{D}$ *sachant* $\mathcal{C}$ *est*

$$H(\mathcal{D}|\mathcal{C}) = E(I_{\mathcal{C}|\mathcal{D}})$$

Un simple calcul donne :

Proposition 4. *Soit* $(\Omega, \mathcal{A}, \mu)$ *un espaces probabilisé et* $\mathcal{C} = \{C_1, ..., C_n\}$, $\mathcal{D} = \{D_1, ..., D_m\}$ *deux partitions mesurables finies de* Ω*. L'entropie de la partition* $\mathcal{D}$ *sachant* $\mathcal{C}$ *vaut*

$$\begin{aligned} H(\mathcal{D}|\mathcal{C}) &= -\sum_{i=1}^{n}\sum_{j=1}^{m} \mu(C_i \cap D_j) \ln \frac{\mu(C_i \cap D_j)}{\mu(C_i)} \\ &= -\sum_{i=1}^{n} \mu(C_i) \sum_{j=1}^{m} \mu(D_j|C_i) \ln \mu(D_j|C_i) \end{aligned}$$

(avec la convention $0 \times \ln \frac{0}{0} = 0$*).*

Exercice. Soit $(\Omega, \mathcal{A}, \mu)$ un espaces probabilisé et $\mathcal{C} = \{C_1, ..., C_n\}$, $\mathcal{D} = \{D_1, ..., D_m\}$ deux partitions mesurables finies de Ω. Appelons pour chaque $C \in \mathcal{C}$, $N(C)$ le nombre des éléments D de $\mathcal{D}$ tels que $D \cap C \neq \emptyset$. Montrer que $H(\mathcal{D}|\mathcal{C}) \leq \max_{C \in \mathcal{C}} \ln N(C)$. (Pour chaque $C \in \mathcal{C}$, l'entropie calculée avec la probabilité conditionnelle $\mu(.|C)$ est $\leq \ln N(C)$).

En remarquant que $\frac{\mu(C_i \cap D_j)}{\mu(C_i)}$ est égale à la probabilité conditionnelle $\mu(D_j|C_i)$, on obtient

$$I_{\mathcal{D}|\mathcal{C}}(x) = -\sum_{i=1}^{n}\sum_{j=1}^{m} \ln \mu(D_j|C_i)\, 1_{C_i \cap D_j}(x)$$

et

$$H(\mathcal{D}|\mathcal{C}) = -\sum_{i=1}^{n}\sum_{j=1}^{m} \mu(C_i \cap D_j) \ln \mu(D_j|C_i).$$

Les expressions de l'information conditionnelle et de l'entropie conditionnelle suggèrent de remplacer la partition $\mathcal{C}$ dans les définitions par une sous tribu $\mathcal{F}$ de $\mathcal{A}$. Regardons d'abord le cas où $\mathcal{F}$ est la tribu engendrée par $\mathcal{C}$. La fonction $\ln(\mu(D_j|\mathcal{F}))$ est égale à $\sum_i \ln(\mu(D_j|C_i))\, 1_{C_i}$ donc

$$\begin{aligned} I_{\mathcal{D}|\mathcal{C}}(x) &= -\sum_{i=1}^{n}\sum_{j=1}^{m} \ln \mu(D_j|C_i)\, 1_{C_i \cap D_j}(x) \\ &= \sum_{j=1}^{m} 1_{D_j}(x) \sum_{i=1}^{n} \ln \mu(D_j|C_i)\, 1_{C_i}(x) \\ &= \sum_{j=1}^{m} 1_{D_j}(x) \ln(\mu(D_j|\mathcal{F})(x)). \end{aligned}$$

D'où la définition :

Définition 6. *Soit* $(\Omega, \mathcal{A}, \mu)$ *un espaces probabilisé,* $\mathcal{D} = \{D_1, ..., D_m\}$*une partition mesurable finie de* Ω *et* $\mathcal{F}$ *une sous tribu de* $\mathcal{A}$*.*

1. L'information conditionnelle de $\mathcal{D}$ sachant $\mathcal{F}$ est la fonction définie presque sûrement par

$$I_{\mathcal{D}|\mathcal{F}}(x) = -\sum_{j=1}^{m} \ln \mu(D_j|\mathcal{F})\, 1_{D_j}(x).$$

2. L'entropie conditionnelle de $\mathcal{D}$ sachant $\mathcal{F}$ est le nombre

$$H(\mathcal{D}|\mathcal{F}) = E(I_{\mathcal{D}|\mathcal{F}}(x)).$$

Remarque. Lorsque $\mathcal{F} = \{\emptyset, X\}$ est la tribu triviale, on a $I_{\mathcal{D}|\mathcal{F}} = I_{\mathcal{D}}$ et $H(\mathcal{D}|\mathcal{F}) = H(\mathcal{D})$.

Proposition 5. *Soit $(\Omega, \mathcal{A}, \mu)$ un espaces probabilisé, $\mathcal{D} = \{D_1, ..., D_m\}$une partition mesurable finie de Ω et $\mathcal{F}$ une sous tribu de $\mathcal{A}$. On a*

$$H(\mathcal{D}|\mathcal{F}) = E(-\sum_{j=1}^{m} \ln \mu(D_j|\mathcal{F})\, \mu(D_j|\mathcal{F}))$$

Dem. On a

$$\begin{aligned} H(\mathcal{D}|\mathcal{F}) &= E(-\sum_{j=1}^{m} \ln \mu(D_j|\mathcal{F})\, 1_{D_j}) \\ &= E(E(-\sum_{j=1}^{m} \ln \mu(D_j|\mathcal{F})\, 1_{D_j}|\mathcal{F})) \\ &= E(-\sum_{j=1}^{m} \ln \mu(D_j|\mathcal{F})\, E(1_{D_j}|\mathcal{F})) \\ &= E(-\sum_{j=1}^{m} \ln \mu(D_j|\mathcal{F})\, \mu(D_j|\mathcal{F})) \end{aligned}$$

(ce qui donne une définition alternative de l'entropie conditionnelle).

Notations. On note $\sigma(\mathcal{C})$ la tribu engendrée par un ensemble de partie et $\mathcal{F}_1 \vee \mathcal{F}_2 = \sigma(\mathcal{F}_1 \cup \mathcal{F}_2)$.

Proposition 6. *Soit $(\Omega, \mathcal{A}, \mu)$ un espaces probabilisé, $\mathcal{C} = \{C_1, ..., C_n\}$ et $\mathcal{D} = \{D_1, ..., D_m\}$ deux partitions mesurables finies de Ω et $\mathcal{F}$ une sous tribu de $\mathcal{A}$. On a presque sûrement*

$$I_{\mathcal{D}\vee\mathcal{C}|\mathcal{F}} = I_{\mathcal{C}|\mathcal{F}} + I_{\mathcal{D}|\mathcal{C}\vee\mathcal{F}}$$

et

$$H(\mathcal{D} \vee \mathcal{C}|\mathcal{F}) = H(\mathcal{C}|\mathcal{F}) + H(\mathcal{D}|\mathcal{C} \vee \mathcal{F}).$$

Dem. Montrons que pour tout A mesurable, on a presque sûrement $\mu(A|\mathcal{C} \vee \mathcal{F}) = \sum_{i=1}^{n} \frac{\mu(A\cap C_i|\mathcal{F})}{\mu(C_i|\mathcal{F})} 1_{C_i}$. Comme le deuxième membre de l'égalité est $\mathcal{C} \vee \mathcal{F}$ mesurable, il suffit de vérifier que pour toute partie $B \in \mathcal{C} \vee \mathcal{F}$ on a

$$\mu(A \cap B) = E(\sum_{i=1}^{n} \frac{\mu(A \cap C_i|\mathcal{F})}{\mu(C_i|\mathcal{F})} 1_{C_i} 1_B).$$

Mais la tribu $\mathcal{C} \vee \mathcal{F}$ est engendrée par les parties de la forme $C_k \cap F$ où $k \in \{1, ..., n\}$ et $F \in \mathcal{F}$, qui forment un ensemble de parties stable par intersection finie, donc la vérification se réduit au partie $B = C_k \cap F$. Cette formule correspond à la formule élémentaire

$$\mu(A \cap C \cap F) = \frac{\mu(A \cap C|F)}{\mu(C|F)} \mu(C \cap F).$$

On a

$$\begin{aligned} E(\sum_{i=1}^{n} \frac{\mu(A \cap C_i|\mathcal{F})}{\mu(C_i|\mathcal{F})} 1_{C_i} 1_B) &= E(\sum_{i=1}^{n} \frac{\mu(A \cap C_i|\mathcal{F})}{\mu(C_i|\mathcal{F})} 1_{C_i} 1_{C_k} 1_F) \\ &= E(\frac{\mu(A \cap C_k|\mathcal{F})}{\mu(C_k|\mathcal{F})} 1_{C_k} 1_F) \end{aligned}$$

or presque sûrement, $E(\frac{\mu(A \cap C_k|\mathcal{F})}{\mu(C_k|\mathcal{F})} 1_{C_k}|\mathcal{F}) = \frac{\mu(A \cap C_k|\mathcal{F})}{\mu(C_k|\mathcal{F})} E(1_{C_k}|\mathcal{F}) = \mu(A \cap C_k|\mathcal{F})$, donc

$$\begin{aligned} E(\frac{\mu(A \cap C_k|\mathcal{F})}{\mu(C_k|\mathcal{F})} 1_{C_k} 1_F) &= E(E(\frac{\mu(A \cap C_k|\mathcal{F})}{\mu(C_k|\mathcal{F})} 1_{C_k}|\mathcal{F}) 1_F) \\ &= E(\mu(A \cap C_k|\mathcal{F}) 1_F) \\ &= \mu(A \cap C_k \cap F). \end{aligned}$$

Finalement, pour chaque j, $\ln \mu(D_j|\mathcal{C} \vee \mathcal{F}) = \ln(\sum_{i=1}^{n} \frac{\mu(D_j \cap C_i|\mathcal{F})}{\mu(C_i|\mathcal{F})} 1_{C_i}) = \sum_{i=1}^{n} \ln(\frac{\mu(D_j \cap C_i|\mathcal{F})}{\mu(C_i|\mathcal{F})})\, 1_{C_i}$ donc

$$\begin{aligned} I_{\mathcal{D}|\mathcal{C} \vee \mathcal{F}} &= -\sum_{j=1}^{m} \ln \mu(D_j|\mathcal{C} \vee \mathcal{F})\, 1_{D_j} = -\sum_{j=1}^{m} \sum_{i=1}^{n} \ln(\frac{\mu(D_j \cap C_i|\mathcal{F})}{\mu(C_i|\mathcal{F})})\, 1_{C_i} 1_{D_j} \\ &= -\sum_{j=1}^{m} \sum_{i=1}^{n} \ln(\mu(D_j \cap C_j|\mathcal{F})\, 1_{C_i} 1_{D_j} + \sum_{j=1}^{m} \sum_{i=1}^{n} \ln \mu(C_j|\mathcal{F})\, 1_{C_i} 1_{D_j} \\ &= I_{\mathcal{D} \vee \mathcal{C}|\mathcal{F}} - I_{\mathcal{C}|\mathcal{F}}. \square \end{aligned}$$

Proposition 7. *Soit* $(\Omega, \mathcal{A}, \mu)$ *un espaces probabilisé,* $\mathcal{C} = \{C_1, ..., C_n\}$ *une partition mesurable finie de* Ω *et* $\mathcal{F}_1$, $\mathcal{F}_2$ *deux sous tribus de* $\mathcal{A}$*. Si* $\mathcal{F}_1 \subset \mathcal{F}_2$ *alors* $H(\mathcal{C}|\mathcal{F}_1) \geq H(\mathcal{C}|\mathcal{F}_2)$.

Dem. Comme $\mathcal{F}_1 \subset \mathcal{F}_2$,

$$\begin{aligned} H(\mathcal{C}|\mathcal{F}_1) &= E(\sum_{i=1}^{n} \phi(E(1_{C_i}|\mathcal{F}_1))) \\ &= E(\sum_{i=1}^{n} \phi(E(E(1_{C_i}|\mathcal{F}_2)|\mathcal{F}_1)) \\ &= E(\sum_{i=1}^{n} \phi(E(f_i|\mathcal{F}_1)) \end{aligned}$$

avec $f_i = E(1_{C_i}|\mathcal{F}_2)$. La fonction $\phi(t) = -t \ln t$ est concave sur $[0, 1]$ donc d'après l'inégalité de Jensen pour les espérances conditionnelles,

$$\phi(E(f_i|\mathcal{F}_1)) \geq E(\phi(f_i)|\mathcal{F}_1)$$

presque sûrement, donc

$$H(\mathcal{C}|\mathcal{F}_1) \geq E(\sum_{i=1}^{n} E(\phi(f_i)|\mathcal{F}_1)) = E(\sum_{i=1}^{n} \phi(f_i)) = H(\mathcal{C}|\mathcal{F}_2). \square$$

Proposition 8. *Soit $(\Omega, \mathcal{A}, \mu)$ un espaces probabilisé, $\mathcal{C} = \{C_1, ..., C_n\}$ et $\mathcal{D} = \{D_1, ..., D_m\}$ deux partitions mesurables finies de Ω et $\mathcal{F}$ une sous tribu de $\mathcal{A}$.*
1. On a $H(\mathcal{C}) \geq H(\mathcal{C}|\mathcal{F})$.
2. Si $\mathcal{C} \prec \mathcal{D}$ alors $H(\mathcal{C}|\mathcal{F}) \leq H(\mathcal{D}|\mathcal{F})$.
3. $H(\mathcal{D} \vee \mathcal{C}|\mathcal{F}) \leq H(\mathcal{C}|\mathcal{F}) + H(\mathcal{D}|\mathcal{F})$.
4. $H(\mathcal{D} \vee \mathcal{C}) \leq H(\mathcal{C}) + H(\mathcal{D})$.
5. Si $T : \Omega \to \Omega$ est mesurable et conserve μ, alors $H(T^{-1}\mathcal{C}|T^{-1}\mathcal{F}) = H(\mathcal{C}|\mathcal{F})$.

Dem. 1. On utilise la proposition précédente avec $\mathcal{F}_1 = \{\emptyset, X\}$ et $\mathcal{F}_2 = \mathcal{F}$.
2. Par hypothèse $\mathcal{C} \vee \mathcal{D} = \mathcal{D}$ et comme $I_{\mathcal{C}\vee\mathcal{D}|\mathcal{F}} = I_{\mathcal{C}|\mathcal{F}} + I_{\mathcal{D}|\mathcal{F}\vee\mathcal{C}}$, on obtient $H(\mathcal{D}|\mathcal{F}) = H(\mathcal{C}|\mathcal{F}) + H(\mathcal{D}|\mathcal{F} \vee \mathcal{C}) \geq H(\mathcal{C}|\mathcal{F})$.
3. D'après les propositions précédentes , $H(\mathcal{D}\vee\mathcal{C}|\mathcal{F}) = H(\mathcal{C}|\mathcal{F}) + H(\mathcal{D}|\mathcal{C} \vee \mathcal{F})$ et $H(\mathcal{D}|\mathcal{C} \vee \mathcal{F}) \leq H(\mathcal{D}|\mathcal{F})$ et le résultat en découle.
4. On utilise 2 avec $\mathcal{F} = \{\emptyset, X\}$.
5. On a $E(1_{C_i}|\mathcal{F}) \circ T = E(1_{T^{-1}(C_i)}|T^{-1}\mathcal{F})$ car $E(1_{C_i}|\mathcal{F}) \circ T$ est $T^{-1}\mathcal{F}$ mesurable et pour tout $A \in \mathcal{F}$,

$$\begin{aligned}
\int_{T^{-1}A} E(1_{C_i}|\mathcal{F}) \circ T \, d\mu &= \int_X 1_A \circ T \, E(1_{C_i}|\mathcal{F}) \circ T \, d\mu \\
&= \int_X 1_A E(1_{C_i}|\mathcal{F}) \, d\mu = \int_X 1_A 1_{C_i} d\mu \\
&= \int_X 1_{T^{-1}A} 1_{T^{-1}C_i} d\mu.
\end{aligned}$$

En intégrant on obtient le résultat désiré. $\square$

Proposition 9. *Soit $(\Omega, \mathcal{A}, \mu)$ un espaces probabilisé, $\mathcal{C} = \{C_1, ..., C_n\}$ et $\mathcal{D} = \{D_1, ..., D_m\}$ deux partitions mesurables finies de Ω et $\mathcal{F}$ une sous tribu de $\mathcal{A}$.*
1. $H(\mathcal{C}|\mathcal{F}) = 0$ ssi $\mathcal{C} \subset \mathcal{F}$ aux ensembles de mesure nulle près, i.e. pour chaque i il existe $A_i \in \mathcal{F}$ tel que $\mu(A_i \Delta C_i) = 0$.
2. $H(\mathcal{C}|\mathcal{D}) = 0$ ssi $\mathcal{C} \prec \mathcal{D}$ aux ensembles de mesure nulle près.

Dem. 2 est une conséquence de 1. Si $\mathcal{C} \subset \mathcal{F}$ on a presque sûrement $\mu(C_i|\mathcal{F}) = E(1_{C_i}|\mathcal{F}) = 1_{C_i}$ donc $\ln \mu(C_i|\mathcal{F}) = 0$ sur C_i et $H(\mathcal{C}|\mathcal{F}) = 0$. Réciproquement, supposons $H(\mathcal{C}|\mathcal{F}) = 0$. Pour chaque i, la fonction $\phi(\mu(C_i|\mathcal{F}))$ doit être presque sûrement nulle. Donc $\mu(C_i|\mathcal{F}) = 0$ ou 1 presque sûrement et par conséquent, il existe $A \in \mathcal{F}$ tel que $\mu(C_i|\mathcal{F}) = 1_A$ presque sûrement. Par définition de la probabilité conditionnelle on doit avoir $E(1_A 1_{C_i}) = E(1_A 1_A)$ donc $A \subset C_i$ à un ensemble de mesure nulle près mais on doit aussi avoir $E(1_A) = E(1_{C_i})$ donc A et C_i diffèrent d'un ensemble de mesure nulle. $\square$

Définition 7. *Soit $(\Omega, \mathcal{A}, \mu)$ un espaces probabilisé, $\mathcal{C} = \{C_1, ..., C_n\}$ et $\mathcal{D} = \{D_1, ..., D_m\}$ deux partitions mesurables finies de Ω. On dit que $\mathcal{C}$ et $\mathcal{D}$ sont indépendantes si pour tout les couples (i, j) on a*

$$\mu(C_i \cap D_j) = \mu(C_i) \cap \mu(D_j).$$

Remarque. Les deux partitions $\mathcal{C}$ et $\mathcal{D}$ sont indépendantes ssi les tribus engendrées le sont.

Proposition 10. *Soit $(\Omega, \mathcal{A}, \mu)$ un espaces probabilisé, $\mathcal{C} = \{C_1, ..., C_n\}$ et $\mathcal{D} = \{D_1, ..., D_m\}$ deux partitions mesurables finies de Ω. Les partitions $\mathcal{C}$ et $\mathcal{D}$ sont indépendantes ssi $H(\mathcal{C} \vee \mathcal{D}) = H(\mathcal{C}) + H(\mathcal{D})$.*

Dem. Si les partitions sont indépendantes on a évidemment $H(\mathcal{C}\vee\mathcal{D}) = H(\mathcal{C})+H(\mathcal{D})$. Réciproquement supposons $H(\mathcal{C} \vee \mathcal{D}) = H(\mathcal{C}) + H(\mathcal{D})$. On a alors $H(\mathcal{D}|\mathcal{C}) = H(\mathcal{D})$ c'est-à-dire

$$\begin{aligned}
\sum_{j=1}^{m} \phi(\mu(D_j)) &= -\sum_{j=1}^{m}\sum_{i=1}^{n} \mu(C_i \cap D_j) \ln \mu(D_j|C_i) \\
&= -\sum_{j=1}^{m}\sum_{i=1}^{n} \mu(C_i)\mu(D_j|C_i) \ln \mu(D_j|C_i) \\
&= \sum_{j=1}^{m}\sum_{i=1}^{n} \mu(C_i)\phi(\mu(D_j|C_i)).
\end{aligned}$$

Or la fonction ϕ est concave donc pour chaque j,

$$\sum_{i=1}^{n} \mu(C_i)\phi(\mu(D_j|C_i)) \leq \phi(\sum_{i=1}^{n} \mu(C_i)\mu(D_j|C_i)) = \phi(\mu(D_j))$$

et par conséquent, $\sum_{i=1}^{n} \mu(C_i)\phi(\mu(D_j|C_i)) = \phi(\mu(D_j))$. Maintenant la stricte concavité de la fonction ϕ entraîne que pour chaque j, les $\mu(D_j|C_i)$ sont indépendants de i. $\square$

Théorème 1. *Soit $(\Omega, \mathcal{A}, \mu)$ un espace probabilisé, $\mathcal{C} = \{C_1, ..., C_n\}$ une partition mesurable finie de Ω et $(\mathcal{F}_m)_{m\geq 1}$ une suite croissante de sous tribus de $\mathcal{A}$ engendrant une sous tribu $\mathcal{F}$. Alors*

$$\lim_{m\to\infty} H(\mathcal{C}|\mathcal{F}_m) = H(\mathcal{C}|\mathcal{F}).$$

Dem. L'application qui à $f \in L^2(\mu)$ associe $E(f|\mathcal{F}_n)$ est la projection orthogonale sur le sous espace H_n de L^2 des fonctions $\mathcal{F}_n$ mesurables. De même, L'application qui à $f \in L^2(\mu)$ associe $E(f|\mathcal{F})$ est la projection orthogonale sur le sous espace H de L^2 des fonctions $\mathcal{F}$ mesurables. La réunion $\cup_{n\geq 1} H_n$ est dense dans H car si $B \in \mathcal{F}$ alors il existe une suite $(B_k)_{k\in\mathbf{N}}$ d'éléments de l'algèbre $\cup_{n\geq 1}\mathcal{F}_n$ qui engendre $\mathcal{F}$, telle que $\lim_{k\to\infty} \|1_{B_k} - 1_B\|_2 = \lim_{k\to\infty} \mu(B_k \Delta B) = 0$ (justification : l'ensemble des parties $B \in \mathcal{A}$ pour lesquelles il existe une suite $(B_k)_k$ d'éléments de $\cup_{n\geq 1}\mathcal{F}_n$ telle que $\lim_{k\to\infty} \mu(B_k \Delta B) = 0$ est une tribu contenant $\cup_{n\geq 1}\mathcal{F}_n$). Pour tout $f \in L^2$, la suite des projections $E(f|\mathcal{F}_n)$ converge donc dans L^2 vers $E(f|\mathcal{F})$. Il suffit maintenant d'utiliser ce résultat avec $E(1_{C_i}|\mathcal{F}_n)$ $i = 1, ..., n$. Comme la convergence L^2 implique la convergence en probabilité, la suite

$$\sum_{i=1}^{n} \phi(E(f|\mathcal{F}_m))$$

converge en probabilité vers la fonction $\sum_{i=1}^{n} \phi(E(f|\mathcal{F}))$. Finalement, comme ces fonctions sont bornées indépendamment de m, la convergence à lieu dans L^1 et donc

$$\lim_{m\to\infty} E(\sum_{i=1}^{n} \phi(E(f|\mathcal{F}_m))) = E(\sum_{i=1}^{n} \phi(E(f|\mathcal{F}))).\square$$

1.3. Entropie d'une application. Soit $\mathcal{C}$ une partition finie d'un ensemble X et $T : X \to X$ une application. Il est facile de voir que deux points x et y de X appartiennent au même élément de la partition $\mathcal{C} \vee T^{-1}\mathcal{C} \vee ... \vee T^{-n}\mathcal{C}$ si pour tout k les éléments $T^k x$ et $T^k y$ appartiennent au même élément de $\mathcal{C}$.

Théorème 2. *Soit $(X, \mathcal{A}, \mu)$ un espace probabilisé, $T : X \to X$ une application mesurable conservant μ et $\mathcal{C}$ une partition mesurable finie de X. Alors la suite*

$$\frac{1}{n}H(\mathcal{C} \vee T^{-1}\mathcal{C} \vee ... \vee T^{-n+1}\mathcal{C})$$

est convergente.

Le théorème découle de l'inégalité

$$\begin{aligned} H(\mathcal{C} \vee T^{-1}\mathcal{C} \vee ... \vee T^{-n-m+1}\mathcal{C}) &\leq H(\mathcal{C} \vee T^{-1}\mathcal{C} \vee ... \vee T^{-n+1}\mathcal{C}) + H(T^{-n}\mathcal{C} \vee ... \vee T^{-n-m+1}\mathcal{C}) \\ &= H(\mathcal{C} \vee T^{-1}\mathcal{C} \vee ... \vee T^{-n+1}\mathcal{C}) + H(T^{-n}(\mathcal{C} \vee T^{-1}\mathcal{C} \vee ... \vee T^{-m+1}\mathcal{C})) \\ &= H(\mathcal{C} \vee T^{-1}\mathcal{C} \vee ... \vee T^{-n+1}\mathcal{C}) + H(\mathcal{C} \vee T^{-1}\mathcal{C} \vee ... \vee T^{-m+1}\mathcal{C}) \end{aligned}$$

et du lemme :

Lemme 1. *Soit $(u_n)_{n\geq 1}$ une suite de réels tels que pour tout m et n, $u_{n+m} \leq u_n + u_m$. Alors la limite de la suite $(\frac{1}{n}u_n)_{n\geq 1}$ existe et vaut $\inf_{n\geq 1} \frac{1}{n}u_n$.*

Dem. Soit n un entier ≥ 1. Pour tout entier m, effectuons la division euclidienne de m par n : $m = q_m n + r_m$ où $r < n$. On a

$$u_m \leq q_m u_n + u_{r_m}$$

donc

$$\frac{1}{m}u_m \leq \frac{q_m}{m}u_n + \frac{1}{m}u_{r_m}$$

et

$$\lim \sup_{m\to\infty} \frac{1}{m}u_m \leq \lim \sup_{m\to\infty} (\frac{q_m}{m}u_n + \frac{1}{m}u_{r_m}) = \frac{1}{n}u_n$$

Ceci étant vrai pour tout $n \geq 1$, on a $\limsup_{m\to\infty} \frac{1}{m}u_m \leq \inf_{n\geq 1} \frac{1}{n}u_n \leq \liminf_{n\to\infty} \frac{1}{n}u_n$ donc $\lim_{n\to\infty} \frac{1}{n}u_n = \inf_{n\geq 1} \frac{1}{n}u_n$. $\square$

Définition 8. *Soit $(X, \mathcal{A}, \mu)$ un espace probabilisé, $T : X \to X$ une application mesurable conservant μ.*
1 Soit $\mathcal{C}$ une partition mesurable finie de Ω. Le nombre $h(T, \mathcal{C}) = \lim_{n\to\infty} \frac{1}{n}H(\mathcal{C} \vee T^{-1}\mathcal{C} \vee ... \vee T^{-n+1}\mathcal{C})$ s'appelle l'entropie de T relativement à la partition $\mathcal{C}$.
2. L'entropie mesurée de T relativement à la probabilité μ est le borne supérieure

$$\sup_{\mathcal{C}} h(T, \mathcal{C})$$

où $\mathcal{C}$ parcourt toutes les partitions finies de X dont tous les éléments sont mesurables. On note l'entropie $h_\mu(T)$ ou $h(T)$ si il n'y a pas d'ambiguïté.

Théorème 3. *L'entropie est invariante par isomorphisme de système dynamique mesuré : Soit $(X, \mathcal{A}, \mu)$ et $(Y, \mathcal{B}, \nu)$ deux espaces probabilisés, T une application mesurable conservant μ et $S : Y \to Y$ une application mesurable conservant ν. Si il existe une bijection $R : X \to Y$ bimesurable telle que $RT = SR$ et telle que l'image de μ par R soit ν, alors $h(T) = h(S)$.*

Dem. Si $\mathcal{C} = (C_1, ..., C_n)$ est une partition finie de X alors $R(\mathcal{C}) = (R(C_1), ..., R(C_n))$ est aussi une partition de Y et on vérifie que $h(T, \mathcal{C}) = h(S, R(\mathcal{C}))$. En passant au sup on obtient $h(T) \leq h(S)$, et comme l'inégalité inverse est une conséquence du même raisonnement, on obtient $h(T) = h(S)$. $\square$

Remarque. Munissons $\mathcal{A}$ (resp. $\mathcal{B}$) de la relation d'équivalence : la différence symétrique de deux parties est de mesure nulle et appelons $\overline{\mathcal{A}}$ (resp. $\overline{\mathcal{B}}$) l'ensemble quotient. On a un théorème plus général, dans lequel la bijection R est remplacée par une bijection de $\overline{\mathcal{A}}$ sur $\overline{\mathcal{B}}$ car R n'intervient que par son action sur les parties.

Exercice. On considère le tore $\mathbf{T}^1 = \mathbf{R}/\mathbf{Z}$ muni de la mesure de Lebesgue, $T : \mathbf{T}^1 \to \mathbf{T}^1$ une translation et $\mathcal{C} = \{I_1, ..., I_m\}$ une partition de $\mathbf{T}^1$ en m intervalles. Montrer que $h(T, \mathcal{C}) = 0$.

1.4. Propriété de l'entropie d'une application.

Proposition 11. *Soit $(X, \mathcal{A}, \mu)$ un espace probabilisé, $T : X \to X$ une application mesurable conservant μ et $\mathcal{C}$, $\mathcal{D}$ deux partitions mesurable finies de X.*
1. $h(T, \mathcal{C}) \leq h(\mathcal{C})$.
2. $h(T, \mathcal{C} \vee \mathcal{D}) \leq h(T, \mathcal{C}) + h(T, \mathcal{D})$.
3. Si $\mathcal{C} \prec \mathcal{D}$ alors $h(T, \mathcal{C}) \leq h(T, \mathcal{D})$.
4. $h(T, \mathcal{C}) \leq h(T, \mathcal{D}) + h(\mathcal{C}|\mathcal{D})$.
5. $h(T, \mathcal{C}) = h(T, T^{-1}\mathcal{C})$.
6. Pour tout $m \geq 0$,, $h(T, \mathcal{C}) = h(T, \bigvee_{i=0}^{m} T^{-i}\mathcal{C})$ et si T est inversible $h(T, \mathcal{C}) = h(T, \bigvee_{i=-m}^{m} T^{-i}\mathcal{C})$.

Dem. 1. On sait que $H(\bigvee_{i=0}^{n-1} T^{-i}\mathcal{C}) \leq \sum_{i=0}^{n-1} H(T^{-i}\mathcal{C}) = nH(\mathcal{C})$ donc $h(T, \mathcal{C}) \leq H(\mathcal{C})$.
2. Pour tout $n \geq 1$, $\bigvee_{i=0}^{n-1} T^{-i}(\mathcal{C} \vee \mathcal{D}) = (\bigvee_{i=0}^{n-1} T^{-i}\mathcal{C}) \vee (\bigvee_{i=0}^{n} T^{-i}\mathcal{D})$ donc

$$H(\bigvee_{i=0}^{n-1} T^{-i}(\mathcal{C} \vee \mathcal{D})) \leq H(\bigvee_{i=0}^{n-1} T^{-i}\mathcal{C}) + H(\bigvee_{i=0}^{n} T^{-i}\mathcal{D})$$

et par passage à la limite $h(T, \mathcal{C} \vee \mathcal{D}) \leq h(T, \mathcal{C}) + h(T, \mathcal{D})$.
3. Si $\mathcal{C} \prec \mathcal{D}$ alors $\bigvee_{i=0}^{n-1} T^{-i}\mathcal{C} \prec \bigvee_{i=0}^{n} T^{-i}\mathcal{D}$. par conséquent, $H(\bigvee_{i=0}^{n-1} T^{-i}\mathcal{C}) \geq H(\bigvee_{i=0}^{n} T^{-i}\mathcal{D})$ et $h(T, \mathcal{C}) \geq h(T, \mathcal{D})$.

4.

$$\begin{aligned} H(\bigvee_{i=0}^{n-1} T^{-i}\mathcal{C}) &\leq H((\bigvee_{i=0}^{n-1} T^{-i}\mathcal{C}) \vee (\bigvee_{i=0}^{n-1} T^{-i}\mathcal{D})) \\ &= H(\bigvee_{i=0}^{n-1} T^{-i}\mathcal{D}) + H(\bigvee_{i=0}^{n-1} T^{-i}\mathcal{C} | \bigvee_{i=0}^{n-1} T^{-i}\mathcal{D}) \\ &\leq H(\bigvee_{i=0}^{n-1} T^{-i}\mathcal{D}) + \sum_{j=0}^{n-1} H(T^{-j}\mathcal{C} | \bigvee_{i=0}^{n-1} T^{-i}\mathcal{D}) \\ &\leq H(\bigvee_{i=0}^{n-1} T^{-i}\mathcal{D}) + \sum_{j=0}^{n-1} H(T^{-j}\mathcal{C} | T^{-j}\mathcal{D}) = H(\bigvee_{i=0}^{n-1} T^{-i}\mathcal{D}) + nH(\mathcal{C}|\mathcal{D}) \end{aligned}$$

d'où le résultat par passage à la limite.
5.

$$H(\bigvee_{i=0}^{n-1} T^{-i}(T^{-1}\mathcal{C}))) = H(T^{-1}\bigvee_{i=0}^{n-1} T^{-i}\mathcal{C}) = H(\bigvee_{i=0}^{n-1} T^{-i}\mathcal{C})$$

d'où $h(T, T^{-1}\mathcal{C}) = h(T, \mathcal{C})$.
6.

$$\lim_{n\to\infty} \frac{1}{n} H(\bigvee_{i=0}^{n-1} T^{-i}(\bigvee_{j=0}^{m} T^{-j}\mathcal{C})) = \lim_{n\to\infty} \frac{n+m}{n} \times \frac{1}{m+n} H(\bigvee_{i=0}^{m+n-1} T^{-i}\mathcal{C})$$

d'où $h(T, (\bigvee_{j=0}^{m} T^{-j}\mathcal{C}) = h(T, \mathcal{C})$. Le cas inversible découle de l'observation $h(T, (\bigvee_{j=-m}^{m} T^{-j}\mathcal{C})) = h(T, T^m(\bigvee_{j=0}^{2m} T^{-j}\mathcal{C}))$. $\square$

Théorème 4. *Soit $(X, \mathcal{A}, \mu)$ un espace probabilisé et $T : X \to X$ une application mesurable conservant μ.*
1. Pour $k \geq 1$, $h(T^k) = kh(T)$.
2. Si T est inversible, alors pour tout entier relatif k, $h(T^k) = |k|\, h(T)$.

Dem. Soit $\mathcal{C}$ une partition de X dont tout les éléments appartiennent à $\mathcal{A}$.

$$\begin{aligned} h(T^k, \bigvee_{i=0}^{k-1} T^{-i}\mathcal{C}) &= \lim_{n\to\infty} \frac{1}{n} H(\bigvee_{i=0}^{n-1} T^{-ik} \bigvee_{j=0}^{k-1} T^{-j}\mathcal{C}) \\ &= \lim_{n\to\infty} \frac{1}{n} H(\bigvee_{i=0}^{nk-1} T^{-i}\mathcal{C}) = k \lim_{n\to\infty} \frac{1}{nk} H(\bigvee_{i=0}^{nk-1} T^{-i}\mathcal{C}) \\ &= kh(T, \mathcal{C}). \end{aligned}$$

Ainsi $h(T^k) = \sup h(T^k, \mathcal{C}) = \sup h(T^k, \bigvee_{i=0}^{k-1} T^{-i}\mathcal{C}) = \sup kh(T, \mathcal{C}) = kh(T)$.
2. D'une part, on a évidement $h(Id) = 0$. D'autre part, $h(T, \mathcal{C}) = h(T^{-1}, \mathcal{C})$ pour toute partition $\mathcal{C}$ car

$$H(\bigvee_{i=0}^{n-1} T^{i}\mathcal{C}) = H(T^{-(n-1)}\bigvee_{i=0}^{n-1} T^{i}\mathcal{C}) = H(\bigvee_{i=0}^{n-1} T^{-i}\mathcal{C}). \square$$

Théorème 5. *Soit $(X,\mathcal{A},\mu)$ un espace probabilisé, $T : X \to X$ une application mesurable conservant μ et $\mathcal{C}$ une partition mesurable finie de X. Appelons $\mathcal{F}$ la tribu engendrée par $\cup_{n\geq 1} \bigvee_{i=1}^{n} T^{-i}\mathcal{C}$. Alors*

$$h(T,\mathcal{C}) = \lim_{n\to\infty} H(\mathcal{C}|\bigvee_{i=1}^{n} T^{-i}\mathcal{C}) = H(\mathcal{C}|\mathcal{F}).$$

Dem. On sait déjà que $\lim_{n\to\infty} H(\mathcal{C}|\bigvee_{i=1}^{n} T^{-i}\mathcal{C}) = H(\mathcal{C}|\mathcal{F})$. Montrons par récurrence que

$$H(\bigvee_{i=0}^{n} T^{-i}\mathcal{C}) = H(\mathcal{C}) + \sum_{j=1}^{n} H(\mathcal{C}|\bigvee_{i=1}^{j} T^{-i}\mathcal{C}).$$

Pour $n = 0$ il n'y a rien à prouver et le passage de n à $n+1$ résulte du calcul

$$\begin{aligned} H(\bigvee_{i=0}^{n+1} T^{-i}\mathcal{C}) &= H(\mathcal{C} \vee \bigvee_{i=1}^{n+1} T^{-i}\mathcal{C}) \\ &= H(\bigvee_{i=1}^{n+1} T^{-i}\mathcal{C}) + H(\mathcal{C}|\bigvee_{i=1}^{n+1} T^{-i}\mathcal{C}) \\ &= H(\bigvee_{i=0}^{n} T^{-i}\mathcal{C}) + H(\mathcal{C}|\bigvee_{i=1}^{n+1} T^{-i}\mathcal{C}). \end{aligned}$$

Maintenant

$$\begin{aligned} H(T,\mathcal{C}) &= \lim_{n\to\infty} \frac{1}{n} H(\bigvee_{i=0}^{n} T^{-i}\mathcal{C}) = \lim_{n\to\infty} \frac{1}{n}(H(\mathcal{C}) + \sum_{j=1}^{n} H(\mathcal{C}|\bigvee_{i=1}^{j} T^{-i}\mathcal{C})) \\ &= \lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} H(\mathcal{C}|\bigvee_{i=1}^{j} T^{-i}\mathcal{C}) \end{aligned}$$

et d'après le théorème de Césaro,

$$h(T,\mathcal{C}) = H(\mathcal{C}|\mathcal{F}). \ \square$$

Exercices. Soit $(X,\mathcal{A},\mu)$ un espace probabilisé et $T : X \to X$ une application mesurable conservant μ. Si $\mathcal{F}$ est une sous-tribu de $\mathcal{A}$, on note $\mathcal{F}_c$ la tribu complétée de $\mathcal{F}$.
1. Soit $\mathcal{C}$ une partition mesurable finie de X. Appelons $\mathcal{F}$ la tribu engendrée par $\bigvee_{i=1}^{\infty} T^{-i}\mathcal{C}$. Montrer que $h(T,\mathcal{C}) = 0$ ssi $\mathcal{C}$ est inclus dans la tribu complétée $\mathcal{F}_c$.
2. Montrer que si $h(T) = 0$ alors $(T^{-1}\mathcal{A})_c = \mathcal{A}_c$.

1.5. Application au codage des textes.

Théorème 6. ***(Shannon-MacMillan-Briemann).*** *Soit $(X,\mathcal{A},\mu)$ un espace probabilisé, $T : X \to X$ une application mesurable conservant μ et ergodique, et $\mathcal{C}$ une partition mesurable finie de X. Pour chaque $x \in X$ et chaque entier $n \geq 1$ notons $C_n(x)$ l'élément de la partition $\bigvee_{i=0}^{n-1} T^{-i}\mathcal{C}$ contenant x. Alors*

$$-\frac{1}{n} \ln \mu(C_n(x)) \to h(T,\mathcal{C})$$

presque sûrement et dans $L^1(\mu)$.

Dem. Petersen page 261.

Corollaire 1. *Soit $(X, \mathcal{F}, \mu)$ un espace probabilisé, $T : X \to X$ une application mesurable conservant μ et $\mathcal{C}$ une partition mesurable finie de X. Pour chaque $\varepsilon > 0$, il existe un entier n_ε tel que pour tout $n \geq n_\varepsilon$, la partition $\bigvee_{i=0}^{n-1} T^{-i}\mathcal{C}$ peut être divisée en deux classes $\mathcal{B}_n$ et $\mathcal{M}_n$ telles que*
i. $\mu(\cup_{A \in \mathcal{M}_n}) \leq \varepsilon$,
ii pour tout $A \in \mathcal{B}_n$,

$$e^{-n(h(T,\mathcal{C})+\varepsilon)} \leq \mu(A) \leq e^{-n(h(T,\mathcal{C})-\varepsilon)}.$$

Exercice Soit $\mathcal{A}$ un alphabet fini. Un source S émet une suite de lettre de l'alphabet $\mathcal{A}$. On désire compresser les messages émis par la source S. Le théorème de Shannon-MacMillan-Briemann permet de déterminer pour une source convenable le taux de compression maximal possible. On modélise la source par une partie X de $\mathcal{A}^{\mathbf{N}}$. Un élément de X est un mot infini écrit dans l'alphabet $\mathcal{A}$, il correspond à une émission sans interruption de la source. La partie X représente l'ensemble de toutes les émissions possibles de la source S. Nous avons besoin de quelques hypothèses sur X. Appelons $T : \mathcal{A}^{\mathbf{N}} \to \mathcal{A}^{\mathbf{N}}$ le décalage c'est-à-dire l'application définie par $T((x_n)_{n\in\mathbf{N}}) = (y_n)_{n\in\mathbf{N}}$ où $y_n = x_{n+1}$ pour tout entier naturel n. On suppose que X est stable par T, cela traduit qu'une émission dont on a tronqué le début est aussi une émission possible. On suppose aussi que X est une partie borélienne de $\mathcal{A}^{\mathbf{N}}$ et que X est muni d'une probabilité μ invariante par T. Cela signifie que la probabilité d'apparition d'un mot fini ne dépend pas de la position du mot dans le message émis. Finalement, on fera aussi l'hypothèse "T est ergodique" : presque tout message émis contient tous les mots de probabilité non nulle avec une fréquence égale à la probabilité du mot. Définissons maintenant un codage.
Rappels. Le langage $\mathcal{L}(X)$ est l'ensemble des sous-mots finis ou facteurs des éléments de X et que $\mathcal{L}_n(X)$ désigne l'ensemble des sous-mots ou facteurs de longueur n des élements de X.
Pour un alphabet $\mathcal{B}$, on désigne par $\mathcal{B}^*$ l'ensemble de tous les mots finis écrits avec l'alphabet $\mathcal{B}$, $\mathcal{B}^* = \cup_{n\geq 0}\mathcal{B}^n$.

Soit $\mathcal{B}$ un alphabet fini. Un codage de la source modélisée par X est une application

$$\phi : \mathcal{L}_{n_0}(X) \to \mathcal{B}^*$$

où n_0 est un entier fixé. Le codage est bon si aucun mot image n'est le début de l'image d'un autre mot. Par concaténation ϕ s'étend à l'ensemble des mots dont la longueur est un multiple de n_0

$$\phi : \cup_{k\geq 1}\mathcal{L}_{kn_0}(X) \to \mathcal{B}^*$$

par

$$\phi(m_1...m_k) = \phi(m_1)...\phi(m_k).$$

La condition précédente entraîne l'injectivité et permet le décodage. Le taux de compression $\tau(\phi)$ du codage ϕ est définie par

$$\tau(\phi) = \limsup_{k\to\infty} \frac{\sum_{m\in\mathcal{L}_{kn}(X)} \mu([m])l(\phi(m))}{kn_0}$$

où $l(m)$ désigne la longueur du mot m et $[m]$ désigne l'ensemble des éléments de X commençant par m (cylindre associé à m). Soit $\mathcal{C} = \{[a] : a \in \mathcal{A}\}$ la partition de X

définie par les cylindres de longueurs 1. Notre but est de prouver que si T est ergodique alors la borne inférieure sur tous les bons codages ϕ à valeur dans $\mathcal{B}^*$ est

$$\tau_X = \inf_{\phi\ bon} \tau(\phi) = \frac{h_\mu(T,\mathcal{C})}{\ln N}$$

où $N = \operatorname{card} \mathcal{B}$.
(Il est facile de prouver grâce au théorème de Kolmogorov-Sinaï que $h_\mu(T) = h_\mu(T,\mathcal{C})$).
1. Montrer que si ϕ est un bon codage alors l'extension de ϕ à $\cup_{k\geq 1}\mathcal{L}_{kn_0}(X)$ est injective.
2. Soit $\varepsilon > 0$ et ϕ une application injective de $\cup_{k\geq 1}\mathcal{L}_{kn_0}(X)$ dans $\mathcal{B}^*$. On veut montrer que $\tau(\phi) > \tau = \frac{h_\mu(T,\mathcal{C})-2\varepsilon}{\ln N}$.
a. Soit n un entier de la forme kn_0. Soit E_n l'ensemble des éléments m de $\mathcal{L}_n(X)$ tels que $l(\phi(m)) \leq n\tau$. Montrer que $\operatorname{card} E_n \leq \frac{N}{N-1}\exp(n(h_\mu(T,\mathcal{C})-2\varepsilon))$. ($\phi$ est injective et le nombre de mots de $\mathcal{B}^*$ de longueur $\leq n\tau$ est inférieur $1+N+...+N^{[n\tau]} = \frac{N^{[n\tau]+1}-1}{N-1} \leq \frac{N}{N-1}\exp(n(h_\mu(T,\mathcal{C})-2\varepsilon)))$.
b. Soit $\mathcal{B}_n$ l'ensemble des mots m de $\mathcal{L}_n(X)$ tel que

$$\mu([m]) \leq e^{-n(h_\mu(T,\mathcal{C})-\varepsilon)}.$$

Posons $\lambda_n = \sum_{m\in\mathcal{B}_n}\mu([m])$. Montrer que

$$\sum_{m\in\mathcal{L}_n(X)} \mu([m])l(\phi(m)) \geq n\tau \times (\lambda_n - e^{-n(h_\mu(T,\mathcal{C})-\varepsilon)} \times \frac{N}{N-1}e^{n(h_\mu(T,\mathcal{C})-2\varepsilon)}).$$

(Les $m \in \mathcal{L}_n(X)$ tels que $l(\phi(m)) \leq n\tau$ sont divisés en deux catégories. Ceux qui ne sont pas dans $\mathcal{B}_n$ et ceux qui sont dans $\mathcal{B}_n$. Le nombres de ces m appartenant à $\mathcal{B}_n$ est au plus $\frac{N}{N-1}\exp(n(h_\mu(T,\mathcal{C})-2\varepsilon))$ et leurs mesures est $\leq e^{-n(h_\mu(T,\mathcal{C})-\varepsilon)}$). Conclure.
3. Soit $\varepsilon > 0$. Nous voulons prouver qu'il existe un codage bon ϕ tel que $\tau(\phi) \leq \tau = \frac{h_\mu(T,\mathcal{C})+2\varepsilon}{\ln N}$.
a. Soit $\mathcal{B}_n$ l'ensemble des mots m de $\mathcal{L}_n(X)$ tel que

$$\mu([m]) \geq e^{-n(h_\mu(T,\mathcal{C})+\varepsilon)}.$$

Montrer que si n est assez grand, il existe une injection ϕ de $\mathcal{B}_n$ dans l'ensemble $\mathcal{B}^{L_n}$ des mots de longueur $L_n = n\frac{h_\mu(T,\mathcal{C})+\varepsilon}{\ln N}$ écrit avec l'alphabet $\mathcal{B}$. (Le nombre de mots de longueur $n\frac{h_\mu(T,\mathcal{C})+\varepsilon}{\ln N}$ écrit dans l'alphabet $\mathcal{B}$ est $N^{L_n} = \exp n(h_\mu(T,\mathcal{C})+\varepsilon)$).
b. Fixons n tel que ϕ soit injective. Sélectionnons un mots w_n de longueur L_n et supposons que w_n ne soit pas dans l'image de ϕ. Soit f une application injective de $\mathcal{A}$ dans $\mathcal{B}^*$ tel que la longueur des images soit une constante c. L'application f s'étend à $\mathcal{A}^*$ par concaténation, $f(a_1...a_k) = f(a_1)...f(a_k)$. Etendons ϕ à $\mathcal{L}_n(X)$ par $\phi(m) = w_nf(m)$ pour les mots m qui ne sont pas dans $\mathcal{B}_n$. Montrer que ϕ est un bon codage.
(pour tous les $m \in \mathcal{B}_n$, les $\phi(m)$ ont la même longueur L_n, comme ϕ est injective sur $\mathcal{B}_n$ aucun mot de $\phi(\mathcal{B}_n)$ ne peut prolonger un autre mot de $\phi(\mathcal{B}_n)$. De plus w_n est de longueur L_n et n'appartient pas à $\phi(\mathcal{B}_n)$. Donc aucun $\phi(m)$ avec $m \in \mathcal{L}_n\backslash\mathcal{B}_n$ ne peut prolonger un mot de $\phi(\mathcal{B}_n)$ et réciproquement. Pour finir il suffit de remarquer que l'extension de f est injective et que toutes les images $f(m)$, $m \in \mathcal{L}_n\backslash\mathcal{B}_n$ ont même longueur.)
c. Posons $\lambda_n = \sum_{m\in\mathcal{B}_n}\mu([m])$. Montrer que

$$\sum_{m\in\mathcal{L}_n(X)} \mu([m])l(\phi(m)) \leq (1-\lambda_n)(cn + n\frac{h_\mu(T,\mathcal{C})+\varepsilon}{\ln N}) + n\frac{h_\mu(T,\mathcal{C})+\varepsilon}{\ln N}.$$

d. Montrer que

$$\sum_{m \in \mathcal{L}_{kn}(X)} \mu([m]) l(\phi(m)) = k \sum_{m \in \mathcal{L}_n(X)} \mu([m]) l(\phi(m))$$

(Posons $l_i(m) = l(\phi(m_i))$ où $m = m_1...m_k$ et $m_i \in \mathcal{L}_n(X)$. Pour $w \in \mathcal{L}_n(X)$, le cylindre $[w]_i$ est l'ensemble des mots de $m_1...m_k \in \mathcal{L}_{kn}(X)$ tel que $m_i = w$. On a pour tout $m \in \mathcal{L}_{nk}(X)$, $l(\phi(m)) = l_1(m) + ... + l_k(m)$ et $\sum_{m \in \mathcal{L}_{kn}(X)} \mu([m]) l_i(m) = \sum_{w \in \mathcal{L}_n(X)} \sum_{m \in [w]_i} \mu([m]) l_i(w)$. De plus par invariance de la mesure μ, $\sum_{m \in [w]_i} \mu([m]) = \mu([w])$ d'où le résultat. On peut aussi interpréter les sommes comme des espérances de variables aléatoires définies sur X). Conclure.

1.6. Calcul de l'entropie. Walters page 94.

Le calcul de l'entropie semble difficile car l'entropie est définie à l'aide d'un borne supérieure sur l'ensemble des partitions. Nous allons voir que dans certains cas il est possible de trouver des partitions réalisant cette borne supérieure.

Théorème 7. ***(Kolmogorov-Sinai)*** *Soit $(X, \mathcal{A}, \mu)$ un espace probabilisé, $T : X \to X$ une bijection bimesurable conservant μ et $\mathcal{C}$ une partition mesurable finie de X. Si la tribu complète engendrée par $\cup_{n \geq 0} \bigvee_{i=-n}^{n} T^{-i}\mathcal{C}$ contient $\mathcal{A}$ alors $h(T) = h(T, \mathcal{C})$.*

Dem. Soit $\mathcal{D} \subset \mathcal{A}$ une partition finie. Nous devons prouver que $h(T, \mathcal{D}) \leq h(T, \mathcal{C})$. On sait que (paragraphe précédent)

$$\begin{aligned} h(T, \mathcal{D}) &\leq h(T, \bigvee_{i=-n}^{n} T^{-i}\mathcal{C}) + H(\mathcal{D} | \bigvee_{i=-n}^{n} T^{-i}\mathcal{C}) \\ &= h(T, \mathcal{C}) + H(\mathcal{D} | \bigvee_{i=-n}^{n} T^{-i}\mathcal{C}). \end{aligned}$$

Soit $\mathcal{F}$ la tribu engendrée par $\cup_{n \geq 0}(\bigvee_{i=-n}^{n} T^{-i}\mathcal{C})$. On a $\lim_{n\to\infty} H(\mathcal{D} | \bigvee_{i=-n}^{n} T^{-i}\mathcal{C}) = H(\mathcal{D}|\mathcal{F})$ et comme $\mathcal{D}$ est inclus dans la complétée de $\mathcal{F}$, $H(\mathcal{D}|\mathcal{F}) = 0$.

Avec la même démonstration on obtient :

Théorème 8. ***(Cas non inversible)*** *Soit $(X, \mathcal{A}, \mu)$ un espace probabilisé, $T : X \to X$ une application mesurable conservant μ et $\mathcal{C}$ une partition mesurable finie de X. Si la tribu complète engendrée par $\cup_{n \geq 0} \bigvee_{i=0}^{n} T^{-i}\mathcal{C}$ contient $\mathcal{A}$ alors $h(T) = h(T, \mathcal{C})$.*

Corollaire 2. *Soit $(X, \mathcal{A}, \mu)$ un espace probabilisé, $T : X \to X$ une bijection bimesurable conservant μ. Supposons qu'il existe une partition mesurable finie $\mathcal{C}$ de X telle que la tribu complète engendrée par $\cup_{n \geq 0} \bigvee_{i=0}^{n} T^{-i}\mathcal{C}$ contienne $\mathcal{A}$ alors $h(T) = 0$.*

Dem. On a $h(T) = h(T, \mathcal{C})$ et $\sigma(\cup_{n \geq 1} \bigvee_{i=1}^{n} T^{-i}\mathcal{C}) = T^{-1}\sigma(\cup_{n \geq 0} \bigvee_{i=0}^{n} T^{-i}\mathcal{C}) = T^{-1}\mathcal{A} = \mathcal{A}$ donc

$$h(T, \mathcal{C}) = \lim_{n\to\infty} H(\mathcal{C} | \bigvee_{i=1}^{n} T^{-i}\mathcal{C}) = H(\mathcal{C}|\mathcal{A}) = 0. \square$$

Remarque : Théorème de Krieger. ([Kr])

On peut étendre la définition de l'entropie aux partitions dénombrables et on peut montrer

que l'entropie d'une application est encore obtenue avec la borne supérieure sur toutes les partitions dénombrables. Dans le cas inversible, une partition dénombrable $\mathcal{C}$ telle que $\cup_{n\geq 0} \bigvee_{i=-n}^{n} T^{-i}\mathcal{C}$ engendre $\mathcal{A}$ modulo les ensembles négligeables s'appelle un générateur. Krieger à montré le résultat :
Soit $(X, \mathcal{A}, \mu)$ un espace de Lebesgue (qui n'est pas isomorphe à un ensemble fini) et $T : X \to X$ une bijection bimesurable conservant μ. Si $h(T) < \infty$ alors il existe un générateur.

Donnons encore quelques résultats qui permettent de calculer l'entropie.

Lemme 2. *Soit $k \geq 1$ un entier et $(X, \mathcal{A}, \mu)$ un espace probabilisé. Pour tout $\varepsilon > 0$, il existe $\delta > 0$ tel que si $\mathcal{C} = \{C_1, ..., C_k\}$ et $\mathcal{D} = \{D_1, ..., D_k\}$ sont deux partitions mesurables à k éléments, alors*

$$\sum_{i=1}^{k} \mu(C_i \Delta D_i) \leq \delta \Longrightarrow H(\mathcal{C}|\mathcal{D}) + H(\mathcal{D}|\mathcal{C}) \leq \varepsilon.$$

Dem. Pour $\varepsilon > 0$ fixé choisissons $\delta > 0$ tel que $\delta < \frac{1}{4}$ et $k(k-1)\phi(\delta) + \phi(1-\delta) \leq \varepsilon/2$ où $\phi(t) = -t \ln t$. Considérons la partition $\mathcal{E}$ de X formée par les parties $C_i \cap D_j$, $i \neq j$, et $\cup_{i=1}^{k}(C_i \cap D_i)$. Le nombre d'éléments de $\mathcal{E}$ est $\leq k(k-1)+1$. L'hypothèse $\sum_{i=1}^{k} \mu(C_i \Delta D_i) \leq \delta$ entraîne que tout pour $E \in \mathcal{E}$ on a soit $\mu(E) \leq \delta$ soit $\mu(E) \geq 1-\delta$. Par conséquent $H(\mathcal{E}) \leq k(k-1)\phi(\delta) + \phi(1-\delta) \leq \varepsilon/2$. De plus les partitions $\mathcal{E} \vee \mathcal{C}$ et $\mathcal{C} \vee \mathcal{D}$ sont indentiques, donc

$$H(\mathcal{C} \vee \mathcal{D}) = H(\mathcal{E} \vee \mathcal{C}) = H(\mathcal{C}) + H(\mathcal{D}|\mathcal{C}) \leq H(\mathcal{E}) + H(\mathcal{C})$$

donc $H(\mathcal{D}|\mathcal{C}) \leq H(\mathcal{E}) \leq \varepsilon/2$. L'autre inégalité se démontre de la même manière. $\square$

Théorème 9. *Soit $(X, \mathcal{A}, \mu)$ un espace probabilisé $\mathcal{B}$ une algèbre de parties telle que la tribu complète engendrée par $\mathcal{B}$ soit égale à la complétée de $\mathcal{A}$. Soit $\mathcal{C}$ une partition mesurable finie de X. Alors pour tout $\varepsilon > 0$, il existe une partition finie $\mathcal{E} \subset \mathcal{B}$ telle que $H(\mathcal{E}|\mathcal{C}) + H(\mathcal{C}|\mathcal{E}) \leq \varepsilon$.*

Dem. $\mathcal{C} = \{C_1, ..., C_k\}$. Soit $\varepsilon > 0$. Choisissons $\delta > 0$ correspondant à k et ε dans le lemme précédent. Soit α un nombre strictement positif que l'on choisira plus loin. Comme la tribu complète engendrée par $\mathcal{B}$ contient $\mathcal{A}$ pour chaque i il existe une partie $\mathcal{D}_i$ appartenant à $\mathcal{B}$ telle que $\mu(C_i \Delta D_i) \leq \alpha$ (l'ensemble des parties mesurables A telles que pour tout $\alpha > 0$, il existe une partie $B \in \mathcal{B}$ avec $\mu(A \Delta B) \leq \alpha$, est une tribu contenant $\mathcal{B}$). Si $i \neq j$, l'intersection $D_i \cap D_j$ est incluse dans la réunion des différences symétriques $D_i \Delta C_i$ et $C_j \Delta D_j$, d'où $\mu(D_i \cap D_j) \leq 2\alpha$. La partie $N = \cup_{i \neq j} D_i \cap D_j$ a donc une mesure $\leq k(k-1)\alpha$. Prenons $E_i = D_i \backslash N$, $i = 1, ..., k-1$, et $E_k = X \backslash \cup_{i<k} E_i$. Les parties $E_1, ..., E_k$ forment une partition $\mathcal{E}$ de X et par construction $\mathcal{E} \subset \mathcal{B}$. Pour chaque $i < k$, on a $D_i \Delta E_i \subset N$ donc

$$\begin{array}{rcl} C_i \Delta E_i & \subset & C_i \Delta D_i \cup D_i \Delta E_i, \\ \mu(C_i \Delta E_i) & \leq & \mu(C_i \Delta D_i) + \mu(D_i \Delta E_i) \leq \alpha + \mu(N) \leq (1 + k(k-1))\alpha. \end{array}$$

De plus,

$$C_k \Delta E_k = C_k^C \Delta E_k^C \subset \cup_{i<k} C_i \Delta E_i$$

donc $\mu(C_k \Delta E_k) \leq \sum_{i<k} \mu(C_i \Delta E_i)$. Ainsi, en choisissant α assez petit on a $\sum_{i=1}^{k} \mu(C_i \Delta E_i) \leq \delta$ et donc $H(\mathcal{E}|\mathcal{C}) + H(\mathcal{C}|\mathcal{E}) \leq \varepsilon$. $\square$

Théorème 10. *Soit $(X,\mathcal{A},\mu)$ un espace probabilisé, $T:X\to X$ une application mesurable conservant μ et $\mathcal{B}$ une sous algèbre de $\mathcal{A}$. Si la tribu complète engendrée par $\mathcal{B}$ contient $\mathcal{A}$ alors*

$$h(T)=\sup_{\mathcal{C}} h(T,\mathcal{C})$$

où $\mathcal{C}$ parcourt l'ensemble des partitions mesurables finies de X dont tous les éléments appartiennent à $\mathcal{B}$.

Dem. Soit $\mathcal{C}$ une partition finie dont tous les éléments appartiennent à $\mathcal{A}$ et $\varepsilon>0$. D'après le théorème précédent, il existe une partition finie $\mathcal{D}$ incluse dans $\mathcal{B}$ telle que $H(\mathcal{C}|\mathcal{D})\leq\varepsilon$. Le théorème découle de l'inégalité

$$h(T,\mathcal{C})\leq h(T,\mathcal{D})+h(\mathcal{C}|\mathcal{D}).\ \square$$

Corollaire 3. *Soit $(X_1,\mathcal{A}_1,\mu_1)$ et $(X_1,\mathcal{A}_1,\mu_1)$ deux espaces probabilisés, $T_i:X\to X$, $i=1,2$, deux applications mesurables conservant μ_i. Soit $T:X_1\times X_2\to X_1\times X_2$ définie par $T(x_1,x_2)=(T_1(x_1),T_2(x_2))$. On a*

$$h(T)=h(T_1)+h(T_2).$$

Dem. Utilisons le théorème précédent avec la sous algèbre $\mathcal{B}$ de $\mathcal{A}_1\otimes\mathcal{A}_2$ engendrée par les parties $A_1\times A_2$ où $A_1\in\mathcal{A}_1$ et $A_2\in\mathcal{A}_2$ qui par définition engendre la tribu produit $\mathcal{A}_1\otimes\mathcal{A}_2$. Soit $\mathcal{C}$ une partition finie de X dont tous les éléments appartiennent à $\mathcal{B}$. Comme les éléments de $\mathcal{B}$ sont tous des réunions finies de rectangles $A_1\times A_2$, il existe une partition $\mathcal{D}_1$ de X_1 incluse dans $\mathcal{A}_1$ et une partition $\mathcal{D}_2$ de X_2 incluse dans $\mathcal{A}_2$ telles que la partition $\mathcal{D}=\mathcal{D}_1\times\mathcal{D}_2=\{A_i\times B_j: i=1,...,n,\ j=1,...,m\}$ soit plus fine que $\mathcal{C}$. Par conséquent, $h(T,\mathcal{C})\leq h(T,\mathcal{D})$ et d'après le théorème précedent, $h(T)$ est inférieure au sup des entropie $h(T,\mathcal{D})$ où $\mathcal{D}$ est une partition produit du type précédent. Pour une telle partition on a

$$H(\bigvee_{i=0}^{n} T^{-i}\mathcal{D})=H(\bigvee_{i=0}^{n}(T_1^{-i}\mathcal{D}_1\times T_2^{-i}\mathcal{D}_2)=H(\bigvee_{i=0}^{n} T_1^{-i}\mathcal{D}_1\times\bigvee_{i=0}^{n} T_2^{-i}\mathcal{D}_2).$$

Pour terminer la démonstration il suffit de vérifier que si $\mathcal{C}_1=\{A_1,...,A_n\}$ et $\mathcal{C}_2=\{B_1,...,B_m\}$ sont des partitions de X_1 et X_2 alors $H(\mathcal{C}_1\times\mathcal{C}_2)=H(\mathcal{C}_1)+H(\mathcal{C}_2)$. Comme $\mathcal{C}_1\times\mathcal{C}_2=(\mathcal{C}_1\times\{X_2\})\vee(\{X_1\}\times\mathcal{C}_2)$, cela résulte de l'indépendance des partitions $\mathcal{C}_1\times\{X_2\}$ et $\{X_1\}\times\mathcal{C}_2$. $\square$

Théorème 11. *Soit (X,d) un espace métrique compact, $\mathcal{A}$ la tribu borélienne de X, μ une probabilité sur $(X,\mathcal{A})$ et $T:X\to X$ une application mesurable conservant μ. Soit $(\mathcal{C}_n)_{n\geq1}$ une suite de partitions finies dont tous les éléments appartiennent à $\mathcal{A}$. Supposons que*

$$\lim_{n\to\infty}\sup_{C\in\mathcal{C}_n}\mathrm{diam}(C)=0.$$

Alors

$$h(T)=\sup_n h(T,\mathcal{C}_n).$$

Dem. Soit $\mathcal{D}=\{D_1,...,D_k\}$ une partition mesurable finie de X et $\varepsilon>0$. Il nous suffit de montrer qu'il existe une partition $\mathcal{C}_n$ telle que $h(T,\mathcal{C}_n)\geq h(T,\mathcal{D})-\varepsilon$. Soit α un nombre réel > 0. Pour chaque i, il existe un compact K_i inclus dans D_i dont la mesure est supérieure à celle de D_i moins α. La distance entre deux quelconques de ces compacts est minorée par un nombre $\delta > 0$. Choisissons n tel que le maximum des diamètres des éléments de $\mathcal{C}_n$ soit $< \delta/2$. Considérons pour chaque $i \leq k-1$, la réunion B_i des $C \in \mathcal{C}_n$ qui rencontre K_i. Définissons B_k comme le complémentaire de $B_1 \cup ... \cup B_{k-1}$. Le choix de n montre que les B_i, $i = 1, ..., k-1$ sont deux à deux disjoints et par construction B_k est disjoint des autres, par conséquent $\mathcal{B} = \{B_1, ..., B_k\}$ est une partition de X. De plus, pour $i = 1, ..., k$,

$$\mu(D_i \Delta B_i) \leq \max(\alpha, \mu(X \backslash K_1 \cup ... \cup K_k) = \max(\alpha, k\alpha) = k\alpha.$$

donc d'après le lemme, si α est assez petit,

$$H(\mathcal{D}|\mathcal{B}) + H(\mathcal{B}|\mathcal{D}) \leq \varepsilon.$$

Comme la partition $\mathcal{C}_n$ est plus fine que la partition $\mathcal{B}$ on a $H(\mathcal{D}|\mathcal{C}_n) \leq H(\mathcal{D}|\mathcal{B}) \leq \varepsilon$, il en résulte que

$$h(T, \mathcal{D}) \leq h(T, \mathcal{C}_n) + H(\mathcal{D}|\mathcal{C}_n) \leq h(T, \mathcal{C}_n) + \varepsilon. \ \square$$

Corollaire 4. *Considérons le tore* $\mathbf{T}^1 = \mathbf{R}/\mathbf{Z}$ *muni de la mesure de Lebesgue. Si* T : $\mathbf{T}^1 \to \mathbf{T}^1$ *est translation alors* $h(T) = 0$.

Dem : exercice.

Corollaire 5. *Soit* $\mathbf{T}^n = \mathbf{T}^1 \times ... \times \mathbf{T}^1$ *et* $T : \mathbf{T}^n \to \mathbf{T}^n$ *une translation. On a* $h(T) = 0$.

1.7. Exemples. Translation d'un groupe abélien compact

Exercice Soit $(G, +)$ un groupe abélien métrique compact, μ la mesure de Haar de G et $T : G \to G$ une translation de G. Notons $\widehat{G} = \{\gamma_1, \gamma_2, ...\}$ le groupe dual de G (on sait que $\widehat{G}$ est dénombrable). Soit $\gamma_1, ..., \gamma_n$ un nombre fini d'éléments de G et $H_n = \cap_{i=1}^n \ker \gamma_i$. H_n est un sous-groupe compact de G.
1. Montrer que $R_n : x \in G/H_n \to (\gamma_1(x), ..., \gamma_n(x)) \in (\mathbf{S}^1)^n \approx \mathbf{T}^n$ est morphisme injectif continue. Appelons $\mathbf{K}_n$ l'image de R_n. $\mathbf{K}_n$ est un sous-groupe compact de $\mathbf{S}_1^n$ dont on admettra qu'il est isomorphe à un groupe de la forme $\mathbf{Z}/p\mathbf{Z} \times \mathbf{S}^q$.
2. Soit $T_n : G/H_n \to G/H_n$ l'application définie par $T_n(g + H_n) = a + g + H_n$. Montrer que $h(T_n) = 0$. (Montrer que T est conjugué à une translation de $\mathbf{K}_n$).
3. Soit $\mathcal{A}_n$ la tribu image réciproque par R_n de la tribu borélienne de $\mathbf{T}^n$. Montrer que la tribu borélienne de G est égale à la tribu $\mathcal{A}$ engendrée par $\cup_{n \geq 1} \mathcal{A}_n$ (les caractères sont $\mathcal{A}$-mesurable et les combinaisons linéaires d'éléments de $\widehat{G}$ sont dense dans $\mathcal{C}(G, \mathbf{R})$ donc tout ouvert est réunion dénombrable d'éléments de $\mathcal{A}$).
4. Montrer que si $\mathcal{C}$ est une partition finie de G incluse dans $\mathcal{A}_n$ alors $h(T, \mathcal{C}) = 0$ (Désignons p_n la projection de G sur G/H_n. Ecrire les éléments C de $\mathcal{C}$ sous la forme $p_n^{-1}(D)$ où D appartient à une partition $\mathcal{D}$ de G/H_n et remarquer que $\bigvee_{i=0}^k T^{-i}\mathcal{C} = p_n^{-1}(\bigvee_{i=0}^k T_n^{-i}\mathcal{D})$ et donc $H(\bigvee_{i=0}^k T^{-i}\mathcal{C}) = H(\bigvee_{i=0}^k T_n^{-i}\mathcal{D})$ car par unicité p_n, envoie la mesure de Haar de G sur la mesure de Haar de G/H_n)
5. Conclure.

Décalage

Theorem 12. *Considérons l'ensemble* $\{1,...,k\}$ *muni d'une probabilité* μ *et notons* p_i *la probabilité de* i*. Soit* $\Omega_b = \{1,...,k\}^{\mathbf{Z}}$ *muni de la probabilité produit* $\nu = \mu^{\otimes \mathbf{Z}}$*. Appelons* $\sigma : \Omega_b \to \Omega_b$ *l'application décalage* $\sigma((x_n)_{n\in\mathbf{Z}}) = (y_n)_{n\in\mathbf{Z}}$ avec $y_n = x_{n+1}$*. On a* $h(T) = -\sum_{i=1}^k p_i \ln p_i$.

Dem : exercice.

Theorem 13. *Considérons l'ensemble* $\{1,...,k\}$ *muni d'une probabilité* μ *et notons* p_i *la probabilité de* i*. Soit* $\Omega_u = \{1,...,k\}^{\mathbf{N}}$ *muni de la probabilité produit* $\nu = \mu^{\otimes \mathbf{N}}$*. Appelons* $\sigma : \Omega_u \to \Omega_u$ *l'application décalage* $\sigma((x_n)_{n\in\mathbf{N}}) = (y_n)_{n\in\mathbf{N}}$ avec $y_n = x_{n+1}$*. On a* $h(T) = -\sum_{i=1}^k p_i \ln p_i$.

Dem : exercice.

Décalage de Markov.

Soit $P = (p_{ij})_{1\le i,j\le k}$ une matrice carrée dont les lignes sont des probabilités et $p = (p_1,...,p_k)$ un vecteur probabilité tel que pour tout i, $p_j = \sum_i p_i p_{ij}$. L'espace $\Omega_b = \{1,...,k\}^{\mathbf{Z}}$ peut-être muni d'une mesure μ définie par

$$\mu([x_n,...,x_m]) = p_{x_n} p_{x_n,x_{n+1}} p_{x_{n+1},x_{n+2}} ... p_{x_{m-1}x_m}$$

où $n < m$ sont deux entiers relatifs. Pour vérifier que cela définie bien une mesure il suffit de vérifier que

$$\mu([x_n,...,x_m]) = \sum_{x_{n-1}=1}^{k} \mu([x_{n-1},x_n...,x_m])$$

et

$$\mu([x_n,...,x_m]) = \sum_{x_{m+1}=1}^{k} \mu([x_n,...,x_m,x_{m+1}]).$$

La première relation est évidente car $p_{x_n} = \sum_{x_{n-1}=1}^k p_{x_{n-1}} p_{x_{n-1}x_n}$, la seconde l'est aussi car les lignes de la matrices P sont des probabilités. La définition même de μ montre que μ est invariante par le décalage σ.

Théorème 12. *Soit* $P = (p_{ij})_{1\le i,j\le k}$ *une matrice carrée dont les lignes sont des probabilités et* $p = (p_1,...,p_k)$ *un vecteur probabilité sur* $\{1,...,k\}$ P*-invariant. Munissons* $\Omega_b = \{1,...,k\}^{\mathbf{Z}}$ *de la probabilité* μ *définie par* P *et* p*. Alors l'entropie du décalage* $\sigma : \Omega_b \to \Omega_b$ *est*

$$h_\mu(\sigma) = -\sum_{i,j} p_i p_{ij} \ln p_{ij}.$$

Dem : exercice.

Automorphisme du tore $\mathbf{T}^2$.

On munit le tore de la mesure de Lebesgue.

Lemme 3. *Soit* $A \in SL(2,\mathbf{Z})$ *et* $\sigma(A) = \{\lambda_1,\lambda_2\}$ *le spectre de* A*. Alors soit* $\sigma(A) \subset \mathbf{R}\backslash\mathbf{Q}$ *soit* $\sigma(A) \subset \mathbf{S}^1 = \{z \in \mathbf{C} : |z| = 1\}$*. De plus, si* $\lambda_1 = \lambda_2$ *alors* $\lambda_1 = \lambda_2 = \pm 1$.

Dem. - Si les valeurs propres λ_1 et λ_2 de A sont ne sont pas réelles alors elles sont conjuguées et $|\lambda_i|^2 = \lambda_i \overline{\lambda_i} = \lambda_1 \lambda_2 = \det A = 1$.
- De même si $\lambda_1 = \lambda_2$ alors $1 = \lambda_1 \lambda_2 = \lambda_1^2$ et $\lambda_1 = \pm 1$.
- Supposons $\sigma(A) = \{\lambda_1, \lambda_2\} \subset \mathbf{Q}$ et $\lambda_1 \neq \lambda_2$.
Alors il existe des vecteurs propres à coordonnées rationnelles. Montrons que cela est impossible si les valeurs propres ont une valeur absolue différente de 1. Après multiplication on peut obtenir des vecteurs propres à coordonnées entières. Appelons X un tel vecteur propre. Pour tout $n \in \mathbf{Z}$, $A^n X \in \mathbf{Z} \backslash \{0\}$ mais cela est impossible car $A^n X = \lambda^n X$ qui tend vers 0 quand n tend vers $+\infty$ ou $-\infty$. $\square$

Lemme 4. *Soit $A \in SL(2, \mathbf{Z})$ et T_A l'automorphisme induit par A sur le tore $\mathbf{T}^2 = \mathbf{R}^2/\mathbf{Z}^2$.*
- L'application T_A est ergodique ssi le spectre de A, $\sigma(A)$ est inclus dans $\mathbf{R} \backslash \mathbf{Q}$.
- Si T_A est ergodique, aucun vecteurs de $\mathbf{Q}^2$ n'est un vecteur propre de A.

Dem. Cas 1 $\sigma(A) = \{\lambda_1, \lambda_2\} \subset \mathbf{S}^1$ et $\lambda_1 \neq \lambda_2$.
La matrice A est conjuguée à la matrice diagonale $\begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$, donc pour tout vecteur $X \in \mathbf{Z}^2$, l'orbite de X sous l'action de la transposée de A, $\mathcal{O}(X) = \{ {}^t A^n X : n \in \mathbf{Z}\}$, est bornée et inclus dans $\mathbf{Z}^2$. Cette orbite est donc finie. Fixons un vecteur non nul X de $\mathbf{Z}^2$ et considérons la fonction $f : \mathbf{T}^2 \to \mathbf{C}$ définie par

$$f(x) = \sum_{Y \in \mathcal{O}(X)} \exp 2i\pi Y.x \ .$$

La fonction f n'est pas constante car les exponentielles sont orthogonales et donc linéairement indépendantes (la fonction 1 est une exponentielle). La fonction f est invariante car

$$\begin{aligned} f(T_A x) &= \sum_{Y \in \mathcal{O}(X)} \exp 2i\pi Y.Ax = \sum_{Y \in \mathcal{O}(X)} \exp 2i\pi \, {}^t A Y.x \\ &= \sum_{Y \in \mathcal{O}(X)} \exp 2i\pi Y.x = f(x), \end{aligned}$$

donc T_A n'est pas ergodique.
Cas 2 $\lambda_1 = \lambda_2 = \pm 1$.
Soit X un vecteur propre associé à λ_1. On peut supposer que le vecteur X est à coordonnées entières. La fonction

$$f(x) = \exp 2i\pi X.x + \exp 2i\pi (-X).x$$

est T_A invariante donc T_A n'est pas ergodique.
Cas 3 $\sigma(A) = \{\lambda_1, \lambda_2\} \subset \mathbf{R} \backslash \mathbf{Q}$.
On sait que $\lambda_1 \neq \lambda_2$ et $\lambda_1 \lambda_2 = 1$. On peut donc supposer que $|\lambda_1| > 1 > |\lambda_2|$.
Montrons que pour tout vecteur non nul $X \in \mathbf{Z}^2$, ${}^t A^{-n} X$ tend vers l'infini quand n tend l'infini. En effet, on démontre comme dans le lemme précédent, que les vecteurs propres ne peuvent pas être à coordonnées rationnelles. Donc si e_1 e_2 est une base de vecteurs propres de ${}^t A$, on a $X = x_1 e_1 + x_2 e_2$ où x_1 et x_2 sont tous les deux non nuls. Par conséquent ${}^t A^{-n} X = \lambda_1^n x_1 e_1 + \lambda_2^n x_2 e_2$ tend vers l'infini. Soit $f \in L^2(\mathbf{T}^2)$ une fonction T_A-invariante

et $\sum_{X\in\mathbf{Z}^2} a_X \exp 2i\pi X.x$ sa série de Fourrier. Pour tout $n \in \mathbf{N}$, on a $f\circ T_A^n = f$ et la série de Fourrier de $f \circ T_A^n$ est

$$\begin{aligned}\sum_{X\in\mathbf{Z}^2} a_X \exp 2i\pi X.A^n x &= \sum_{X\in\mathbf{Z}^2} a_X \exp 2i\pi\, {}^tA^n X.x \\ &= \sum_{X\in\mathbf{Z}^2} a_{{}^tA^{-n}X} \exp 2i\pi X.x,\end{aligned}$$

par unicité du développement en série de fourrier on en déduit que pour tout $X \in \mathbf{Z}^2$, $a_X = a_{{}^tA^{-n}X}$. Comme les coefficients de Fourrier tendent vers 0 à l'infini, si X est non nul, $a_{{}^tA^{-n}X}$ tend vers 0. Par conséquent, $a_X = 0$ est f est constante. $\square$

Théorème 13. *Soit $A \in SL(2,\mathbf{Z})$ et T_A l'automorphisme induit par A sur le tore $\mathbf{T}^2 = \mathbf{R}^2/\mathbf{Z}^2$. Si T_A est ergodique alors*

$$h(T,dx) = \ln|\lambda|$$

où λ est la plus grande valeur propre de A.

Dem.

Etape 1 : $h(T,dx) \geq \ln|\lambda|$.

Soit $\varepsilon > 0$ assez petit. Soit $\mathcal{C}$ une partition de $\mathbf{T}^2$ dont tous les éléments ont un diamètre $\leq \varepsilon$. Appelons D_1 et D_2 les droites propres associées aux valeurs propres λ_1 et λ_2 où $|\lambda_1| > 1$. Pour $n \geq 0$, montrons que les diamètres des éléments de la partition $T^{-n}\mathcal{C} \vee T^{-n+1}\mathcal{C} \vee ... \vee T^n\mathcal{C}$ sont tous inférieure à $2\varepsilon\lambda_1^{-n}$. Soit C un éléments de cette partition, pour chaque k compris entre $-n$ et n il existe $C_k \in \mathcal{C}$ tel que $C = \cap_{k=-n}^n T^{-k}C_k$. Soit x et $y \in C$, on a $T^k(x)$ et $T^k(y) \in C_k$. Ainsi pour tout k, $d(T^kx, T^ky) \leq \varepsilon$. Prenons des représentants X et Y de x et y dans $\mathbf{R}^2$. On peut choisir X et Y de tel sorte que $d(x,y) = d(X,Y)$. Appelons Z le point d'intersection des droites $X + D_1$ et $Y + D_2$ et z la projection de Z dans $\mathbf{T}^2$. On a

$$\max(|X-Z|, |Y-Z|) \leq K|X-Y|$$

où K ne dépend que de la matrice A. Si ε est assez petit on a aussi $d(x,z) = d(X,Z)$ et $d(y,z) = d(Y,Z)$. Pour chaque $k \in \{0,...,n\}$, on a

$$\begin{aligned} d(T^kx, T^kz) &\leq d(T^kx, T^ky) + d(T^ky, T^kz) \\ &\leq \varepsilon + d(A^kY, A^kZ) \\ &= \varepsilon + \left|A^k(Y-Z)\right| = \varepsilon + \lambda_2^k|Y-Z| \leq \varepsilon + |Y-Z| \leq (K+1)\varepsilon. \end{aligned}$$

Posons $L = K+1$. Par définition de la distance sur le tore

$$d(T^kx, T^kz) = d(A^kX - A^kZ, \mathbf{Z}^2) = \left|A^kX - A^kZ + P_k\right| \leq L\varepsilon$$

où P_k est un point à coordonnées entières. Montrons par récurrence que si ε est assez petit, alors $\left|A^kX - A^kZ\right| \leq K\varepsilon$, $k = 0,...,n$, sont tous nuls. En effet, pour $U \in \mathbf{R}^2$ on a

$$|U| \leq L\varepsilon \Rightarrow d(\lambda_1 U, \mathbf{Z}^2) = |\lambda_1 U|$$

dès que ε est petit. Par hypothèse $|X - Z| \leq K\varepsilon$ et si $\left|A^k X - A^k Z\right| \leq L\varepsilon$ alors

$$\begin{aligned} d(T^{k+1}x, T^{k+1}z) &= d(A^{k+1}X - A^{k+1}Z, \mathbf{Z}^2) = d(\lambda_1(A^k X - A^k Z), \mathbf{Z}^2) \\ &= \left|\lambda_1(A^k X - A^k Z)\right| \\ &= \left|A^{k+1}X - A^{k+1}Z\right| \end{aligned}$$

et donc $\left|A^{k+1}X - A^{k+1}Z\right| \leq L\varepsilon$. Finalement, $|A^n X - A^n Z| \leq L\varepsilon$, on en déduit que

$$|X - Z| = \left|A^{-n}(A^n X - A^n Z)\right| = \left|\lambda_1^{-n}(A^n X - A^n Z)\right| \leq K\varepsilon \left|\lambda_1\right|^{-n}.$$

De la même manière en considérant les distances $d(T^{-k}y, T^{-k}z)$, $k = 0, ..., n$, on montre que $|Y - Z| \leq L\varepsilon \left|\lambda_1\right|^{-n}$. Il en résulte que $d(x, y) \leq 2L\varepsilon$. Le diamètre de C est donc inférieure à $2K\varepsilon$ et l'aire de C est inférieure à $(2K\varepsilon)^2 \left|\lambda_1\right|^{-2n}$. Ceci étant vrai pour tout les éléments de la partition $T^{-n}\mathcal{C} \vee T^{-n+1}\mathcal{C} \vee ... \vee T^n\mathcal{C}$, on a

$$H(T^{-n}\mathcal{C} \vee T^{-n+1}\mathcal{C} \vee ... \vee T^n\mathcal{C}) \geq 2n \ln|\lambda_1| - 2\ln 2K\varepsilon$$

et donc

$$h(T) \geq h(T, \mathcal{C}) \geq \ln|\lambda_1|$$

Etape 2 : $h(T, dx) \leq \ln|\lambda|$.
Pour chaque entier N, considérons l'ensemble $\mathcal{P}_N$ des partitions de $\mathbf{T}^2$ obtenues en découpant $\mathbf{T}^2$ en N^{2q} carrés de côtés $1/N^q$ parallèles aux axes où q est un entier positif. Comme ces partitions engendrent la tribu borélienne de $\mathbf{T}^2$, $h(T) = \sup_{\mathcal{C} \in \mathcal{P}_N} h(T, \mathcal{C})$. Soit $\mathcal{C}$ une partition de $\mathbf{T}^2$ appartenant à $\mathcal{P}_N$, pour tout entier m, on a

$$h(T, \mathcal{C}) = \frac{1}{m} h(T^m, \mathcal{C}) = \lim_{n \to \infty} H(\mathcal{C}| \bigvee_{i=1}^n T^{-mi}\mathcal{C}),$$

et comme $H(\mathcal{C}| \bigvee_{i=1}^n T^{-mi}\mathcal{C}) \leq H(\mathcal{C}|T^{-m}\mathcal{C})$ donc

$$h(T) \leq H(\mathcal{C}|T^{-m}\mathcal{C})$$

Or

$$H(\mathcal{C}|T^{-m}\mathcal{C}) \leq \max_{D \in T^{-m}\mathcal{C}} \ln N_{\mathcal{C}}(D)$$

où $N_{\mathcal{C}}(D)$ est le nombre des $C \in \mathcal{C}$ tels que $D \cap C \neq \emptyset$ donc

$$h(T) \leq \min_{m \geq 1} \frac{1}{m} \max_{\mathcal{C} \in \mathcal{P}_N} \max_{D \in T^{-m}\mathcal{C}} \ln N_{\mathcal{C}}(D).$$

Fixons m. Soit $\mathcal{C} \in \mathcal{P}_N$. Les éléments C de $\mathcal{C}$ sont des carrées de côtés $r \leq 1/N$. Les éléments D de $T^{-m}\mathcal{C}$ sont les projections dans le tore $\mathbf{T}^2$ de parallélogrammes de $\mathbf{R}^2$ dont le diamètre est compris entre $K_1 r \left|\lambda_1\right|^m$ et $K_2 r \left|\lambda_1\right|^m$ où $K_1 \leq K_2$ sont des constantes ne dépendant que de la matrice A (K_1 et K_2 dépendent uniquement des axes propres de A). L'aire de D est r^2 et la grande diagonale de D à une longueur d'au moins $K_1 r \left|\lambda_1\right|^m$ donc D est compris dans une bande d'épaisseur $e \leq r^2/(K_1 r \left|\lambda_1\right|^m) = r\left|\lambda_1\right|^{-m}/K_1$. Choisissons m suffisamment grand pour que $\left|\lambda_1\right|^{-m}/K_1 \leq 1$, le nombre $N_{\mathcal{C}}(D)$ d'éléments de $\mathcal{C}$ coupant D est alors au plus

$$\frac{3\sqrt{2}\,\mathrm{diam}(D)}{r} \leq K_3 \left|\lambda_1\right|^m$$

où K_3 ne dépend de K_2. Finalement,

$$h(T) \leq \frac{1}{m} \max_{\mathcal{C} \in \mathcal{P}_N} \max_{D \in T^{-m}\mathcal{C}} \ln N_{\mathcal{C}}(D) \leq |\lambda_1| + \frac{1}{m} \ln K_3. \ \square$$

Remarque Dans le cas général d'un automorphisme du tore $\mathbf{T}^n$ donner par une matrice $A \in SL(n, \mathbf{Z})$ de valeur propre $\lambda_1, ..., \lambda_n$ dont aucune n'est une racine de l'unité, Sinai à démontrer en 1959 la formule

$$h(T) = \sum_{|\lambda_i|>1} \ln |\lambda_i|$$

2. Entropie topologique

2.1. Définition à partir des recouvrements.

Définition 9. *Soit X un espace topologique compact et α, β deux recouvrements ouverts de X.*
1. Le recouvrement
$$\alpha \vee \beta = \{U \cap V : U \in \alpha,\ V \in \beta\}$$
s'appelle le recouvrement engendré par les recouvrements α et β.
2. On dit que le recouvrement β est plus fin que α et on écrit $\alpha < \beta$ si tout élément de β est inclus dans un élément de α.
3. Si $T : X \to X$ est une application continue. $T^{-1}\alpha$ désigne le recouvrement dont les éléments sont les $T^{-1}U$, $U \in \alpha$.

Définition 10. *Soit X un espace topologique compact non vide et α un recouvrement ouvert de X. Désignons par $N(\alpha)$ le nombre minimum d'éléments d'un sous recouvrement de α. Le nombre $H(\alpha) = \ln N(\alpha)$ s'appelle l'entropie de α.*

Proposition 14. *Soit X un espace topologique compact (non vide) et α, β deux recouvrements ouverts de X.*
1. $H(\alpha) \geq 0$.
2. $H(\alpha) = 0$ ssi $X \in \alpha$.
3. Si $\alpha < \beta$ alors $H(\alpha) < H(\beta)$.
4. $H(\alpha \vee \beta) \leq H(\alpha) + H(\beta)$.

Dem. Seul 4 n'est pas évident. Soit α' et β' des sous recouvrements de α et β de cardinal minimum. Le recouvrement $\alpha' \vee \beta'$ à un cardinal inférieur à $N(\alpha)N(\beta)$. □

Proposition 15. *Soit X un espace topologique compact, $T : X \to X$ une application continue et α recouvrement ouvert de X.*
1. $H(T^{-1}\alpha) \leq H(\alpha)$.
2. Si T est surjective alors $H(T^{-1}\alpha) = H(\alpha)$.

Dem. 1 est évident. Supposons T surjective. Soit α' une partie de α tel que $T^{-1}\alpha'$ recouvre X. Si $y \in X$ il existe $x \in X$ tel que $Tx = y$. Comme $T^{-1}\alpha'$ recouvre X, il existe $U \in \alpha'$ tel que $x \in T^{-1}U$. On a alors $y = Tx \in U$. α' est donc un recouvrement de X. □

Proposition 16. *Soit X un espace topologique compact, $T : X \to X$ une application continue et α recouvrement ouvert de X. Alors la limite $\lim_{n\to\infty} \frac{1}{n} H(\alpha \vee T^{-1}\alpha \vee ...vT^{-n+1}\alpha)$ existe.*

Dem. Comme pour l'entropie mesuré , il suffit de prouver que la suite $(H(\alpha \vee T^{-1}\alpha \vee ...vT^{-n+1}\alpha))_{n\geq 1}$ est sous additive. Pour tout entier $m \geq 1$ et tout entier $n \geq 1$ on a

$$\begin{aligned} H(\alpha \vee T^{-1}\alpha \vee ... \vee T^{-n-m+1}\alpha) &= H((\alpha \vee ... \vee T^{-n+1}\alpha) \vee T^{-n}(\alpha \vee T^{-1}\alpha \vee ...T^{-n+1})) \\ &\leq H(\alpha \vee ... \vee T^{-n+1}\alpha) + H(T^{-n}(\alpha \vee T^{-1}\alpha \vee ...T^{-m+1})) \\ &\leq H(\alpha \vee ... \vee T^{-n+1}\alpha) + H(\alpha \vee T^{-1}\alpha \vee ...T^{-m+1}). \ \square \end{aligned}$$

Définition 11. *Soit X un espace topologique compact, $T : X \to X$ une application continue.*
1. Soit α un recouvrement ouvert de X. Le nombre $H(T,\alpha) = \lim_{n\to\infty} \frac{1}{n} H(\alpha \vee T^{-1}\alpha \vee ...vT^{-n+1}\alpha)$ s'appelle l'entropie de T relativement au recouvrement α.
2. L'entropie topologique de T est le borne supérieure

$$\sup_{\alpha} H(T,\alpha)$$

où α parcourt tous les recouvrement ouvert de X. On note l'entropie topologique $h_{top}(T)$.

Lemme 5. *Soit X un espace topologique compact, $T : X \to X$ une application continue et α, β deux recouvrements ouverts de X. Si $\alpha < \beta$ alors $H(T,\alpha) \leq H(T,\beta)$.*

Dem. Pour tout n, $\alpha \vee ... \vee T^{-n+1}\alpha < \beta \vee ... \vee T^{-1}\beta$ d'où le résultat. □

Théorème 14. *Soit X_1 et X_2 deux espaces topologique compacts et $T_i : X_i \to X_i$, $i = 1, 2$, deux applications continues. Si il existe une application continue surjective $\phi : X_1 \to X_2$ telle que $\phi \circ T_1 = T_2 \circ \phi$ alors $h_{top}(T_1) \geq h_{top}(T_2)$. Si de plus ϕ est un homéomorphisme alors $h_{top}(T_1) = h_{top}(T_2)$.*

Dem. On vérifie facilement que si α est un recouvrement de X_2 alors $H(T_2,\alpha) = H(T_1, \phi^{-1}\alpha)$ donc $h_{top}(T_2) \leq h_{top}(T_1)$. Si T est un homéomorphisme on a aussi l'inégalité contraire. □

Théorème 15. *Soit X un espace topologique compact et $T : X \to X$ une application continue..*
1. Pour tout entier naturel k, on a $h_{top}(T^k) = k h_{top}(T)$.
2. Si de plus T est un homéomorphisme alors pour tout entier relatif k, $h_{top}(T^k) = |k|\, h_{top}(T)$.

Dem. 1. Soit α un recouvrement ouvert de X et $\beta = \alpha \vee ... \vee T^{-k+1}\alpha$. On a

$$\bigvee_{i=0}^{kn-1} T^{-i}\alpha = \bigvee_{i=0}^{n} T^{-ki}\beta$$

donc

$$\lim_{m\to\infty} \frac{1}{m} H(\bigvee_{i=0}^{m-1} T^{-i}\alpha) = \lim_{n\to\infty} \frac{1}{nk} H(\bigvee_{i=0}^{nk-1} T^{-i}\alpha) = \lim_{n\to\infty} \frac{1}{nk} H(\bigvee_{i=0}^{n-1} T^{-ki}\beta) \leq \frac{1}{k} h_{top}(T^k).$$

Par conséquent, $k h_{top}(T) \leq h_{top}(T^k)$. Inversement, le recouvrement $\bigvee_{i=0}^{kn-1} T^{-i}\alpha$ est plus fin que le recouvrement $\bigvee_{i=0}^{n-1} T^{-ki}\alpha$, donc

$$\frac{1}{n} H(\bigvee_{i=0}^{n-1} T^{-ki}\alpha) \leq \frac{1}{n} H(\bigvee_{i=0}^{kn-1} T^{-i}\alpha) = k \times \frac{1}{kn} H(\bigvee_{i=0}^{kn-1} T^{-i}\alpha).$$

En passant à la limite, on obtient $h(T^k,\alpha) \leq kh(T,\alpha)$ puis $h_{top}(T^k) \leq k h_{top}(T)$.
2. Il suffit de prouver que $h(T) = h(T^{-1})$. Soit α un recouvrement ouvert de X. On a

$$\begin{aligned} H(\alpha \vee T\alpha \vee ... \vee T^{n-1}\alpha) &= H(T^{n-1}(\alpha \vee T^{-1}\alpha \vee ... \vee T^{-n+1}\alpha)) \\ &= H(\alpha \vee T^{-1}\alpha \vee ... \vee T^{-n+1}\alpha) \end{aligned}$$

et le résultat en découle. □

2.2. Définitions de Bowen.

Définition 12. *Soit (X,d) un espace métrique compact, $T : X \to X$ une application continue, ε un nombre réel > 0 et n un entier naturel.*
1. Une partie E de X est dite (n,ε)-recouvrante si pour tout y de X, il existe $x \in E$ tel que $d(T^i x, T^i y) < \varepsilon$, $i = 0, ..., n-1$.
2. Une partie F de X est dite (n,ε)-séparante si pour tout couple (x,y) d'élément de F on a $d(T^i x, T^i y) \geq \varepsilon$ pour au moins un $i \in \{0, ..., n-1\}$.

Remarques. 1. En considérant la distance d_n sur X définie par

$$d_n(x,y) = \max_{i=0}^{n-1} d(T^i x, T^i y)$$

on voit que :
- une partie E de X est (n,ε)-recouvrante si pour tout y de X il existe $x \in E$ tel que $d_n(x,y) < \varepsilon$,
- une partie F de X est (n,ε)-séparante si pour tout couple (x,y) d'éléments de F, on a $d_n(x,y) \geq \varepsilon$.
2. On peut considerer les recouvrements α dont les éléments ont un diamètre pour la distance d_n inférieure à ε, $r(n,\varepsilon)$ est alors le minimum des cardinaux de ces recouvrements.

Notation. Soit (X,d) un espace métrique compact, $T : X \to X$ une application continue, ε un nombre réel > 0 et n un entier naturel. On pose

$$r(n,\varepsilon) = \min\{\mathrm{card}\, E : \ E \text{ est } (n,\varepsilon)\text{-recouvrante}\}$$

et

$$s(n,\varepsilon) = \max\{\mathrm{card}\, E : \ E \text{ est } (n,\varepsilon)\text{-séparante}\}.$$

Théorème 16. ***(Bowen-Dinaburg)*** *Soit (X,d) un espace métrique compact, $T : X \to X$ une application continue. On a*

$$\begin{aligned} h_{top}(T) &= \lim_{\varepsilon\to 0,\ \varepsilon>0} \limsup_{n\to\infty} \frac{1}{n} \ln r(n,\varepsilon) \\ &= \lim_{\varepsilon\to 0,\ \varepsilon>0} \limsup_{n\to\infty} \frac{1}{n} \ln s(n,\varepsilon). \end{aligned}$$

Dem. 1. Montrons d'abord l'égalité

$$\lim_{\varepsilon\to 0,\ \varepsilon>0} \limsup_{n\to\infty} \frac{1}{n} \ln r(n,\varepsilon) = \lim_{\varepsilon\to 0,\ \varepsilon>0} \limsup_{n\to\infty} \frac{1}{n} \ln s(n,\varepsilon).$$

Soit F_n un ensemble (n,ε)-séparant de cardinal maximal. Pour tout $y \in X$, l'ensemble $F_n \cup \{y\}$ n'est plus séparant donc il existe $x \in F_n$ tel que $d_n(x,y) < \varepsilon$, donc F_n est recouvrant et

$$r(n,\varepsilon) \leq s(n,\varepsilon).$$

Soit E_n un ensemble $(n,\varepsilon/2)$-recouvrant et F_n un ensemble (n,ε)-séparant. Tout point de F_n est dans une boule ouverte pour la distance d_n centré en un point de E_n et de rayon $\varepsilon/2$. Comme une telle boule contient au plus un point de F_n le cardinal de F_n est inférieur à celui de E_n, d'où

$$s(n,\varepsilon) \leq r(n,\varepsilon/2)$$

et l'égalité précédente en découle.
2. Montrons que

$$h_{top}(T) = \lim_{\varepsilon \to 0,\ \varepsilon > 0} \lim \sup_{n \to \infty} \frac{1}{n} \ln r(n, \varepsilon).$$

Soit $\varepsilon > 0$ et α un recouvrement ouvert de X dont tous les ouverts ont un diamètre $\leq \varepsilon$. Si F_n est (n, ε)-séparant alors pour tout $x \neq y \in F_n$ il existe $k \leq n - 1$ tel que $d(T^k x, T^k y) \geq \varepsilon$ donc il n'existe pas de U appartenant à α contenant simultanément $T^k x$ et $T^k y$. Ainsi un ouvert $U_0 \cap T^{-1} U_1 \cap ... \cap T^{-n+1} U_{n+1}$ de $\alpha \vee T^{-1}\alpha \vee ... \vee T^{-n+1}\alpha$ ne peut pas contenir plus d'un point de F_n donc

$$\operatorname{card} F_n \leq N(\alpha \vee T^{-1}\alpha \vee ... \vee T^{-n+1}\alpha).$$

Il en résulte que

$$s(n, \varepsilon) \leq N(\alpha \vee T^{-1}\alpha \vee ... \vee T^{-n+1}\alpha),$$

puis que

$$\lim_{\varepsilon \to 0,\ \varepsilon > 0} \lim \sup_{n \to \infty} \frac{1}{n} \ln s(n, \varepsilon) \leq h_{top}(T).$$

Inversement soit α un recouvrement ouvert de X. Ce recouvrement à un nombre de Lebesgue $\varepsilon > 0$. Pour chaque entier n, soit E_n un ensemble (n, ε)-recouvrant de cardinal minimal. Pour chaque $x \in F$ et chaque $k \leq n - 1$ il existe un ouvert $U_{x,k} \in \alpha$ tel que la boule $B(T^k x, \varepsilon)$ soit incluse dans $U_{x,k}$. Pour tout $y \in X$, il existe $x \in E_n$ tel que $d_n(x, y) \leq \varepsilon$ donc $y \in U_{x,0} \cap T^{-1} U_{x,1} \cap ... \cap T^{-n+1} U_{x,n-1}$. Par conséquent,

$$\{U_{x,0} \cap T^{-1} U_{x,1} \cap ... \cap T^{-n+1} U_{x,n-1} : x \in E_n\}$$

est un sous recouvrement de $\alpha \vee ... \vee T^{-n+1}\alpha$ et

$$N(\alpha \vee ... \vee T^{-n+1}\alpha) \leq \operatorname{card} E_n. \ \square$$

Remarque. La définition de Bowen de l'entropie topologique semble dépendre de la distance grâce au théorème précédent on voit qu'elle ne dépend que de la topologie.

Exercice Les limites $\lim_{n \to \infty} \frac{1}{n} \ln s(n, \varepsilon)$ et $\lim_{n \to \infty} \frac{1}{n} \ln r(n, \varepsilon)$ n'existent pas toujours. On peut néanmoins introduire un troisième nombre $q(n, \varepsilon)$ pour lequel la limite $\lim_{n \to \infty} \frac{1}{n} \ln r(n, \varepsilon)$ existe. Appelons $\operatorname{diam}_n A$ le diamètre d'une partie A mesuré avec la distance d_n ; $q(n, \varepsilon)$ est alors le minimum des cardinaux des recouvrements $\mathcal{E} = \{E_1, ..., E_k\}$ de X par des parties $E_1, ..., E_k$ telles que $\operatorname{diam}_n E_i < \varepsilon$, pour $i = 1, ..., k$ (les parties E_i sont quelconques). Pouver que $q(n + m, \varepsilon) \leq q(n, \varepsilon) \times q(m, \varepsilon)$ (Cela se montre à l'aide de l'observation :
si le $\operatorname{diam}_n A \leq \varepsilon$ et $\operatorname{diam}_m B \leq \varepsilon$ alors $\operatorname{diam}_{n+m} A \cap T^{-n} B \leq \varepsilon$). Conclure.

Proposition 17. *Soit X_1 et X_2 deux espaces métriques compacts et $T_i : X_i \to X_i$, $i = 1, 2$, deux applications continues. Appelons $T : X_1 \times X_2 \to X_1 \times X_2$ l'application définie par $T(x_1, x_2) = (T_1 x_1, T_2 x_2)$. Alors*

$$h_{top}(T) = h_{top}(T_1) + h_{top}(T_2).$$

Dem : exercice.
1. Montrer en utilisant les ensembles (n, ε) recouvrant que $h_{top}(T) \leq h_{top}(T_1) + h_{top}(T_2)$.
2. Montrer en utilisant les ensembles (n, ε) séparant que $h_{top}(T) \geq h_{top}(T_1) + h_{top}(T_2)$.

2.3. Relation avec l'entropie mesurée. **Notation rappel** : Soit X un espace topologique et $T : X \to X$ une application continue. On désigne par $P(X,T)$ l'ensemble des probabilités μ définie sur la tribu borélienne de X invariantes par T.

Théorème 17. ***Principe variationnel.*** *Soit X un espace métrique compact et $T : X \to X$ une application continue. Alors*

$$h_{top}(T) = \sup_{\mu \in P(X,T)} h_\mu(T).$$

Dem. **Etape 1.** Montrons que l'entropie topologique est supérieure à l'entropie relativement à n'importe quelle probabilté invariante. Pour cela il suffit de montrer que si μ est une mesure invariante, $\mathcal{C} = \{C_1, ..., C_k\}$ une partition mesurable finie de X alors pour tout entier $n \geq 1$,

$$H_\mu(T, \mathcal{C}) \leq h_{top}(T) + nK$$

où K est une constante ne dépendant ni de n ni de $\mathcal{C}$ ni de μ ni de T. En effet, cela montre que pour tout entier p on a

$$ph_\mu(T) = h_\mu(T^p) \leq h_{top}(T^p) + K = ph_{top}(T) + K$$

d'où $h_\mu(T) \leq h_{top}(T)$.
Soit $\varepsilon > 0$. Par régularité, il existe des compacts $F_1, ..., F_k$ tels que pour tout i, F_i soit inclus dans C_i et $\sum_{i=1}^k \mu(C_i \backslash F_i) \leq \varepsilon$. La partie $U = X \backslash (F_1 \cup F_2 ... \cup F_k)$ est un ouvert de X et pour chaque i, $U_k = F_k \cup U$ est aussi un ouvert. Considérons le recouvrement ouvert $\alpha = \{U_1, ..., U_k\}$ et la partition $\mathcal{D} = \{F_0 = U, F_1, ..., F_k\}$. On a

$$\begin{aligned}
H_\mu(\mathcal{C}|\mathcal{D}) &= \sum_{i=0}^k \sum_{j=1}^k \mu(C_j \cap F_i) \ln \mu(C_j|F_i) \\
&= \sum_{j=1}^k \mu(C_j \cap F_0) \ln \mu(C_j|F_0) + \sum_{i=1}^k \mu(F_i) \ln \mu(C_i|F_i) \\
&\leq \mu(F_0) \sum_{j=1}^k \mu(C_j|F_0) \ln \mu(C_j|F_0) + 0 \\
&\leq \varepsilon \ln k.
\end{aligned}$$

Notons $\mathcal{D}_n = \mathcal{D} \vee T^{-1}\mathcal{D} \vee ... \vee T^{-n+1}\mathcal{D}$ et $\alpha_n = \alpha \vee T^{-1}\alpha \vee ... \vee T^{-n+1}\alpha$. On sait que

$$H_\mu(\mathcal{D}_n) \leq \ln \operatorname{card} \mathcal{D}_n.$$

. Il nous reste maintenant à majorer $N(\mathcal{D}_n)$ à l'aide de $N(\alpha_n)$. Soit $A = U_{i_0} \cap T^{-1}U_{i_1} \cap ... \cap T^{-n+1}U_{i_{n-1}}$ un élément de α_n. On a

$$A = \bigcap_{p=0}^{n-1} T^{-p}(U \cup F_{i_p})$$

qui est la réunion d'au plus 2^n parties du type $\bigcap_{p=0}^{n-1} T^{-p}(B_p)$ où $B_p = U$ ou F_{i_p}. Toutes ces parties appartient à $\mathcal{D}_n$. A chaque $A \in \alpha_n$ on peut donc faire correspondre une partie $\mathcal{D}_A$ de $\mathcal{D}_n$ de cardinal $\leq 2^n$ telle que la réunion des éléments de $\mathcal{D}_A$ soit A. Soit β un sous recouvrement de α_n. Comme β recouvre X et que $\mathcal{D}_n$ est une partition de X, l'ensemble des $\mathcal{D}_A$, $A \in \beta$, recouvre $\mathcal{D}_n$. Par conséquent, $\operatorname{card} \beta \times 2^n \geq \operatorname{card} \mathcal{D}_n$. On en déduit que $N(\alpha)2^n \geq \operatorname{card} \mathcal{D}_n$. Finalement,

$$H(\mathcal{D}_n) \leq \ln \operatorname{card} \mathcal{D}_n \leq \ln(2^n N(\alpha_n)) \leq n \ln 2 + \ln N(\alpha_n)$$

et

$$h_\mu(T, \mathcal{D}) \leq h_{top}(T) + \ln 2.$$

Mais $h_\mu(T, \mathcal{C}) \leq h_\mu(T, \mathcal{D}) + H_\mu(\mathcal{C}|\mathcal{D})$, d'où

$$h_\mu(T, \mathcal{C}) \leq h_{top}(T) + \ln 2 + \varepsilon \ln k.$$

Pour conclure, il suffit donc de choisir ε tel que $\varepsilon \ln k \leq 1$.

Etape 2. Deux lemmes sont utiles.

Lemme 6. *Soit X un espace métrique compact muni de sa tribu borélienne et μ une probabilité sur X. Alors pour tout $\varepsilon > 0$, il existe une partition $\mathcal{C}$ de X dont tous les éléments ont un diamètre $\leq \varepsilon$ et une frontière de mesure nulle.*

Dem : exercice.

Lemme 7. *Soit $(X, \mathcal{A})$ un espace mesuré et $\mathcal{C}$ une partition mesurable de X. Alors l'application qui à une probabilité μ sur X associe l'entropie $H_\mu(\mathcal{C})$ est concave.*

Dem : la fonction $\phi(t) = -t \ln t$ est concave.

Fin de la démonstration du théorème. Nous devons construire des probabilité dont l'entropie approche l'entropie topologique de T. Fixons $\varepsilon > 0$. Pour chaque entier n, soit E_n une partie de cardinal maximal X telle que pour tout $x \neq y$ appartenant à E, $d_n(x, y) = \sup_{k=0}^{n-1} d(T^k x, T^k y) \geq \varepsilon$. Désignons par δ_x la mesure de Dirac en x. Comme la partie E_n est finie, la mesure

$$\sigma_n = \frac{1}{\operatorname{card} E_n} \sum_{x \in E_n} \delta_{x_n}$$

est une probabilité sur X. Considérons la probabilité

$$\mu_n = \frac{1}{n} \sum_{k=0}^{n-1} \sigma_n \circ T^{-k}.$$

Choisissons une suite d'entiers $(n_p)_{p \in \mathbf{N}}$ telle que $\lim_{p \to \infty} \frac{1}{n_p} \ln \operatorname{card} E_{n_p} = \limsup_{n \to \infty} \frac{1}{n} \ln \operatorname{card} E_n$. Par compacité, avec une nouvelle extraction, on peut supposer que la suite $(\mu_{n_p})_{p \in \mathbf{N}}$ converge vers une probabilité μ. Par définition des mesures μ_n, cette probabilité est T-invariante. Grâce à la définition de Bowen de l'entropie topologique, il suffit de prouver l'inégalité $h_\mu(T) \geq \limsup_{n \to \infty} \frac{1}{n} \ln \operatorname{card} E_n$. Choisissons une partition $\mathcal{C} = \{C_1, ..., C_k\}$ de X telle que $\operatorname{diam} C_i < \varepsilon$ et $\mu(\partial C_i) = 0$, $i = 1, ..., k$. Nous allons montrer que

$$h_\mu(T, \mathcal{C}) \geq \limsup_{n \to \infty} \frac{1}{n} \ln \operatorname{card} E_n.$$

Pour tout entier m, notons $\mathcal{C}_m = \mathcal{C} \vee T^{-1}\mathcal{C} \vee ... \vee T^{-m+1}\mathcal{C}$. Chaque élément de $\mathcal{C}$ est une partie de diamètre $< \varepsilon$ et pour chaque couple (x, y) d'éléments de E_n, il existe $k \leq n-1$

tel que $d(T^k x, T^k y) \geq \varepsilon$ donc il existe un $k \leq n-1$ tel que $T^k x$ et $T^k y$ n'appartiennent pas au même élément de $\mathcal{C}$. Deux tels points ne peuvent donc appartenir au même élément de $\mathcal{C}_n$. Par conséquent, si A est un élément de $\mathcal{C}_n$, on a $\sigma_n(A) = \frac{1}{\operatorname{card} E_n}$ ou 0, d'où $H_{\sigma_n}(\mathcal{C}_n) = \ln \operatorname{card} E_n$.
Fixons un entier $q \leq n$.
Comme l'application $\nu \to H_\nu(\mathcal{C}_q)$ est convexe, on a

$$H_{\mu_n}(\mathcal{C}_q) \geq \frac{1}{n} \sum_{k=0}^{n-1} H_{\sigma_n \circ T^{-k}}(\mathcal{C}_q).$$

De plus, chaque entier k s'écrit de manière unique sous la forme $p = rq + j$ où $j \in \{0, ..., q-1\}$ donc

$$H_{\mu_n}(\mathcal{C}_q) \geq \frac{1}{n} \sum_{j=0}^{q-1} \sum_{r \geq 0\,:\, rq+j \leq n-1} H_{\sigma_n \circ T^{-rq-j}}(\mathcal{C}_q) = \frac{1}{n} \sum_{j=0}^{q-1} \sum_{r \geq 0\,:\, rq+j \leq n-1} H_{\sigma_n}(T^{-rq-j}\mathcal{C}_q).$$

Or

$$\sum_{j=0}^{q-1} \sum_{r \geq 0\,:\, rq+j \leq n-1} H_{\sigma_n}(T^{-rq-j}\mathcal{C}_q) \geq \sum_{j=0}^{q-1} H_{\sigma_n}(\bigvee_{r \geq 0\,:\, rq+j \leq n-1} T^{-rq-j}\mathcal{C}_q)$$

donc

$$H_{\mu_n}(\mathcal{C}_q) \geq \frac{1}{n} \sum_{j=0}^{q-1} H_{\sigma_n}(\mathcal{C}_n^j)$$

où $\mathcal{C}_n^j = \bigvee_{r \geq 0\,:\, rq+j \leq n-1} T^{-rq-j}\mathcal{C}$. Les partitions $\mathcal{C}_n^j$ ne diffèrent pas trop de la partition $\mathcal{C}_n$:

$$\mathcal{C}_n \prec \mathcal{C}_n^j \vee \mathcal{C}_q \ ;$$

Par conséquent,

$$\begin{aligned} H_{\sigma_n}(\mathcal{C}_n) &\leq H_{\sigma_n}(\mathcal{C}_n^j) + H_{\sigma_n}(\bigvee_{i=0}^{q-1} T^{-i}\mathcal{C}) \\ &\leq H_{\sigma_n}(\mathcal{C}_n^j) + q \ln k \end{aligned}$$

et

$$H_{\mu_n}(\mathcal{C}_q) \geq \frac{1}{n} \sum_{j=0}^{q-1} H_{\sigma_n}(\mathcal{C}_n^j) \geq \frac{1}{n} \sum_{j=0}^{q-1} (H_{\sigma_n}(\mathcal{C}_n) - 2q \ln k) = \frac{q}{n} \ln \operatorname{card} E_n - \frac{2q \ln k}{n}.$$

Montrons que pour tout q, $\lim_{p \to \infty} H_{\mu_{n_p}}(\mathcal{C}_q) = H_\mu(\mathcal{C}_q)$. Pour cela il suffit de montrer que pour chaque $C \in \mathcal{C}$, $\mu(\partial C) = 0$. Or si $C = C_{i_0} \cap T^{-1}C_{i_1} \cap ... \cap T^{-(q-1)}C_{i_{q-1}}$ alors

$$\begin{aligned} \partial C &\subset \partial C_{i_0} \cup \partial T^{-1} C_{i_1} \cup ... \cup \partial T^{-(q-1)} C_{i_{q-1}} \\ &\subset \partial C_{i_0} \cup T^{-1} \partial C_{i_1} \cup ... \cup T^{-(q-1)} \partial C_{i_{q-1}} \end{aligned}$$

et $\mu(\partial C) = 0$ d'après le choix de la partition $\mathcal{C}$ et la T-invariance de μ. Finalement, pour chaque q fixé, on a

$$\begin{aligned} H_\mu(\mathcal{C}_q) &\geq \lim_{p \to \infty} (\frac{q}{n_p} \ln \operatorname{card} E_{n_p} - \frac{2q \ln k}{n_p}) = q \limsup_{n \to \infty} \frac{1}{n} \ln \operatorname{card} E_n \\ &= q \limsup_{n \to \infty} \frac{1}{n} \ln s(n, \varepsilon), \end{aligned}$$

d'où
$$h_\mu(T) \geq h_\mu(T,\mathcal{C}) = \lim_{q\to\infty} \frac{1}{q} H_\mu(\mathcal{C}_q) \geq \lim \sup_{n\to\infty} \frac{1}{n} \ln r(n,\varepsilon). \square$$

Corollaire 6. *Soit X_1 et X_2 deux espaces métriques compacts munis de leurs tribus boréliennes et $T_i : X_i \to X_i$, $i = 1,2$, deux applications continues. Si il existe une bijection bimesurable $\phi : X_1 \to X_2$ telle que $T_2 \circ \phi = \phi \circ T_1$ alors $h_{top}(T_1) = h_{top}(T_2)$.*

Dem. Il suffit de remarquer qu'une probabilité μ sur X_1 est T_1-invariante ssi son image par ϕ, $\mu \circ \phi^{-1}$, est T_2-invariante. Pour tout $\mu \in P(X_1,T_1)$, ϕ est donc un isomorphisme entre les systèmes dynamiques mesurés (X_1,μ,T_1) et $(X_2,\mu\circ\phi^{-1},T_2)$ et $h_\mu(T_1) = h_{\mu\circ\phi^{-1}}(T_2)$. Grâce au principe variationnel, en passant à la borne supérieure, on en déduit que $h_{top}(T_1) = h_{top}(T_2)$. $\square$

Corollaire 7. *Soit X un espace métrique compact et $T : X \to X$ une application continue.*
1. $h_{top}(T) = h_{top}(T_{|\Omega(T)})$.
2. Soit $X_\infty = \cap_{n\in\mathbf{N}} T^n X$. Alors $h_{top}(T) = h_{top}(T_{|X_\infty})$.

Dem. 1. Toutes les probabilités T-invariantes sont portées par $\Omega(T)$ donc si μ est une probabilité T-invariante on a $h_\mu(T) = h_{\mu_{|\Omega(T)}}(T_{|\Omega(T)})$. On conclut grâce au principe variationnel.
2. Si μ est une probabilité T-invariante alors pour tout $n \in \mathbf{N}$, $\mu(T^nX) = \mu(T^{-n}(T^nX)) = \mu(X)$ donc μ est portée par T^nX. On conclut comme précédemment. $\square$

2.4. L'application entropie d'une mesure.

Théorème 18. *Soit X un espace métrique compact et $T : X \to X$ une application continue. Alors l'application $\mu \in P(X,T) \to h_\mu(T) \in \mathbf{R}$ est affine.*

Dem. Notons ϕ l'application $x \in [0,1] \to x\ln x$. Soit μ_1 et μ_2 deux probabilités T-invariante et λ_1 et λ_2 deux réels ≥ 0 de somme 1. Comme ϕ est concave, pour tout borélien B de X, on a

$$\begin{aligned}
0 &\leq \phi(\lambda_1\mu_1(B)+\lambda_2\mu_2(B)) - \lambda_1\phi(\mu_1(B)) - \lambda_2\phi(\mu_2(B)) \\
&= -(\lambda_1\mu_1(B)+\lambda_2\mu_2(B))\ln(\lambda_1\mu_1(B)+\lambda_2\mu_2(B)) + \lambda_1\mu_1(B)\ln\mu_1(B) + \lambda_2\mu_2(B)\ln\mu_2(B) - \\
&= -\lambda_1\mu_1(B)(\ln(\lambda_1\mu_1(B)+\lambda_2\mu_2(B)) - \ln\lambda_1\mu_1(B)) \\
&\quad -\lambda_2\mu_2(B)(\ln(\lambda_1\mu_1(B)+\lambda_2\mu_2(B)) - \ln\lambda_2\mu_2(B)) \\
&\quad +\lambda_1\mu_1(B)(\ln\mu_1(B) - \ln\lambda_1\mu_1(B)) + \lambda_2\mu_2(B)(\ln\mu_2(B) - \ln\lambda_2\mu_2(B)) \\
&\leq 0 + 0 - \lambda_1\mu_1(B)\ln\lambda_1 - \lambda_2\mu_2(B)\ln\lambda_2
\end{aligned}$$

où la deuxième inégalité résulte de croissance la fonction logarithme. Si $\mathcal{C}$ est une partition mesurable finie, en sommant sur tout les éléments de la partition on obtient

$$\begin{aligned}
0 &\leq H_{\lambda_1\mu_1+\lambda_2\mu_2}(\mathcal{C}) - \lambda_1 H_{\mu_1}(\mathcal{C}) - \lambda_2 H_{\mu_2}(\mathcal{C}) \\
&\leq -\lambda_1\ln\lambda_2 - \lambda_2\ln\lambda_2 \leq \ln 2.
\end{aligned}$$

En prenant une partition de la forme $\mathcal{C}\vee\ldots\vee T^{-n+1}\mathcal{C}$ et en passant à la limite on obtient

$$h_{\lambda_1\mu_1+\lambda_2\mu_2}(T,\mathcal{C}) = \lambda_1 h_{\mu_1}(T,\mathcal{C}) + \lambda_2 h_{\mu_2}(T,\mathcal{C}).$$

En passant à la borne supérieure sur l'ensemble des partitions on obtient seulement l'inégalité $h_{\lambda_1\mu_1+\lambda_2\mu_2}(T) \leq \lambda_1 h_{\mu_1}(T)+\lambda_2 h_{\mu_2}(T)$. Soit $\mathcal{C}_1$ et $\mathcal{C}_2$ deux partitions mesurables finies. On a

$$\begin{array}{rcl} h_{\lambda_1\mu_1+\lambda_2\mu_2}(T) & \geq & h_{\lambda_1\mu_1+\lambda_2\mu_2}(T, \mathcal{C}_1 \vee \mathcal{C}_2) \\ & = & \lambda_1 h_{\mu_1}(T, \mathcal{C}_1 \vee \mathcal{C}_2) + \lambda_2 h_{\mu_2}(T, \mathcal{C}_1 \vee \mathcal{C}_2) \\ & \geq & \lambda_1 h_{\mu_1}(T, \mathcal{C}_1) + \lambda_2 h_{\mu_2}(T, \mathcal{C}_2) \end{array}$$

et l'inégalité inverse s'en déduit en passant au sup sur $\mathcal{C}_1$ et $\mathcal{C}_2$. $\square$

Corollaire 8. *Soit X un espace métrique compact et $T : X \to X$ une application continue. Alors l'ensemble $P_{\max}(X, T)$ des probabilité μ T-invariantes telles que $h_\mu(T) = h_{top}(T)$ est convexe.*

Remarque. Soit X un espace métrique compact et $T : X \to X$ une application continue. En utilisant le théorème de représentation intégrale de Choquet, on peut démontrer les résultats suivants (cf Walters page 186 et 191) :
1. Les points extrémaux $P_{\max}(X, T) \neq \emptyset$ sont des probabilités ergodiques.
2. Si $h_{top}(T) < \infty$ et si $P_{\max}(X, T) \neq \emptyset$ alors $P_{\max}(X, T)$ contient une mesure ergodique.

2.5. Applications expansives.

Définition 13. *Soit X un espace métrique compact et $T : X \to X$ une application continue.*
1. On dit que T est expansive si il existe un réel $\delta > 0$ tel que pour tout $x, y \in X$, si $d(T^n x, T^n y) \leq \delta$ pour tout $n \in \mathbf{N}$ alors $x = y$.
2. Si T est un homéomorphisme on dit que T est expansif si il existe un réel $\delta > 0$ tel que pour tout $x, y \in X$, si $d(T^n x, T^n y) \leq \delta$ pour tout $n \in \mathbf{Z}$ alors $x = y$.

Exemples 1. Les décalages unilatéral et bilatéral sur $\mathcal{A}^{\mathbf{N}}$ et $\mathcal{A}^{\mathbf{Z}}$ où $\mathcal{A}$ est un ensemble fini.
2. Toute restriction d'une application expansive à un fermé stable est expansive. Toute restriction d'un homéomorphisme expansif à un fermé invariant est expansif. Tous les sous décalages sont expansifs.
3. Soit $A \in SL_2(\mathbf{Z})$ et $T_A : \mathbf{T}^2 = \mathbf{R}^2/\mathbf{Z}^2 \to \mathbf{T}^2$ l'application associée à la matrice A. Si les valeurs propres de A ne sont pas de modules 1 alors T_A est expansive **(Exercice)**.

Proposition 18. *Soit X un espace métrique compact et $T : X \to X$ une application continue expansive ou un homéomorophisme expansif. Soit $\delta > 0$. Supposons dans le cas d'une application, que pour tout $x, y \in X$, $d(T^n x, T^n y) \leq \delta$ pour tout $n \in \mathbf{N}$ entraine $x = y$ et dans le cas d'un homéomorphisme, que pour tout $x, y \in X$, $d(T^n x, T^n y) \leq \delta$ pour tout $n \in \mathbf{Z}$ entraine $x = y$.*
1.

$$h_{top}(T) = \limsup_{n\to\infty} \frac{1}{n} \ln s(n, \delta)$$

où $s(n, \varepsilon)$ est le maximum des cardinaux des parties (n, ε) séparantes.
2. Si α est un recouvrement ouvert dont tous les ouverts ont un diamètre $< \delta$ alors

$$h_{top}(T) = \lim_{n\to\infty} \frac{1}{n} \ln N(\alpha \vee T^{-1}\alpha ... \vee T^{-n+1}\alpha)$$

où $N(\alpha \vee T^{-1}\alpha ... \vee T^{-n+1}\alpha)$ désigne le minimum des cardinaux des sous recouvrements de $\alpha \vee T^{-1}\alpha ... \vee T^{-n+1}\alpha$.

Dem. 1. Faisons la démonstration dans le cas des applications. Rappelons que $d_n(x,y) = \max_{i=0}^n d(T^i x, T^i y)$.
Montrons que pour tout $\varepsilon > 0$, il existe un entier $n_\varepsilon \geq 0$ tel que pour tout $x, y \in X$, $d(x,y) \geq \varepsilon$ entraine $d_{n_\varepsilon}(x,y) \geq \delta$. En effet, soit K_ε l'ensemble des couples (x,y) de X tel que $d(x,y) \geq \varepsilon$ et $U_n = \{(x,y) \in X : d_n(x,y) > \delta\}$. Les U_n sont des ouverts et par définition de l'expansivité, leur réunion, $\cup_{n\in\mathbf{N}} U_n$, recouvre le complémentaire de la diagonale dans $X \times X$. La compacité de K_ε montre qu'il existe n_ε tel que K_ε soit inclus dans U_{n_ε}.
Si x et y sont deux points tels que $d_n(x,y) \geq \varepsilon$ alors $d_{n+n_\varepsilon}(x,y) \geq \delta$, par conséquent tout ensemble (n,ε) séparant est $(n+n_\varepsilon, \delta)$ séparant, d'où

$$s(n,\varepsilon) \leq s(n+n_\varepsilon, \delta).$$

Finalement,

$$\limsup_{n\to\infty} \frac{1}{n} \ln s(n,\varepsilon) \leq \limsup_{n\to\infty} \frac{1}{n} \ln s(n+n_\varepsilon, \delta) = \limsup_{n\to\infty} \frac{1}{n} \ln s(n,\delta).$$

2. Soit α un recouvrement ouvert de X dont tous les ouverts ont un diamètre $< \delta$. Si F_n est (n,δ)-séparant alors pour tout $x \neq y \in F_n$ il existe $k \leq n-1$ tel que $d(T^k x, T^k y) \geq \delta$ donc il n'existe pas de U appartenant à α contenant simultanément $T^k x$ et $T^k y$. Ainsi un ouvert $U_0 \cap T^{-1}U_1 \cap ... \cap T^{-n+1}U_{n-1}$ de $\alpha \vee T^{-1}\alpha \vee ... \vee T^{-n+1}\alpha$ ne peut pas contenir plus d'un point de F_n donc

$$\operatorname{card} F_n \leq N(\alpha \vee T^{-1}\alpha \vee ... \vee T^{-n+1}\alpha).$$

Il en résulte que

$$s(n,\varepsilon) \leq N(\alpha \vee T^{-1}\alpha \vee ... \vee T^{-n+1}\alpha)$$

et

$$h_{top}(T) = \limsup_{n\to\infty} \frac{1}{n} \ln s(n,\delta) \leq \lim_{n\to\infty} \frac{1}{n} \ln N(\alpha \vee T^{-1}\alpha ... \vee T^{-n+1}\alpha) \leq h_{top}(T). \ \square$$

Exercice. Faire la démonstration de la proposition précédente dans le cas des homéomorphismes. On pourra remarquer que si T est un homéomorphisme alors on a $r_k(n,\varepsilon) = r(n,\varepsilon)$ pour tout $k \in \mathbf{Z}$ où $r_k(n,\varepsilon)$ est définie comme $r(n,\varepsilon)$ avec la distance $d_{n,k}(x,y) = \max_{i=0}^{n-1} d(T^{k+i}x, T^{k+i}y)$ à la place de la distance d_n.

2.6. Entropie d'un sous décalage.

Proposition 19. *Soit $\mathcal{A}$ un ensemble fini et X un sous décalage de $\mathcal{A}^{\mathbf{N}}$ ou $\mathcal{A}^{\mathbf{Z}}$. Désignons par S le décalage sur $\mathcal{A}^{\mathbf{Z}}$. Alors*

$$h_{top}(S_{|X}) = \lim_{n\to\infty} \frac{1}{n} \ln \operatorname{card} \mathcal{L}_n(X).$$

Dem. Faisons la démonstration dans le cas d'un sous décalage bilatéral. Munissons $\mathcal{A}$ de la distance à valeur dans $\{0,1\}$ et $\mathcal{A}^{\mathbf{Z}}$ de la distance

$$d(x,y) = \sum_{n\in\mathbf{Z}} 4^{-|n|} d(x(n), y(n)).$$

On a $\delta = 1$ car pour tout $x \neq y$ appartenant à $\mathcal{A}^{\mathbf{Z}}$, il existe $n \in \mathbf{Z}$ tel que $x(n) \neq y(n)$ et par définition de la distance on obtient

$$d(S^n x, S^n y) \geq d((S^n x)(0), (S^n y)(0)) = d(x(n), y(n)) = 1.$$

Pour chaque $a \in \mathcal{A}$ le cylindre $[a] = \{\omega : \omega(0) = a\}$ est de diamètre $2\sum_{n>0} 4^{-n} < 1$. Le diamètre de tous les éléments du recouvrement ouvert $\alpha = \{[a] \cap X : a \in \mathcal{A}\}$ est inférieur à 1 donc d'après la proposition sur l'entropie de homéomorphismes expansifs,

$$h_{top}(S_{|X}) = \lim_{n\to\infty} \frac{1}{n} \ln N(\alpha \vee S^{-1}\alpha ... \vee S^{-n+1}\alpha).$$

Déterminons $N(\alpha \vee S^{-1}\alpha ... \vee S^{-n+1}\alpha)$. Soit $m = a_0 a_1 ... a_{n-1}$ un mot appartenant à $\mathcal{A}^n$. Soit l'ouvert

$$[m] \cap X = [a_0] \cap S^{-1}[a_1] ... \cap S^{-n+1}[a_{n-1}] \cap X$$

est vide soit il appartient à $\alpha \vee S^{-1}\alpha ... \vee S^{-n+1}\alpha$. Si $[m] \cap X$ n'est pas vide alors m appartient à $\mathcal{L}_n(X)$. Réciproquement si m est un facteur de longueur n d'un élément $\omega \in X$ avec un bon choix de l'entier $k \in \mathbf{Z}$, $S^k\omega(0) = a_0, ..., S^k\omega(n-1) = a_{n-1}$ donc $[m] \cap X$ n'est pas vide. Comme les ouverts $[m]$, $m \in \mathcal{A}^n$ sont deux à deux disjoints, on a

$$N(\alpha \vee S^{-1}\alpha ... \vee S^{-n+1}\alpha) = \mathrm{card}\, \mathcal{L}_n(X). \ \square$$

Théorème 19. *Soit $\mathcal{A}$ un ensemble fini à k éléments et $A = (A_{ij})$ une matrice carrée d'ordre k de 0 et de 1. Supposons que chaque colonne de la matrice A possède au moins un 1. Soit X_A la chaine de Markov topologique unilatéral associée à la matrice A. Soit S_A la restriction du décalage S à X_A. On a*

$$h_{top}(S_A) = \ln r(A)$$

où $r(A)$ désigne le rayon spectral de A.

Dem. Soit M une matrice. Notons $N(M)$ la somme des valeurs absolues de tous les coefficients de la matrice M. La norme $N(A^n)$ vaut

$$\sum_{1 \leq i_1, ..., i_{n+1} \leq k} A_{i_1 i_2} A_{i_2 i_3} ... A_{i_n i_{n+1}},$$

et chaque terme de cette somme vaut 0 ou 1. Le terme $A_{i_1 i_2} A_{i_2 i_3} ... A_{i_n i_{n+1}}$ vaut 1 ssi tous les $A_{i_k i_{k+1}}$ valent tous 1. Comme les colonnes de la matrice possède un 1, chaque mot $j_1 ... j_{n+1}$ vérifiant $A_{j_k j_{k+1}}$, $k = 1, ..., n$, est prolongeable en un élément de X_A. Par conséquent, $A_{i_1 i_2} A_{i_2 i_3} ... A_{i_n i_{n+1}} = 1$ ssi le mot $m = i_1 i_2 ... i_{n+1} \in \mathcal{L}_{n+1}(X_A)$. Avec la proposition précédente on obtient

$$h_{top}(S_A) = \lim_{n\to\infty} \frac{1}{n+1} \ln N(A^n).$$

Soit $|.|$ une norme sur l'ensemble des matrices telle que $|PQ| \leq |P|\,|Q|$. D'après le théorème du rayon spectral,

$$r(A) = \lim \inf_{n\to\infty} |A^n|^{\frac{1}{n}}.$$

Comme la norme N est équivalente à la norme $|.|$, on a

$$\lim_{n\to\infty} \frac{|A^n|^{\frac{1}{n}}}{N(A^n)^{\frac{1}{n}}} = 1$$

et

$$\ln r(A) = \ln \lim \inf_{n\to\infty} |A^n|^{\frac{1}{n}} = \lim \inf_{n\to\infty} \ln N(A^n)^{\frac{1}{n}},$$

donc

$$h_{top}(S_A) = \ln r(A). \square$$

3. Produits croisés, cobords

Nous nous sommes placé dans le cadre des systèmes dynamiques topologiques, une simple adaptations des définitions suivantes convient pour les systèmes dynamiques mesurés.

Définition 14. ***(Extension en groupe)*** *Soit X un espace métrique compact, $T : X \to X$ une application continue, $(G, .)$ un groupe localement compact et $\phi : X \to G$ une application continue. L'application $T_\phi : X \times G \to X \times G$ définie par $T_\phi(x, g) = (Tx, \phi(x).g)$ s'appelle une extension en groupe de T.*

Remarques 1. On parle aussi de produit gauche ou de produit croisé.
2. Le système dynamique $(X \times G, T_\phi)$ est une extension du système dynamique (X, T). En effet, si on note $p : X \times G \to X$ la projection $p(x, g) = x$ on a $T \circ p = T_\phi \circ p$.
3. Pour tout $n \geq 1$, $T_\phi^n(x, g) = (T^n x, \phi(T^{n-1}x)\phi(T^{n-2}x)...\phi(Tx)\phi(x).g)$.
4. On peut aussi définir l'extension de l'action d'un groupe H sur un espace X à un produit $X \times G$ où G est un groupe. Pour cela il faut se donner un "cocycle" de $X \times H$ dans G, c'est à dire une application $\phi : H \times X \to G$ vérifiant

$$\forall x \in X,\ \forall h_1, h_2 \in H,\ \phi(h_2 h_1, x) = \phi(h_2, h_1 x)\phi(h_1, x).$$

On obtient ainsi une action de H sur $X \times G$:

$$h(x, g) = (hx, \phi(h, x)g)$$

et on a bien

$$\begin{aligned} h_2(h_1(x, g)) &= h_2(h_1 x, \phi(h_1, x)) = (h_2 h_1 x, \phi(h_2, h_1 x)\phi(h_1, x)) \\ &= (h_2 h_1 x, \phi(h_2 h_1, x)) = (h_2 h_1)(x, g). \end{aligned}$$

5. Lorsque $\mathbf{Z}$ agit sur X par l'intermédiaire d'un homéomorphisme $T : X \to X$ toute application $\phi : X \to G$ engendre un cocycle :

$$\begin{aligned} \phi(n, x) &= \phi(T^{n-1}x)\phi(T^{n-2}x)...\phi(Tx)\phi(x) \text{ pour } n \geq 1, \\ \phi(0, x) &= 1, \\ \phi(n, x) &= \phi(T^{-m}x)...\phi(T^{-1}x) \text{ pour } n \leq -1. \end{aligned}$$

On vérifie aisément que

$$\phi(n + m, x) = \phi(n, T^m x)\phi(m, x).$$

Lorsque le groupe G est abélien on obtient en notation additive

$$\begin{aligned} \phi(n, x) &= \sum_{k=0}^{n-1} \phi(T^k x) \text{ pour } n \geq 1 \\ \phi(0, x) &= 0, \\ \phi(n, x) &= \sum_{k=n}^{-1} \phi(T^k x) \text{ pour } n \leq -1. \end{aligned}$$

Vocabulaire. La fonction ϕ s'appelle un cocycle.

Définition 15. ***(Cobord)*** *Soit X un espace métrique compact, $T : X \to X$ une application continue, $(G, .)$ un groupe localement compact.*
1. Soit $\phi : X \to G$ une application continue. On dit que ϕ est un cobord d'une application (non nécessairement continue) si il existe une application $\psi : X \to G$ continue telle que $\phi = (\psi^{-1} \circ T)\psi$ où ψ^{-1} désigne l'inverse dans G. Si ψ est continue, on dit que ϕ est le cobord d'une fonction continue.
2. Soit $\phi_1, \phi_2 : X \to G$ deux applications continues. On dit que ϕ_1 et ϕ_2 sont cohomologues si il existe une application continue $\psi : X \to G$ telle que $\phi_1 = (\psi^{-1} \circ T)\, \phi_2 \psi$.

Remarques. 1. Dans le cas des systèmes dynamiques mesurés on remplace la continuité par la mesurabilité et l'égalité partout par l'égalité presque partout.
2. En notation additive, on écrit pour un cobord

$$\phi = \psi - \psi \circ T.$$

1. La relation de cohomologie est une relation d'équivalence.

Proposition 20. ***(Equivalence des systèmes dynamiques)*** *Soit X un espace métrique compact, $T : X \to X$ une application continue, $(G, .)$ un groupe localement compact et $\phi_1, \phi_2 : X \to G$ deux applications continues. Si ϕ_1 et ϕ_2 sont cohomologues alors les systèmes dynamiques $(X \times G, T_{\phi_1})$ et $(X \times G, T_{\phi_2})$ sont topologiquement conjugués.*

Dem. Soit $\Psi : X \to G$ une application continue telle que $\phi_1 \psi \circ T = \psi \phi_2$. Considérons l'application

$$\begin{aligned} S \quad &: \quad X \times G \to X \times G \\ &: \quad (x, g) \to (x, \psi(x)g). \end{aligned}$$

L'application S est clairement un homéomorphisme. On a

$$T_{\phi_2} \circ S(x, g) = (Tx, \phi_2(x)\psi(x)g),$$

or $\phi_2(x)\psi(x) = \psi \circ T(x)\phi_1(x)$, donc

$$\begin{aligned} T_{\phi_2} \circ F(x, g) &= (Tx, \psi \circ T(x)\phi_1(x)g) \\ &= S(Tx, \phi_1(x)g) \\ &= S \circ T_\phi(x, g). \ \square \end{aligned}$$

Proposition 21. ***(Mesure invariante)*** *Soit X un espace métrique compact, $T : X \to X$ une application continue, $(G, .)$ un groupe localement compact et $\phi : X \to G$ une application continue. Appelons μ la mesure de Haar sur G invariante à gauche et supposons que m soit une mesure sur X T-invariante. Alors la mesure $\nu = m \otimes \mu$ est invariante par l'application $T_\phi : X \times G \to X \times G$ définie par $T_\phi(x, g) = (Tx, \phi(x).g)$.*

Dem. On utilise le théorème de Fubini.

$$\begin{aligned} \int_{X \times G} f \circ T_\phi \, d\nu &= \int_X \Big(\int_G f(Tx, \phi(x)g) \, d\mu(g)\Big) dm(x) \\ &= \int_X \Big(\int_G f(Tx, g) \, d\mu(g)\Big) dm(x) \\ &= \int_X \Big(\int_G f(x, g) \, d\mu(g)\Big) dm(x) \\ &= \int_{X \times G} f \, d\nu. \ \square \end{aligned}$$

Proposition 22. *Soit X un espace métrique compact, $T : X \to X$ une application continue, $(G, .)$ un groupe compact de mesure de Haar μ et $\phi : X \to G$ une application continue. Supposons T uniquement ergodique et appelons m l'unique probabilité T-invariante sur X. Si l'extension $T_\phi : X \times G \to X \times G$ définie par $T_\phi(x, g) = (Tx, \phi(x).g)$ est ergodique pour la mesure $\nu = m \otimes \mu$ alors elle est uniquement ergodique.*

Remarque préliminaire. Comme G est compact la mesure μ est invariante à gauche et à droite.

Dem. 1. Soit d'un système dynamique topologique (Y, S). Rappelons qu'un point y est générique pour une probabilité invariante λ si pour toute fonction continue $f : Y \to \mathbf{R}$

$$\lim_{n\to\infty} \frac{1}{n}(f(y) + ... + f \circ T^{n-1}(y)) = \int_Y f\, d\lambda.$$

Montrons que si (x, g) est générique pour ν alors pour tout $h \in G$, le point (x, gh) est générique. En effet, appelons $R : X \times G \to X \times G$ définie par $R(x, g) = (x, gh)$, on a $T_\phi R = RT_\phi$ et la mesure ν est R-invariante. Par conséquent, si $f : X \times G \to \mathbf{R}$ est continue, la suite

$$\frac{1}{n}\sum_{k=0}^{n-1} f(T_\phi^k R(x, g)) = \frac{1}{n}\sum_{k=0}^{n-1} f \circ R(T_\phi^k(x, g))$$

converge vers $\int f \circ R\, d\nu = \int f\, d\nu$ ce qui montre que $R(x, g)$ est un point générique.
2. Comme T_ϕ est ergodique presque tout (x, g) de $X \times G$ est un point générique. Donc il existe une partie A de X, de mesure nulle telle que pour tout $x \notin A$ il existe $g \in G$ avec (x, g) générique. Grâce au 1, nous savons donc que tout les $(x, g) \in (X\backslash A) \times G$ sont génériques. Si T_ϕ n'est pas uniquement ergodique il existe une autre probabilité λ T_ϕ-invariante. Choisissons une fonction continue $f : X \to \mathbf{R}$ telle que $\int_{X\times G} f\, d\nu \neq \int_{X\times G} f\, d\lambda$. Comme pour λ-presque (x, g) on a

$$\lim_{n\to\infty} \frac{1}{n}(f(x, g) + f(T_\phi(x, g)) + ... + f(T_\phi^{n-1}(x, g)) = \int f\, d\lambda,$$

$\lambda((X\backslash A) \times G) = 0$. Par conséquent, la mesure λ est portée par $A \times G$. Mais ceci est impossible car la projection de λ sur X doit être m car T est uniquement ergodique. $\square$

3.1. Un exemple, théorème d'équidistribution de Weyl.

Théorème 20. *Si P est un polynôme à coefficients réels avec au moins un coefficient autre que le terme constant, irrationnel, alors la suite $(P(n))_{n\in\mathbf{N}}$ est équidistribuée modulo 1.*

On peut supposer que le coefficients du terme de plus haut degré est irrationnel. En effet, supposons que ce coefficients soit le rationnel $\frac{p}{q}$. Appelons d le degré de P et Q le polynôme P moins son monôme de plus haut degré. Pour tout entier n et $r = 0, ..., q-1, P(qn + r) = \frac{p}{q}r^d + Q(nq + r) \bmod 1$. Comme l'equirépartition des q suites $(P(qn + r))_{n\in\mathbf{N}}$ entraîne celle de la suite $(P(n))_{n\in\mathbf{N}}$, en continuant ce processus, on se ramène a prouver l'équirépartition des suites de la forme $(R(n))_{n\in\mathbf{N}}$ où $R(n)$ est polynôme non constant dont le coefficient du terme de plus haut degré est irrationnel.

L'idée générale de la démonstration est d'utiliser une transformation du tore $\mathbf{T}^d$ dont l'unique probabilité invariante est la mesure de Lebesgue et pour laquelle la convergence des moyennes de Birkhoff de toute fonction continue f de $\mathbf{T}^d \to \mathbf{R}$ vers $\int_{\mathbf{T}^d} f\,dx$ entraîne l'équirépartition de la suite $P(n)$. Cette transformation est obtenue par une succession d'extensions d'une translation du tore $\mathbf{T}^1$ à $\mathbf{T}^1 \times \mathbf{T}^1$, $\mathbf{T}^1 \times \mathbf{T}^1 \times \mathbf{T}^1$,....

Soit θ un réel que l'on précisera plus loin. On peut considérer l'extension T_2 à $\mathbf{T}^2$ de la translation $T_1 : x \in \mathbf{T}^1 \to x + \theta \in \mathbf{T}^1$ définie par

$$T_2(x_1, x_2) = (x_1 + \theta, x_2 + x_1).$$

De même on peut étendre T_2 à $\mathbf{T}^3$ par

$$T_3(x_1, x_2, x_3) = (x_1 + \theta, x_1 + x_2, x_3 + x_2),$$

le cocycle correspondant est la fonction $\phi(x_1, x_2) = x_2$. Ainsi de suite, on construit des extensions successives $T_d : \mathbf{T}^d \to \mathbf{T}^d$ définies par

$$T_d(x_1, ..., x_d) = (x_1 + \theta, x_2 + x_1, ..., x_d + x_{d-1}).$$

Soit P un polynôme de degré d à coefficients réels. Posons

$$\begin{aligned} P_d &= P \\ P_{d-1}(x) &= P_d(x+1) - P_d(x) \\ &\vdots \\ P_0(x) &= P_1(x+1) - P_1(x). \end{aligned}$$

Chaque P_i est de degré i et P_0 est une constante θ. Prenons $\theta = P_0$. Pour tout n, on a

$$\begin{aligned} T_d(P_1(n), P_2(n), ..., P_d(n)) &= (P_1(n) + \theta, P_2(n) + P_1(n), ..., P_d(n) + P_{d-1}(n)) \\ &= (P_1(n+1), P_2(n+1), ..., P_d(n+1)), \end{aligned}$$

donc

$$T_d^n(P_1(0), P_2(0), ..., P_d(0)) = (P_1(n), P_2(n), ..., P_d(n)).$$

Par conséquent, si $f : \mathbf{T}^d \to \mathbf{R}$ ne dépend que de la dernière variable, la convergence des moyennes de Birkhoff de f au point $(P_1(0), ..., P_d(0))$ correspond exactement à l'équirépartition de la suite $(P(n))_{n \in \mathbf{N}}$.

Lemme 8. *Si θ est irrationnel alors T_d est ergodique pour la mesure de Lebesgue sur $\mathbf{T}^d$.*

Dem. Pour montrer que T_d est ergodique il suffit de montrer que les seules fonctions T-invariantes de $L^2(\mathbf{T}^d)$ sont les fonctions presque sûrement constantes. Soit f une fonction appartenant à $L^2(\mathbf{T}^d)$. Appelons A la matrice

$$A = \begin{pmatrix} 1 & 0 & . & . & 0 \\ 1 & 1 & . & . & 0 \\ 0 & 1 & 1 & . & . \\ . & . & . & . & . \\ 0 & . & . & 1 & 1 \end{pmatrix}$$

et $\Theta = (\theta, 0, ..., 0)$. On a $T_d(x) = \Theta + Ax$. Comme la mesure de Lebesgue du tore $\mathbf{T}^d$ est invariante par l'application $x \in \mathbf{T}^d \to Ax \in \mathbf{T}^d$, pour chaque $n \in \mathbf{T}^d$,

$$\begin{aligned} c_n(f \circ T_d) &= \int_{\mathbf{T}^d} f(\Theta + Ax) \exp(2i\pi n.x)\, dx \\ &= \int_{\mathbf{T}^d} f(\Theta + x) \exp(2i\pi n.A^{-1}x)\, dx \\ &= \int_{\mathbf{T}^d} f(\Theta + x) \exp(2i\pi\, {}^t A^{-1} n.x)\, dx \\ &= \int_{\mathbf{T}^d} f(x) \exp(2i\pi\, {}^t A^{-1} n.(x - \Theta))\, dx \\ &= \exp(-2i\pi\, {}^t A^{-1} n.\Theta) \int_{\mathbf{T}^d} f(x) \exp(2i\pi\, {}^t A^{-1} n.x)\, dx. \end{aligned}$$

Or

$$A^{-1} = \begin{pmatrix} 1 & 0 & . & . & 0 \\ -1 & 1 & . & . & 0 \\ 0 & -1 & 1 & . & . \\ . & . & . & . & . \\ 0 & . & . & -1 & 1 \end{pmatrix}, \quad {}^t A^{-1} = \begin{pmatrix} 1 & -1 & . & . & 0 \\ 0 & 1 & -1 & . & 0 \\ 0 & 0 & 1 & . & . \\ . & . & . & . & -1 \\ 0 & . & . & 0 & 1 \end{pmatrix},$$

donc

$$c_n(f \circ T) = \exp(2i\pi\, (n_1 - n_2)\theta)\, c_{{}^t A^{-1} n}(f).$$

où $n = (n_1, ..., n_d) \in \mathbf{Z}^d$. Si f est invariante alors pour tout $n \in \mathbf{Z}^d$, on a

$$c_n(f) = \exp(2i\pi\, (n_1 - n_2)\theta)\, c_{{}^t A^{-1} n}(f)$$

et donc

$$|c_n(f)| = |c_{{}^t A^{-1} n}(f)|$$

et comme la somme $\sum_{n \in \mathbf{Z}^d} |c_n(f)|^2$ est fini pour chaque n tel que $c_n(f) \neq 0$, les $({}^t A^{-1})^k n$, $k \in \mathbf{Z}$, ne peuvent être deux à deux distincts. On en déduit que pour chaque n tel que $c_n(f) \neq 0$, il existe k tel que $({}^t A^{-1})^k n = n$, or cela implique que ${}^t A^k n = n$ mais ${}^t A^k$ à un unique vecteur propre associé à la valeur propre 1 qui est $(1, 0, ..., 0)$, par conséquent $n = (n_1, 0, ..., 0)$. Finalement, pour les n de cette forme on doit avoir

$$c_n(f) = \exp(2i\pi n_1 \theta) c_n(f)$$

et comme θ est irrationnel, si $n_1 \neq 0$ on doit avoir $c_n(f) = 0$. $\square$

Fin de la démonstration du Théorème de Weyl.

Soit P un polynôme de degré $d \geq 1$ dont le coefficient du terme de plus haut degré est un l'irrationnel α. Le polynôme P_0 est constant est vaut $\theta = d!\alpha$ qui est irrationnel. Grâce au lemme et à la proposition, on montre successivement que toutes les extensions $T_1, T_2, ..., T_d$ sont uniquement ergodiques ce qui achève la démonstration. Finalement si $f : \mathbf{T}^1 \to \mathbf{R}$ est une fonction continue, on étend f à $\mathbf{T}^d$ par $F(x_1, ..., x_d) = f(x_d)$ et on obtient

$$\begin{aligned}\lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{\infty}f(P(k)) &= \lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{\infty}F(P_1(k),...,P_d(k))\\ &= \lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{\infty}F\circ T_d(k)=\int_{\mathbf{T}^d}F(x)\,dx\\ &= \int_{\mathbf{T}^1}f(x_d)\,dx_d.\ \square\end{aligned}$$

3.2. Cobords.

Théorème 21. *(Gottschalk-Hedlund [G, H]) Soit X un espace métrique compact, $T : X\to X$ une application continue ou un homéomorphisme et $f : X\to\mathbf{R}$ une application continue. Si T est minimale alors f est le cobord d'une fonction continue ssi les sommes $\sum_{k=0}^{n}f\circ T^k$, $n\geq 0$, sont uniformément bornées sur X.*

Dem. Si f est un cobord d'une fonction continue ψ alors les sommes $\sum_{k=0}^{n}f\circ T^k = \psi-\psi\circ T^{n+1}$ sont uniformément bornées. Démontrons la réciproque dans le cas où T est une application continue minimale. Si T est homéomorphisme il suffira d'utiliser que les seuls fermés invariants sont le vide et X alors que nous utiliserons que les seuls fermés stables sont le vide et X. Si les sommes $\sum_{k=0}^{n}f\circ T^k$, $n\geq 0$ sont uniformément bornées sur X alors la fonction $\phi : X\to\mathbf{R}$ définie par

$$\phi(x)=\limsup_{n\to\infty}\sum_{k=0}^{n}f\circ T^k(x)$$

est bornée sur X et on a

$$\begin{aligned}\phi\circ T(x) &= \limsup_{n\to\infty}\sum_{k=0}^{n}f\circ T^k(Tx)\\ &= \limsup_{n\to\infty}(-f(x)+\sum_{k=0}^{n+1}f\circ T^k(x))\\ &= -f(x)+\limsup_{n\to\infty}\sum_{k=0}^{n+1}f\circ T^k(x)=-f(x)+\phi(x).\end{aligned}$$

Pour terminer la démonstration il faudrait prouver que ϕ est continue. Malheureusement ce n'est pas évident et on aura besoin de fonctions auxiliaires.. Pour une fonction bornée $h : X\to\mathbf{R}$, notons h^+ la plus petite fonction semi-continue supérieurement (SCS) supérieure ou égale à h et h^- la plus grande fonction semi-continue inférieurement (SCI) inférieure ou égale à h. En fait, on a simplement

$$h^+(x)=\limsup_{y\to x}h(y),\ \textit{et}\ h^-(x)=\liminf_{y\to x}h(x).$$

Comme f est continue,

$$(\phi\circ T)^+=(\phi-f)^+=\phi^+-f.$$

Or $\phi^+\circ T$ est SCS et supérieure ou égale à $\phi\circ T$, donc $\phi^+\circ T\geq(\phi\circ T)^+$ et

$$\phi^+\circ T\geq\phi^+-f.$$

De la même manière on montre que

$$\phi^- \circ T \leq \phi^- - f.$$

On en déduit que

$$(\phi^+ - \phi^-) \circ T \geq \phi^+ - \phi^-.$$

La fonction $\psi = \phi^+ - \phi^-$ est donc une fonction SCS vérifiant $\psi \circ T \geq \psi$. Pour $t \in \mathbf{R}$, considérons l'ensemble $F_t = \{x \in X : \psi(x) \geq t\}$. Comme ψ est SCS, F_t est fermé et comme pour tout x, $\psi(Tx) \geq \psi(x)$, le fermé F_t est stable. La minimalité de T entraîne que pour tout t, F_t est soit vide soit égale à X. Si il existe deux points x_1 et x_2 de X tels que $t_1 = f(x_1) < t_2 = f(x_2)$ alors pour $t \in]t_1, t_2[$, F_t n'est ni vide ni égale à X, donc l'application ψ est constante. On en déduit que $\phi^+ = \phi^- + c$ est à la fois SCS et SCI donc continue. On obtient ainsi

$$f \geq \phi^+ - \phi^+ \circ T = \phi^- - \phi^- \circ T \geq f. \square$$

3.3. Produits gauches, minimalité, mesures invariantes. Références : Katok-Hasselblatt page 149-150 et Furstenberg.

Proposition 23. *Soit X un espace métrique compact, $T : X \to X$ une application continue et $\phi : X \to \mathbf{T}^1$ une application continue. Supposons T uniquement ergodique et appelons m l'unique probabilité T-invariante sur X. Considérons l'extension T_ϕ : $X \times \mathbf{T}^1 \to X \times \mathbf{T}^1$ définie par $T_\phi(x,t) = (Tx, \phi(x) + t)$. Alors T_ϕ est ergodique pour la mesure $\mu = m \otimes dt$ ssi pour tout entier relatif $k \neq 0$, l'équation*

$$k\phi = \psi - \psi \circ T$$

m-presque sûrement, n'a pas de solution $\psi : X \to \mathbf{T}^1$ mesurable.

Remarque. On sait que l'unique ergodicité de T_ϕ est équivalente à son ergodicité.
Dem. Si il existe $k \neq 0$ tel que l'équation

$$k\phi = \psi - \psi \circ T$$

m-presque sûrement, ait une de solution $\psi : X \to \mathbf{T}^1$ mesurable alors la fonction f : $X \times \mathbf{T}^1 \to \mathbf{T}^1$ définie par $f(x,t) = \psi(x) + kt$ est T_ϕ-invariante. En effet

$$\begin{aligned} f(T_\phi(x,t)) &= f(Tx, \phi(x)+t) = \psi(Tx) + k(\phi(x)+t) \\ &= \psi(x) + kt \end{aligned}$$

pour presque tout x et tout t. Comme cette fonction n'est presque sûrement constante sur aucune fibre $\{x\} \times \mathbf{T}^1$, la transformation T_ϕ n'est pas ergodique.
Réciproquement, supposons qu'il existe $f \in L^2(\mu)$ non presque sûrement constante telle que $f = f \circ T_\phi$. Pour presque tout x la fonction $t \to f(x,t)$ appartient à $L^2(\mathbf{T}^1)$, donc on peut déterminer le développement en série de Fourier de f à x fixé :

$$c_k(x) = \int_{\mathbf{R}/\mathbf{Z}} f(x,t) \exp(-2ik\pi t)\, dt.$$

Comme $f = f \circ T_\phi$, pour presque tout x, on a

$$\begin{aligned} c_k(x) &= \int_{\mathbf{R}/\mathbf{Z}} f \circ T_\phi(x,t) \exp(-2ik\pi t)\, dt \\ &= \int_{\mathbf{R}/\mathbf{Z}} f(Tx, \phi(x)+t) \exp(-2ik\pi t)\, dt \\ &= \exp 2ik\pi\phi(x) \int_{\mathbf{R}/\mathbf{Z}} f(Tx,t) \exp(-2ik\pi t)\, dt \\ &= \exp 2ik\pi\phi(x)\, c_k(Tx). \end{aligned}$$

La fonction $|c_k(x)|$ est donc presque sûrement constante pour chaque k. Si cette fonction est nulle pour chaque $k \neq 0$ alors la fonction f est presque sûrement constante, donc il existe un $k \neq 0$ tel que $|c_k(x)|$ soit presque sûrement une constante $c > 0$. En posant $g(x) = \frac{1}{c} c_k(x)$ on obtient une fonction mesurable de module 1 telle que

$$g = g \circ T \, \exp 2ik\pi\phi$$

presque sûrement. On peut relever g en une fonction mesurable $\psi : X \to \mathbf{T}^1$ telle que $g = \exp 2i\pi\psi$. Ainsi on obtient

$$k\phi = \psi - \psi \circ T$$

presque sûrement. $\square$

Proposition 24. *Soit X un espace métrique compact connexe, m une probabilité sur X, $T : X \to X$ une application continue conservant m et $\phi : X \to \mathbf{T}^1$ une application continue. Appelons Y le produit $X \times \mathbf{T}^1$ et T_ϕ l'extension $T_\phi : Y \to Y$ définie par $T_\phi(x,t) = (Tx, \phi(x) + t)$. Soit $g : Y \to \mathbf{T}^1$ une application continue. Si pour tout $x \in X$ l'application $g_x : t \in \mathbf{T}^1 \to g(x,t) \in \mathbf{T}^1$ est lipschitzienne de rapport M indépendant de x et à un degré $d \neq 0$, alors il n'existe pas de fonction mesurable $f : Y \to \mathbf{T}^1$ telle que*

$$f - f \circ T_\phi = g$$

presque-sûrement la mesure $m \otimes dt$.

Remarque. On peut facilement prouver que les degrés de deux applications proches de $\mathbf{T}^1 \to \mathbf{T}^1$ sont les mêmes. Grâce à la connexité de X, on obtient donc que le degré de l'application g_x ne dépend pas de $x \in X$.

Dem. Notons μ la mesure $m \otimes dt$. Supposons qu'il existe une application mesurable $f : Y \to \mathbf{T}^1$ telle que

$$f - f \circ T_\phi = g$$

μ-presque sûrement. On peut retirer à Y un ensemble $\mathcal{N}$ négligeable tel que l'équation précédente soit satisfaite partout sur $Y \backslash \mathcal{N}$ et tel que $Y \backslash \mathcal{N}$ soit stable par T_ϕ.
1. Soit ε un réel strictement positif que l'on choisira plus tard. D'après le théorème de Lusyn il existe une partie mesurable E de $Y \backslash \mathcal{N}$ dont la mesure est supérieure à $1-\varepsilon$ et telle que la restriction de f à E soit continue. Comme Y est métrique compact, la mesure μ est régulière et on peut supposer que E est un fermé. Appelons A l'ensemble des éléments x de X telle que la mesure de Lebesgue de $E_x = E \cap \{x\} \times \mathbf{T}^1$ soit supérieure à $1 - \sqrt{\varepsilon}$. La partie A est mesurable et d'après le théorème de Fubini, $(1 - m(A)) \times \sqrt{\varepsilon} \leq \mu(E^C) \leq \varepsilon$ donc $m(A) \geq 1 - \sqrt{\varepsilon}$. Il en resulte que la mesure de $E \cap A \times \mathbf{T}^1$ est supérieure à $1 - \varepsilon - \sqrt{\varepsilon}$.
2. Comme f est continue sur le compact E, il existe $\delta > 0$ tel que pour tout couple de points (x, y) de E dont la distance est $\leq \delta$, on ait $d(f(x), f(y)) \leq \varepsilon$. Comme Y est recouvert par un nombre fini de boule de rayon $\delta/2$, il existe un boule B de rayon $\delta/2$ telle que la mesure de $C = B \cap E \cap A \times \mathbf{T}^1$ soit strictement positive. D'après le théorème de récurrence de Poincaré, il existe un point $x_0 = (a_0, t_0) \in C$ dont l'orbite $(T_\phi^n(x_0))_{n \geq 0}$ passe une infinité de fois par C. Appelons $(n_p)_p$ la suite des entiers tels que $T_\phi^{n_p}(x_0) \in C$.
3. Pour n entier posons

$$\phi_n = \frac{1}{n} \sum_{k=0}^{n-1} \phi \circ T^k.$$

Pour tout $(a,t) \in Y$, on a

$$T_\phi^n(a,t) = (T^n a, t + \phi_n(a))$$

donc

$$d(T_\phi^n(a,t), (a,t))$$

ne dépend pas de t mais que de a. Ainsi, pour tout $t \in \mathbf{T}^1$ et tout $p \in \mathbf{N}$,

$$d(T_\phi^{n_p}(a_0,t), (a_0,t)) = d(T_\phi^{n_p}(a_0,t_0), (a_0,t_0)) \leq \delta.$$

Remarquons aussi que (a_0, t_0) et $T_\phi^{n_p}(a_0, t_0)$ appartiennent à $A \times \mathbf{T}^1$ et donc (a_0, t) et $T_\phi^{n_p}(a_0, t)$ appartiennent à $A \times \mathbf{T}^1$.
4. Par définition de A l'ensemble des $t \in \mathbf{T}^1$ tel que $(a_0, t) \notin E$ est de mesure inférieure à $\sqrt{\varepsilon}$. De même, comme

$$T_\phi^{n_p}(a_0, t) = (T^{n_p} a_0, \phi_{n_p}(a_0) + t),$$

l'ensemble des $t \in \mathbf{T}^1$ tels que $T_\phi^{n_p}(a_0, t) \notin E$ est de mesure inférieure à $\sqrt{\varepsilon}$. Fixons p. L'ensemble F des $t \in \mathbf{T}^1$ tels que $(a_0, t) \in E$ et $T_\phi^{n_p}(a_0, t) \in E$, est donc de mesure $\geq 1 - 2\sqrt{\varepsilon}$.
Si $t \in F$, alors (a_0, t) et $T_\phi^{n_p}(a_0, t) \in E$, et $d((T_\phi^{n_p}(a_0, t)), (a_0, t)) \leq \delta$, donc par continuité

$$d(f(T_\phi^{n_p}(a_0, t)), f(a_0, t)) \leq \varepsilon.$$

5. Comme

$$\sum_{k=0}^{n_p-1} g \circ T_\phi^k(a_0, t) = f(a_0, t) - f(T_\phi^{n_p}(a_0, t)),$$

pour tout $t \in F$,

$$d(\sum_{k=0}^{n_p-1} g \circ T_\phi^k(a_0, t), 0) \leq \varepsilon.$$

6. Pour chaque k, la fonction $h_k : t \in \mathbf{T}^1 \to h_k(t) = g(T_\phi^k(a_0, t)) = g(T^k a_0, \phi_k(a_0) + t)$ est continue et sont degré est le même que celui de la fonction $g_{T^k_{a_0}}$ qui est un entier d indépendant de k, non nul par hypothèse. Appelons H_k un relèvement de la fonction $h_k \circ \mathrm{pr} : \mathbf{R} \to \mathbf{T}^1$ où pr désigne la projection de $\mathbf{R}$ dans $\mathbf{T}^1$. Par définition du degré $H_k(s+1) = H_k(s) + d$ pour tout t. Le 5 se traduit sur H_k par

$$\forall s \in \mathbf{R},\ \mathrm{pr}(s) \in F \Rightarrow d(\sum_{k=0}^{n_p-1} H_k(s), \mathbf{Z}) \leq \varepsilon.$$

Posons $H(t) = \sum_{k=0}^{n_p-1} H_k(t)$. Examinons $H([0,1])$. On a

$$H(1) = \sum_{k=0}^{n_p-1} H_k(1) = n_p d + H(0)$$

donc

$$|H([0,1])| \geq n_p d$$

où $|.|$ désigne la mesure de Lebesgue. D'une part, l'ensemble G des s de $[0,1]$ dont la projection appartient à F a une mesure $\geq 1 - 2\sqrt{\varepsilon}$. D'autre part, tous les points de $H(G)$ sont à une distance $\leq \varepsilon$ de $\mathbf{Z}$, donc

$$|H(G)| \leq 2M\varepsilon.$$

Par conséquent, $|H([0,1]\backslash G)| \geq n_p d - 2M\varepsilon$ mais comme H est lipschitzienne de rapport $n_p M$ on a aussi $|H([0,1]\backslash G)| \leq n_p M 2\sqrt{\varepsilon}$. D'où la contradiction en choisissant ε assez petit puis n_p assez grand. □

Corollaire 9. *Soit $T : \mathbf{T}^n \to \mathbf{T}^n$ une application donnée par*

$$T(x_1, ..., x_n) = (x_1 + \theta, \phi_1(x_1) + x_2, \phi_2(x_1, x_2) + x_3, ..., \phi_{n-1}(x_1, ..., x_{n-1}) + x_n)$$

où $\theta \in \mathbf{T}^1 \backslash \mathbf{Q}$ et chaque $\phi_k : \mathbf{T}^k \to \mathbf{T}^1$ est continue lipschitzienne par rapport à x_k, de rapport de lipschitz indépendant de k de $(x_1, ..., x_{k-1})$ et de degré non nul. Alors T est uniquement ergodique et l'unique probabilité invariante du tore est la mesure de Lebesgue du tore.

Dem. Appelons $\mathbf{T}^0 = \{0\}$ un espace réduit à un point et définissons $\phi_0 : \mathbf{T}^0 \to \mathbf{T}^1$ par $\phi_0(0) = \theta_0$. Considérons pour les applications $T_k : \mathbf{T}^k \to \mathbf{T}^k$, $k = 0, ..., n$, définies par

$$T_0(0) = 0,\ T_1(x_1) = (x_1 + \theta)$$

et pour $k = 2, ..., n$

$$T_k(x_1, ..., x_k) = (x_1 + \theta, \phi_1(x_1) + x_2, \phi_2(x_1, x_2) + x_3, ..., \phi_{k-1}(x_1, ..., x_{k-1}) + x_k).$$

On a $T = T_n$. Montrons par récurrence sur k, $k = 1, ..., n$, que l'unique mesure T_k-invariante est la mesure de Lebesgue du tore $\mathbf{T}^k$. L'application T_1 translation irrationnelle du tore $\mathbf{T}^1$ donc la seule probabilité T_1-invariante est la mesure de Lebesgue de $\mathbf{T}^1$. Il y une seule probabilité sur $\mathbf{T}^0$ donc T_0 admet une unique probabilité invariante. Soit $k \geq 1$. Supposons que la seule probabilité T_k invariante soit la mesure de Lebesgue du tore $\mathbf{T}^k$. En posant $x = (x_1, ..., x_k)$, on a

$$T_{k+1}(x, x_{k+1}) = (T_k(x), \phi_k(x) + x_{k+1}).$$

Appliquons la proposition précédente à $X = \mathbf{T}^{k-1}$, $Y = X \times \mathbf{T}^1 = \mathbf{T}^k$, $\phi = \phi_{k-1}$ et $g = \phi_k$. On en déduit que pour tout entier relatif $m \neq 0$, l'équation

$$f - f \circ T_k = m\phi_k\ p.s.$$

n'admet pas n'admet pas de solution mesurable $f : \mathbf{T}^k \to \mathbf{T}^1$. D'après l'hypothèse de recurrence, T_k est uniquement ergodique, la première proposition du paragraphe assure donc que T_k est aussi unquement ergodique.

Proposition 25. *Soit θ et x_0 deux éléments de $\mathbf{T}^1 \backslash \mathbf{Q}$, $\phi : \mathbf{T}^1 \to \mathbf{T}^1$ une application continue et $T = T_\phi : \mathbf{T}^2 \to \mathbf{T}^2$ définie par $T(x, y) = (x + \theta, y + \phi(x))$. Si T n'est pas minimale alors il existe $\psi : \mathbf{T}^1 \to \mathbf{T}^1$ continue sauf en x_0 et un rationnel r telle que pour tout x de $\mathbf{T}^1$, $\phi(x) = r + \psi(x) - \psi(x + \theta)$.*

Preuve. On sait que T admet un fermé invariant minimal M. Appelons $p : (x, y) \in \mathbf{T}^2 \to x \in \mathbf{T}^1$ la projection de $\mathbf{T}^2$ sur $\mathbf{T}^1$. La projection $p(M)$ est invariant par la translation $\tau(x) = x + \theta$ de $\mathbf{T}^1$. En effet, soit $x \in p(M)$ alors il existe y tel que $(x, y) \in M$ et comme $T(M) \subset M$ on a $(x + \theta, y + \phi(x)) \in M$ et donc $x + \theta \in p(M)$. Comme la projection d'un fermé est un fermé et que la translation τ est minimale, $p(M) = \mathbf{T}^1$.
1. Soit $x \in \mathbf{T}^1$ et considérons la "section" $M_x = M \cap \{x\} \times \mathbf{T}^1$. Nous allons montrer que si cette section contient deux points y et $y + \alpha$ alors elle est invariante par la translation $z \to z + \alpha$. En effet, comme M est minimal l'orbite du point (x, y) passe arbitrairement proche de $(x, y + \alpha)$, c-a-d il existe une suite $(n_q)_q$ tendant vers l'infini telle que $d(T^{n_q}(x, y), (x, y + \alpha)) \to 0$ quand q tend vers l'infini. Or pour tout z,

$$T^n(x, z) = T^n(x, y) + (0, z - y),$$

donc

$$T^{n_q}(x,z) = T^{n_q}(x,y) + (0, z-y) \to (x, y+\alpha) + (0, z-y) = (x, z+\alpha).$$

Comme M_x est fermé, si $(x,z) \in M_x$ alors $(x, z+\alpha) \in M_x$.
2. Il y a donc deux cas possibles, ou M_x est de la forme$\{x\} \times (h(x) + G_x)$ où G_x est un sous groupe fini de $\mathbf{T}^1$ ou $M_x = \{x\} \times \mathbf{T}^1$. Comme M est fermé, le cardinal de G_x est une fonction SCS donc borélienne, d'autre part on a

$$T(M_x) \subset M_{x+\theta}$$

de même

$$T^{-1}(M_{x+\theta}) \subset M_x,$$

par conséquent la fonction $N : x \to card G_x \in \mathbf{N} \cup \{\infty\}$ est invariante par la translation $x \to x + \theta$. La fonction N est donc presque-sûrement constante et comme elle est SCS, elle est constante. Si N est la constante $+\infty$ alors $M = \mathbf{T}^2$ donc $N = q \in \mathbf{N}$ et $G_x = H_0$ est l'unique sous-groupe de $\mathbf{T}^1$ à q éléments.
2. Quotientons $\mathbf{T}^2$ par le sous-groupe $H = \{0\} \times \{0, \frac{1}{q}, ..., \frac{q-1}{q}\} = \{0\} \times H_0$. Identifions $\mathbf{T}^2/H$ à $\mathbf{T}^1 \times \mathbf{T}^1/H_0$. Pour chaque $x \in \mathbf{T}^1$ la section $(M/H)_x$ a exactement un point : $h(x)$ modulo H. Avec l'identification, $h(x) = (x, g(x))$ où $g(x) \in \mathbf{T}^1/H_0$. Comme l'ensemble des couples $(x, g(x))$ est le fermé M/H, la fonction $x \to g(x)$ est continue. On peut relever l'application g en une application $\psi : \mathbf{T}^1 \to \mathbf{T}^1$ continue sur $\mathbf{T}^1 \backslash \{x_0\}$ et continue à droite en x_0. Pour cela, il suffit de relever l'application de $[x_0, x_0+1]$ dans $\mathbf{T}^1/H_0$ qui à x associe l'image par $g(\pi(x))$ où π est la projection $\mathbf{R}$ sur $\mathbf{T}^1$, puis de composée avec la réciproque de π à valeur dans $[x_0, x_0+1[$.
3. Pour tout x de $\mathbf{T}^1$, $M_x = \{x\} \times (\{\psi(x)\} + H_0)$ et comme $T(M_x) = M_{x+\theta}$, on a

$$T(x, \psi(x)) = (x + \theta, \psi(x+\theta) + d(x))$$

où d est un application de $\mathbf{T}^1$ dans H_0. En identifiant les secondes coordonnées on obtient

$$\psi(x) + \phi(x) = \psi(x+\theta) + d(x),$$

par conséquent d est continue sur $\mathbf{T}^1 \backslash \{x_0\}$ et continue à droite en x_0, d doit donc être constante. On conclut en prenant un rationnel r se projetant sur d (et changeant ψ en $-\psi$). $\square$

Application : Katok Hasselblatt, un exemple de difféomorphisme minimal non ergodique, page 417-418.

Lemme 9. *Soit θ un nombre irrationnel. Il existe une application mesurable $\Phi : \mathbf{T}^1 \to \mathbf{T}^1$ telle que l'application $\Phi(x) - \Phi(x+\theta)$ soit presque partout égale à une fonction $\phi(x)$ de classe C^∞ et telle que Φ ne soit pas presque partout égale à une fonction continue en* 0.

Dem. (Adaptée de Furstenberg) 1. Comme la suite $(n\theta \mod 1)$ est dense dans $[0,1]$, il existe une suite d'entiers $(n_k)_{k\geq 1}$ tendant vers l'infini telle que pour tout k, $n_k\theta \mod 1 \in]2^{-k-1}, 2^{-k}[$. Considérons la fonction $f : \mathbf{T}^1 \to \mathbf{R}$ définie par

$$f(x) = \sum_{k\geq 1} \frac{1}{k}(1 - e^{2i\pi n_k \theta})\, e^{2i\pi n_k x}.$$

Le choix de la suite $(n_k)_k$ assure que f est de classe C^∞. La fonction

$$F(x) = \sum_{k\geq 1} \frac{1}{k}\, e^{2i\pi n_k x}$$

est dans $L^2(\mathbf{T}^1)$ car la série $\sum_{k\geq 1} \frac{1}{k^2}$ converge et pour presque tout x de $\mathbf{T}^1$,

$$F(x) - F(x+\theta) = f(x).$$

2. Montrons que la fonction F n'est pas presque partout égale à une fonction continue en 0. D'après le théorème de Fejer, si F est presque partout égale à une fonction continue en 0 alors la suite des moyennes de Césaro de la série de Fourier de F en 0 est convergente. Pour $p \in \mathbf{Z}$, notons $c_p(F)$ le $p^{ième}$ coefficients de Fourier de F. Calculons les moyennes de Césaro en 0 ($n_0 = 0$) :

$$\begin{aligned}
\frac{1}{n_k}\sum_{m=0}^{n_k-1}\sum_{p=-m}^{m} c_p(F) &= \frac{1}{n_k}\sum_{q=0}^{k-1}\sum_{n_q\leq m<n_{q+1}}\sum_{p=-m}^{m} c_p(F)\\
&= \frac{1}{n_k}\sum_{q=0}^{k-1}\sum_{n_q\leq m<n_{q+1}}\sum_{j=1}^{q}\frac{1}{j}\\
&= \frac{1}{n_k}\sum_{q=0}^{k-1}(n_{q+1}-n_q)\sum_{j=1}^{q}\frac{1}{j}\ ;
\end{aligned}$$

comme la série $\sum_{j\geq 1}\frac{1}{j}$ diverge, la moyenne $\frac{1}{n_k}\sum_{m=0}^{n_k-1}\sum_{p=-m}^{m} c_p(F)$ tend vers l'infini quand k tend vers l'infini.
3. Appelons pr la projection de $\mathbf{R}$ sur $\mathbf{T}^1$. La fonction $\Phi = \mathrm{pr}\circ F$ peut être presque partout égale à une fonction continue en 0 alors que F ne l'est pas. On ne peut donc pas conclure directement. Posons $\phi_\lambda(x) = \mathrm{pr}(\lambda f(x))$ et $\Phi_\lambda(x) = \mathrm{pr}(\lambda F(x))$ où λ est un réel que nous allons choisir. On a évidement $\Phi_\lambda(x) - \Phi_\lambda(x+\theta) = \phi_\lambda(x)$, il suffit donc de prouver qu'il existe λ tel que Φ_λ ne soit pas presque partout égale à une fonction continue en 0.
Supposons que la fonction Φ_1 soit presque partout égale à une fonction continue en 0. En ajoutant une constante à F on peut supposer qu'il existe une partie négligeable E de $\mathbf{T}^1$ telle que

$$\lim_{\substack{x\to 0\\ x\notin E}} \Phi(x) = 0.$$

Si il existe un voisinage V de 0 tel que $|F|$ soit essentiellement bornée sur V alors il suffit de prendre λ strictement positif assez petit pour que la fonction Φ_λ conviennent. On peut donc supposer que F n'est essentiellement bornée sur aucun voisinage de 0. Pour chaque entier $k \geq 1$, il existe un voisinage V_k de 0 inclus dans la boule de centre 0 et rayon $\frac{1}{k}$, tel que

$$\forall x \in V_k\backslash E,\ d(\Phi(x), 0) \leq \frac{1}{k},$$

pour F cela signifie que

$$\forall x \in V_k\backslash E,\ d(F(x), \mathbf{Z}) \leq \frac{1}{k}.$$

Comme F n'est pas essentiellement sur V_k, il y a une infinité d'entiers n tels que la mesure de

$$\{x \in V_k \backslash E : |F(x) - n| \leq \frac{1}{k}\},$$

soit strictement positive. On peut donc construire une suite d'entiers $(n_k)_{k\geq 1}$ deux à deux distincts telle que quelque soit k la mesure de

$$A_k = \{x \in V_k \backslash E : |F(x) - n_k| \leq \frac{1}{k}\}$$

soit strictement positive. Si λ est tel que Φ_λ soit presque partout égale à une fonction continue alors il existe une partie négligeable E_λ telle

$$\lim_{\substack{x \to 0 \\ x \notin E_\lambda}} \Phi_\lambda(x) = l \in \mathbf{T}_1,$$

par conséquent si (x_k) est une suite de points telle que pour tout k, $x_k \in A_k \backslash E_\lambda$, alors $\lim_{k\to\infty} \Phi(x_k)$. Une telle suite existe bien car les parties $A_k \backslash E_\lambda$ sont non vide. Pour un telle suite on a

$$\lim_{\substack{p\to\infty \\ q\to\infty}} d(\lambda(F(x_p) - F(x_q)), \mathbf{Z}) = \lim_{\substack{p\to\infty \\ q\to\infty}} d(\Phi_\lambda(x_p), \Phi_\lambda(x_q)) = 0.$$

Or $F(x_k) = n_k + \varepsilon_k$ où $|\varepsilon_k| \leq \frac{1}{k}$ donc $\lambda(F(x_p) - F(x_q)) = \lambda(n_p - n_q) + \lambda(\varepsilon_p - \varepsilon_q)$. Par conséquent,

$$\lim_{\substack{p\to\infty \\ q\to\infty}} \mathrm{pr}(\lambda(n_p - n_q)) = 0$$

et

$$\lim_{k\to\infty} \mathrm{pr}(\lambda(n_{k+1} - n_k)) = 0$$

Pour achever la démonstration, il suffit donc de prouver qu'il existe un λ tel que la suite $(\mathrm{pr}(\lambda(n_{k+1} - n_k)))$ ne converge pas vers 0 quand k tend vers l'infini. Soit

$$B_k = \{\lambda \in]0,1] : d(\mathrm{pr}(\lambda(n_{k+1} - n_k)), 0) \geq \frac{1}{4}\}.$$

La mesure de chaque B_k est $\frac{1}{2}$, par conséquent la mesure de $\mathcal{B} = \limsup_{k\to} B_k$ est strictement positive.. Ainsi $\mathcal{B}$ est non vide et pour chaque $\lambda \in \mathcal{B}$, $\mathrm{pr}(\lambda(n_{k+1} - n_k))$ ne converge pas vers 0 quand k tend vers l'infini. $\square$

Corollaire 10. *Il existe un difféomorphisme de $\mathbf{T}^2$ minimal non uniquement ergodique.*

Dem : Soit θ un nombre irrationnel. Considérons l'extension T à $\mathbf{T}^2$ de la translation $x \in \mathbf{T}^1 \to x + \theta \in \mathbf{T}^1$ définie par définie par $T(x, y) = (x + \theta, y + \phi(x))$ où $\phi(x) = \Phi(x) - \Phi(x + \theta)$ est donnée par le lemme précédent. D'après la première proposition, T admet une infinité de mesure invariante. T est un difféomorphisme et d'après la proposition précédente, si T n'est pas minimal alors il existe $\psi : \mathbf{T}^1 \to \mathbf{T}^1$ continue sauf un point que l'on peut choisir différent de 0 et un rationnel r tels que pour tout x de $\mathbf{T}^1$, $r + \psi(x) - \psi(x + \theta) = \phi(x)$ (modulo 1). Considérons l'application $F : \mathbf{T}^1 \to \mathbf{T}^1$ définie par $F(x) = \Phi(x) - \psi(x)$. On a pour presque tout x de $\mathbf{T}^1$,

$$\begin{aligned} F(x) - F(x + \theta) &= \Phi(x) - \Phi(x + \theta) - (\psi(x) - \psi(x + \theta)) \\ &= \phi(x) - (\phi(x) - r) = r. \end{aligned}$$

Si $r = 0$ alors la fonction F est invariante et comme la translation $x \to x + \theta$ est ergodique cela signifie que F est presque-sûrement constante et donc Φ est presque sûrement égale à une fonction continue en 0 ce qui contredit le lemme, par conséquent $r \neq 0$. Ainsi $F(x+\theta) = F(x) + r$ pour presque tout x de $\mathbf{T}^1$ où $r = \frac{p}{q}$ est un rationnel différent de 0. On en déduit que presque tout x de $\mathbf{T}^1$, $F(x + q\theta) = F(x) + qr = F(x)$. Par conséquent, F est invariante par la translation $x \to x + q\theta$ et on conclut comme précédemment. $\square$

4. Ergodicité des produits croisés

4.1. Lemme de Rokhlin-Halmos. Référence : Aaronson page 47.

Hypothèse (*) $(X, \mathcal{B}, m)$ vérifie la propriété (*) il existe une suite décroissante de parties mesurables $(B_i)_{i \in \mathbf{N}}$ d'intersection vide et telle que pour tout i, $0 < m(B_i) < \infty$.

Lemme 10. *Soit $(X, \mathcal{B}, m)$ un espace mesuré σ-fini vérifiant (*), $T : X \to X$ une transformation non singulière conservative, ergodique et telle que $m(T^{-1}E) < \infty$ lorsque $m(E) < \infty$. Alors pour tout $n \geq 1$ et tout $\varepsilon > 0$, il existe une partie mesurable E de X telle que les ensembles $T^{-k}E$, $k = 0, ..., n-1$, soit deux à deux disjoints et telle que $m(X \setminus \cup_{k=0}^{n-1} T^{-k}E) \leq \varepsilon$.*

Dem. Soit $\varepsilon > 0$ et $n \geq 1$. Utilisons (*), lorsque i tend vers l'infini , $m(\cup_{k=0}^{n-1} T^{-k}B_i)$ tend vers 0 car $\cup_{k=0}^{n-1} T^{-k}B_i$ décroit vers le vide et $m(\cup_{k=0}^{n-1} T^{-k}B_0) < \infty$. Il existe donc i_0 tel que $m(\cup_{k=0}^{n-1} T^{-k}B_{i_0}) \leq \varepsilon$. Posons $A = A_0 = B_{i_0}$ et pour tout $k \geq 1$, $A_k = (T^{-k}A) \setminus \cup_{i=0}^{k-1} A_i$. A_k est l'ensemble des x tels que la suite $(T^i(x))$ rentre dans A la première fois en $i = k$. Vérifions que $E = \cup_{p \geq 1} A_{np}$ convient.
Soit $i \in \{1, ..., n-1\}$. $T^{-i}E \cap E$ est vide car si il existe x appartenant à $T^{-i}E \cap E$ alors il existe p et $q \geq 1$ tels que $x \in T^{-i}A_{np}$ et $x \in A_{nq}$. $x \in T^{-i}A_{np}$ entraîne que $T^{np+i}x \in A$ et comme $x \in A_{nq}$ on a $nq \leq np + i$. De même $T^{nq}x \in A$ entraîne que $T^{nq-i}T^i x \in A$ or $T^i x \in A_{np}$ donc $nq - i \geq np$ et $np + i = nq$ ce qui est impossible. On en déduit que les ensembles $T^{-k}E$, $k = 0, ..., n-1$, soit deux à deux disjoints.
Comme T est ergodique et conservative, on a $m(X \setminus \cup_{k \geq 0} T^{-k}A) = 0$ et donc d'après le choix de A on a

$$m(X \setminus \cup_{k \geq n} T^{-k}A) \leq \varepsilon.$$

Or $\cup_{p \geq 1} A_{np+i} \subset T^{-i}E$ donc $\cup_{i \geq n} A_i \subset \cup_{i=0}^{n-1} T^{-i}E$ et

$$m(X \setminus \cup_{i=0}^{n-1} T^{-i}E) \leq m(X \setminus \cup_{k \geq n} T^{-k}A) \leq \varepsilon.$$

Corollaire 11. *Soit $(X, \mathcal{B}, m)$ un espace mesuré probabilisé vérifiant (*), $T : X \to X$ une application mesurable conservant la mesure et ergodique. Si A appartient à $\mathcal{B}$ et $0 < \mu(A) < \infty$ alors pour tout entier $N \geq 1$, il existe B appartenant à $\mathcal{B}_+$ inclus dans A tel que $B, ..., T^{-N+1}B$ soient deux à deux disjoints.*

Dem. Utilisons le lemme avec $\varepsilon = \frac{m(A)}{2}$ et $n \geq \frac{2N}{m(A)}$. Soit E une partie mesurable telle que les ensembles $T^{-k}E$, $k = 0, ..., n-1$, soit deux à deux disjoints et telle que $m(X \setminus \cup_{k=0}^{n-1} T^{-k}E) \leq \frac{m(A)}{2}$. La mesure de $A \cap (\cup_{k=0}^{n-1} T^{-k}E)$ est supérieure à $\frac{m(A)}{2}$ et les parties $T^{-k}E$ ont toutes une mesures inférieures à $\frac{1}{n} \leq \frac{m(A)}{2N}$ donc il y a au moins N entiers k compris entre 0 et $n-1$ tels que $m(T^{-k}E \cap A) > 0$. Soit k_0 le plus petit de ces entiers, k_0 est inférieure à $n - N$. Il suffit de prendre $B = A \cap T^{-k_0}E$ car pour chaque $i \leq N-1$, $T^{-i}B \subset T^{-k_0-i}E$ et les parties $T^{-k_0}E, ..., T^{-k_0-(N-1) \geq -n+1}E$ sont deux à deux disjointes. $\square$

4.2. Valeurs essentielles. Référence : Aaronson page 247-261.

Définition 16. *Soit groupe abélien topologique $(G,+)$. Un norme est une application $\|.\| : G \to [0,\infty[$ telle que $d(x,y) = \|x-y\|$ soit une distance définissant la topologie de G (cette distance est automatiquement invariante par translation).*

Notation. Soit $(G,+)$ un groupe commutatif, X un ensemble, $T : X \to X$ une application. Pour toute application $\phi : X \to G$, on note

$$\phi_n = \sum_{k=0}^{n-1} \phi \circ T^k$$

lorsque $n \geq 0$ et

$$\phi_n = -\sum_{k=n}^{-1} \phi \circ T^k$$

lorsque $n < 0$. De sorte que pour tout entiers m et n, on a

$$\phi_{n+m} = \phi_n + \phi_m \circ T^n.$$

Définition 17. *(Shmidt) Soit $(G,+)$ un groupe localement compacte muni d'une norme $\|.\|$, $(X,\mathcal{B},m)$ un espace probabilisé, $T : X \to X$ un endomorphisme de (X,B,m) et $\phi : X \to G$ une application mesurable. On définit les ensembles suivants :*
1. Les valeurs persistantes

$$\Pi(\phi) = \{a \in G : \forall A \in \mathcal{B}_+,\ \forall \varepsilon > 0,\ \exists n \geq 1,\ m(A \cap T^{-n}A \cap [\|\phi_n - a\| < \varepsilon]) > 0\},$$

2. Si de plus T est inversible, les valeurs essentielles

$$E(\phi) = \{a \in G : \forall A \in \mathcal{B}_+,\ \forall \varepsilon > 0,\ \exists n \in \mathbf{Z},\ m(A \cap T^{-n}A \cap [\|\phi_n - a\| < \varepsilon]) > 0\},$$

3. Notons $T_\phi : X \times G \to X \times G$, $T_\phi(x,y) = (Tx, y + \phi(x))$ le produit croisé, l'ensemble des périodes de T_ϕ

$$Per(\phi) = \{a \in G : \forall A \in \mathcal{I}_{T_\phi},\ Q_a A = A \mod m \times m_G\}$$

où $\mathcal{I}_{T_\phi}$ désigne la tribu des invariants de T_ϕ et $Q_a : (x,y) \in X \times G \to (x, y+a) \in X \times G$.

Remarques 1. $\Pi(\phi) \subset E(\phi)$.
2. $0 \in E(\phi)$ car $\phi_0 = 0$.

Proposition 26. *Soit $(G,+)$ un groupe localement compacte muni d'une norme $\|.\|$, $(X,\mathcal{B},m)$ un espace probabilisé, $T : X \to X$ un endomorphisme de (X,B,m) et $\phi : X \to G$ une application mesurable. Alors $\Pi(\phi)$ est un fermé de G.*

Dem. Soit a un élément de G adhérent à $\Pi(\phi)$, $\varepsilon > 0$ et $A \in \mathcal{B}_+$. Il existe $b \in \Pi(\phi)$ tel que $\|a-b\| \leq \frac{\varepsilon}{2}$ et $n \geq 1$ tel que

$$m(A \cap T^{-n}A \cap [\|\phi_n - b\| < \frac{\varepsilon}{2}]) > 0.$$

Par conséquent,

$$m(A \cap T^{-n}A \cap [\|\phi_n - a\| < \varepsilon]) > m(A \cap T^{-n}A \cap [\|\phi_n - b\| < \varepsilon/2]) > 0 > 0$$

et a appartient à $\Pi(\phi)$ qui est donc un fermé.

Proposition 27. *Soit $(G,+)$ un groupe localement compacte muni d'une norme $\|.\|$, $(X,\mathcal{B},m)$ un espace probabilisé, $T : X \to X$ un endomorphisme de (X,B,m) et $\phi : X \to G$ une application mesurable. Supposons que le lemme de Rokhlin soit valable pour (X,T). Alors $\Pi(\phi)$ est soit vide soit un sous-groupe fermé de G.*

Dem. Soit $a,b \in \Pi(\phi)$. Montrons que $a-b \in \Pi(\phi)$. Soit $\varepsilon > 0$ et $A \in \mathcal{B}_+$. Comme $a \in \Pi(\phi)$, il existe $n \geq 1$ tel que

$$m(A \cap T^{-n}A \cap [\|\phi_n - a\| < \frac{\varepsilon}{2}]) > 0.$$

D'après le corollaire du lemme de Rokhlin (**à éclaircir**), il existe $B \in \mathcal{B}_+$ inclus dans $A \cap T^{-n}A \cap [\|\phi_n - b\| < \frac{\varepsilon}{2}]$ tel que $B \cap T^{-k}B = \emptyset$ pour $1 \leq k \leq n$. Comme b est une valeur persistante, il existe $N \geq 1$ tel que

$$m(B \cap T^{-N}B \cap [\|\phi_N - b\| < \frac{\varepsilon}{2}]) > 0.$$

La condition $B \cap T^{-k}B = \emptyset$ pour $1 \leq k \leq n$, implique $N > n$. Puisque B est inclus dans $[\|\phi_n - a\| < \frac{\varepsilon}{2}]$, on a

$$\begin{aligned} B \cap T^{-N}B \cap [\|\phi_N - b\| &< \frac{\varepsilon}{2}] = B \cap T^{-N}B \cap [\|\phi_n - a\| < \frac{\varepsilon}{2}] \cap [\|\phi_N - b\| < \frac{\varepsilon}{2}] \\ &\subset B \cap T^{-N}B \cap [\left\|\phi_{N-n} \circ T^n - (b-a)\right\| < \varepsilon]. \end{aligned}$$

Finalement, comme T conserve la mesure,

$$\begin{aligned} m(A \cap T^{-(N-n)}A \cap [\left\|\phi_{N-n} - (b-a)\right\| &< \varepsilon]) = m(T^{-n}(A \cap T^{-(N-n)}A \cap [\left\|\phi_{N-n} - (b-a)\right\| < \varepsilon])) \\ &= m(T^{-n}A \cap T^{-N}A \cap [\left\|\phi_{N-n} \circ T^n - (b-a)\right\| < \varepsilon]) \\ &\geq m(B \cap T^{-N}B \cap [\left\|\phi_{N-n} \circ T^n - (b-a)\right\| < \varepsilon]) \\ &\geq m(B \cap T^{-N}B \cap [\|\phi_N - b\| < \frac{\varepsilon}{2}]) \\ &> 0 \end{aligned}$$

donc $b-a$ est une valeur persistante.

Proposition 28. *Supposons que le lemme de Rokhlin soit valable pour (X,T) et que T soit inversible. Alors $E(\phi) = \Pi(\phi) \cup \{0\}$.* ***(admise)***

Théorème 22. *Soit $(X,\mathcal{B},m)$ un espace probabilisé, $T : X \to X$ un isomorphisme de $(X,\mathcal{B},m)$ et $\phi : X \to G$ une application mesurable. Alors $E(\phi) = Per(\phi)$.*

Dem. Montrons que $Per(\phi) \subset E(\phi)$. Supposons que a n'appartienne pas à $E(\phi)$. Alors il existe $A \in \mathcal{B}_+$ tel que $m(A \cap T^{-n}A \cap [\|\phi_n - a\| < \varepsilon]) = 0$ pour tout $n \in \mathbf{Z}$. Posons

$$\begin{aligned} B_1 &= A \times B_G(0, \frac{\varepsilon}{2}),\ B_2 = A \times B_G(a, \frac{\varepsilon}{2}), \\ C_1 &= \cup_{n \in \mathbf{Z}} T_\phi^{-n} B_1,\ C_2 = \cup_{n \in \mathbf{Z}} T_\phi^{-n} B_2. \end{aligned}$$

Montrons que $C_1 \cap C_2$ est de mesure nulle dans $X \times G$. Soit $D = T_\phi^{-m}B_1 \cap T_\phi^{-n}B_2$. Si $n \geq m \geq 0$ alors

$$D = T_\phi^{-m}(B_1 \cap T_\phi^{m-n}B_2)$$

et

$$\begin{aligned} B_1 \cap T_\phi^{m-n} B_2 &= \{(x,b) : x \in A,\ \|b\| < \frac{\varepsilon}{2},\ x \in T^{m-n}A,\ \big\|b + \phi_{n-m}x - a\big\| < \frac{\varepsilon}{2}\} \\ &\subset \{(x,b) : x \in A,\ x \in T^{m-n}A,\ \big\|\phi_{n-m}x - a\big\| < \varepsilon\} \\ &= (A \cap T^{m-n}A \cap [\big\|\phi_{n-m}x - a\big\| < \varepsilon]) \times G. \end{aligned}$$

Donc D est de mesure nulle. Le même raisonnement convient si $m \geq n \geq 0$. Dans le cas général on choisit N un entier assez grand pour que $-N-m$ et $-N-n$ soit négatif et on utilise l'invariance de $\mu = m \times m_G$: $\mu(D) = \mu(T^{-N}D)$. Dans tous les cas on trouve $\mu(D) = 0$, par conséquent $\mu(C_1 \cap C_2) = 0$. Or C_1 et C_2 sont des ensembles T_ϕ invariants de mesures strictement positives et l'image de C_1 par $(x,b) \to (x, b+a)$ est C_2 donc a n'appartient pas à $Per(\phi)$.

Montrons que $E(\phi) \subset Per(\phi)$. Supposons que a n'appartienne pas à $Per(\phi)$. Il existe alors une partie mesurable C de $X \times G$ telle que $T_\phi^{-1}C = C \mod m \times m_G$ et $m \times m_G(C \Delta Q_a C) > 0$.
1. Posons $A = C \backslash Q_a C$ et $B = Q_a A$. Comme les mesures de C et $Q_a C$ sont égales, A est de mesure strictement positive. Les parties A et B sont disjointes par construction. Comme l'image par Q_a d'une partie T_ϕ invariante est T_ϕ invariante et comme les parties invariantes forment une tribu, A et B sont T_ϕ invariante.
2. Pour tout x appartenant à X, posons

$$A_x = \{g \in G : (x,g) \in A\}.$$

Notons que

$$\begin{aligned} g &\in A_{Tx} \Leftrightarrow (Tx, g) \in A \Leftrightarrow T_\phi(x, g - \phi(x)) \in A \\ &\Leftrightarrow (x, g - \phi(x)) \in T_\phi^{-1}A \\ &\Leftrightarrow g \in (T_\phi^{-1}A)_x + \phi(x). \end{aligned}$$

De même, on a pour $n \in \mathbf{Z}$,

$$g \in A_{T^n x} \Leftrightarrow g \in (T_\phi^{-n}A)_x + \phi_n(x).$$

Or $T_\phi^{-1}A = A \mod m \times m_G$ donc pour presque tout x, $A_x = (T_\phi^{-1}A)_x \mod m_G$; par conséquent, pour presque tout x,

$$A_{Tx} = A_x + \phi(x) \mod m_G$$

et

$$m_G(A_{Tx}) = m_G(A_x).$$

La fonction $x \to m_G(A_x)$ est donc T invariante, grâce à l'ergodicité de T on en déduit que $m(A_x)$ est presque sûrement constant sur X. Cette constante n'est pas nulle car la mesure de A est strictement positive.
3. Montrons qu'il existe une partie mesurable D incluse dans A telle que pour presque tout x de X

$$0 < m_G(D_x) < \infty.$$

Pour chaque x de X, on a

$$m_G(A_x) = \lim_{n \to \infty} m_G(A_x \cap B_G(0,n)).$$

Donc pour presque tout x, il existe un entier $n = n(x)$ minimal tel que $m_G(A_x \cap B_G(0,n)) > 0$. Posons

$$D = \{(x,g) \in A : \|g\| \leq n(x)\}.$$

La partie D est mesurable car la fonction $x \to n(x)$ est mesurable (exercice).
4. Considérons la fonction $w : X \times X \times G \to \mathbf{R}_+$ définie par

$$\begin{aligned} w(x,y,g) &= m_G(D_x \cap (D_y + g)) \\ &= \int_G 1_D(x,h) 1_D(y,h-g) dm_G(h) \end{aligned}$$

La fonction w est mesurable. En effet, la fonction $(x,y,g,h) \to 1_D(x,h)1_D(y,h-g)$ est mesurable et le théorème de Fubini donne la mesurabilité de w.
5. Pour x et y fixés, la fonction partielle $g \to w(x,y,g)$ est continue et non identiquement nulle car c'est un produit de convolution de fonctions indicatrices de parties de G de mesures non nulles. Donc, pour tout x, y, l'ensemble $\{g \in G : w(x,y,g) > 0\}$ est de mesure strictement positive. Par conséquent, $\{w > 0\}$ est une partie de mesure strictement positive de $X \times X \times G$. Soit $E = \{(x,y) \in X \times X : m_G(1_{D_x} 1_{D_y}) > 0\}$. On a

$$\begin{aligned} \int_X \int_X w(x,y,0) dm(x) dm(y) &= \int_X \int_X \int_G 1_D(x,h) 1_D(y,h) dm(x) dm(y) dm_G(h) \\ &= \int_G (\int_X 1_D(x,h) dm(x)) (\int_X 1_D(y,h) dm(y)) dm_G(h) \\ &= \int_G (\int_X 1_D(x,h) dm(x))^2 dm_G(h) \end{aligned}$$

Or

$$m \times m_G(D) = \int_G \int_X 1_D(x,h) dm(x) dm_G(h) > 0,$$

donc $\int_X \int_X w(x,y,0) dm(x) dm(y) > 0$. La fonction $(x,y) \to w(x,y,0)$...
6. Montrons l'existence d'une partie Y de X de mesure strictement positive et d'un réel $\delta > 0$ tels que

$$\forall x,y \in Y,\ \forall g \in B_G(0,\delta),\ w(x,y,g) > 0.$$

Comme la tribu de X est la tribu borélienne et que $L^2(G)$ est séparable, d'après le théorème de Lusyn, il existe une partie mesurable Z de X de mesure strictement positive telle que l'application $x \in X \to 1_{D_x} \in L^2(G)$ soit continue sur X. En outre, comme X est à base dénombrable, on peut supposer que Z est inclus dans le support de m. De plus, l'application $\tau : L^2(G) \times G \to L^2(G)$ définie par $\tau(\phi,g)(h) = \phi(g-h)$ est continue donc l'application

$$\begin{aligned} F &: L^2(G) \times L^2(G) \times G \to \mathbf{R} \\ &: (\phi,\psi,g) \to \int_G \phi\, \tau(\psi,g) dm_G \end{aligned}$$

est continue. Or $w(x,y,g) = F(1_{D_x}, 1_{D_y}, g)$ donc w est continue sur $Z \times Z \times G$. Soit $s \in Z$. On sait que $w(x,x,0)$ est strictement positif, par continuité, il existe donc un voisinage $W = Y \times Y \times B_G(0,\varepsilon)$ de $(x,x,0)$ tel que la fonction w soit strictement positive sur W. Finalement, $m(Y)$ est strictement positive car Z est inclus dans le support de m.
7. Pour montrer que a n'appartient pas à $E(\phi)$ il suffit de montrer que pour tout n on a

$$Y \cap T^{-n} Y \cap \{\|\phi_n(x) - a\| < \delta/2\} = \emptyset.$$

Or si x, $T^nx \in Y$ et $\phi_n(x) \in B_G(a, \delta/2)$,

$$m_G(D_{T^nx} \cap (D_x + \phi_n(x) - a)) = w(T^nx, x, \phi_n(x) - a) > 0,$$

donc,

$$m_G((D_{T^nx} + a) \cap (D_x + \phi_n(x))) > 0.$$

Mais

$$D_{T^nx} + a \subset A_{T^nx} + a = B_{T^nx},$$

et

$$D_x + \phi_n(x) \subset A_x + \phi_n(x) = A_{T^nx}$$

pour presque tout x et comme A et B sont des parties disjointes, les sections A_{T^nx} et B_{T^nx} sont disjointes pour tout x ce qui est contradictoire.

5. Liardet Volny

Définition 18. *Soit $(X, \mathcal{A}, \mu, T)$ un système dynamique mesuré. On dit que T est apériodique si la mesure de l'ensemble des point périodique est nulle, c-a-d si*

$$\mu(\{x \in X : \exists n \geq 1,\ T^n(x) = x\}) = 0.$$

Exercice : Soit $(X, \mathcal{A}, \mu, T)$ un système dynamique mesuré inversible et tel que pour tout $\varepsilon > 0$, il existe une partie mesurable de mesure $\leq \varepsilon$. Si T est egordique alors T est apériodique.

Dem. Soit n un entier ≥ 1 et $A_n = \{x \in X : T^nx = x\}$. A_n est une partie mesurable de X. Si $Tx \in A_n$ alors $T^{n+1}x = Tx$ donc $x \in A_n$ et $T^{-1}A_n \subset A_n$. De même, $x \in A_n$ alors $T^{-1}x = T^{n-1}x \in A_n$ donc $A_n \subset T^{-1}A_n$. La partie A_n est donc invariante. Comme T est ergodique, $\mu(A_n) = 0$ ou $\mu(X \backslash A_n) = 0$. Si $\mu(X \backslash A_n) = 0$ alors A_n contient une partie mesurable B de mesure strictement positive $\leq \min(\frac{1}{n+1}, \frac{\mu(X)}{n+1})$. La partie $C = B \cup TB \cup ... \cup T^{n-1}B$ est invariante de mesure strictement positive et strictement inférieure à celle de X ce qui contredit l'ergodicité de T.

Notations : Soit X un espace métrique compact, $T : X \to X$ une application continue.
1. On désigne par $\mathcal{C}(X, \mathbf{R})$ l'espace des applications continue de X dans $\mathbf{R}$.
2. On désigne $M_T(X)$ l'ensemble des probabilités T-invariante sur X.
3. On désigne par $\mathcal{C}_0(X, \mathbf{R})$, l'ensemble des applications continues de X dans $\mathbf{R}$ d'intégrale nulle pour toute probabilité T-invariante, $\mathcal{C}_0(X, \mathbf{R}) = \{f \in \mathcal{C}(X, \mathbf{R}) : \forall \mu \in M_T(X),\ \int f\,d\mu = 0\}$.
4. On désigne par U l'application de $\mathcal{C}(X, \mathbf{R}) \to \mathcal{C}(X, \mathbf{R})$ associée à T, $Uf = f \circ T$.

Théorème 23. *Soit X un espace métrique compact, $T : X \to X$ un homéomorphisme et λ une probabilité T-invariante telle que T soit apériodique. Si $(c_k)_{k \in \mathbf{N}}$ est une suite de réels strictement positifs tendant vers $+\infty$ et telle que $\lim_{k\to\infty} \frac{c_k}{k} = 0$ alors il existe un G_δ dense de fonctions f de $\mathcal{C}_0(X, \mathbf{R})$ telle que toute probabilité ν sur $\mathbf{R}$ soit adhérente pour la convergence en loi à la suite des distributions de*

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

La démonstration utilise deux lemmes et les tours de Rokhlin..

Rappel sur la convergence en loi :

1. Une suite de probabilités (μ_n) sur $\mathbf{R}$ converge en loi vers une probabilité μ si pour toute fonction continue bornée $f : \mathbf{R} \to \mathbf{R}$ la suite de réels $\int f(x)d\mu_n(x)$ converge vers $\int f(x)d\mu(x)$. Cette convergence correspond à une topologie sur l'ensemble probabilités.
2. On démontre que si $\int f(x)d\mu_n(x)$ converge vers $\int f(x)d\mu(x)$ pour toute fonction $f : \mathbf{R} \to \mathbf{R}$ continue à support compact alors (μ_n) converge en loi vers μ.
3. Toute probabilité sur $\mathbf{R}$ est limite en loi d'une suite de probabilités combinaisons linéaires finie de mesures de Dirac.

Proposition 29. *Soit X un espace métrique compact, $T : X \to X$ un homéomorphisme alors $(I - U)\mathcal{C}(X, \mathbf{R})$ est dense dans $\mathcal{C}_0(X, \mathbf{R})$.*

Dem. Par invariance on a clairement l'inclusion $(I - U)\mathcal{C}(X, \mathbf{R}) \subset \mathcal{C}_0(X, \mathbf{R})$. Pour prouver que le sous-espace $F = (I - U)\mathcal{C}_0(X, \mathbf{R})$ est dense dans $\mathcal{C}_0(X, \mathbf{R})$ il suffit d'après le théorème de Hahn-Banach, de montrer que si $\phi : \mathcal{C}_0(X, \mathbf{R}) \to \mathbf{R}$ est forme linéaire continue nulle sur F alors elle est nulle sur $\mathcal{C}_0(X, \mathbf{R})$. Soit ϕ une forme linéaire continue nulle sur F. D'après le théorème de Hahn-Banach, la forme ϕ se prolonge en une forme linéaire continue sur $\mathcal{C}(X, \mathbf{R})$ tout entier. Notons encore ϕ ce prolongement. La forme linéaire ϕ est une mesure de Radon sur le compact X. Appelons ϕ^+ et ϕ^- les parties positives et négatives de ϕ (**Référence ?**). Les mesures ϕ^+ et ϕ^- sont caractérisées par les propriétés suivantes

$$\phi = \phi^+ - \phi^-$$

et si ϕ_1 et ϕ_2 sont deux mesures de Radon telles que

$$\phi = \phi_1 - \phi_2$$

alors $\phi_1 - \phi^+$ et $\phi_2 - \phi^-$ sont deux mesures de Radon positives. Or si $f \in \mathcal{C}(X, R)$ alors $\phi(f - f \circ T) = 0$ et

$$\phi(f \circ T) = \phi^+(f \circ T) - \phi^-(f \circ T) = \phi(f) = \phi^+(f) - \phi^-(f),$$

donc les mesures $\mu^+(f) = \phi^+(f \circ T) - \phi^+(f)$ et $\mu^-(f) = \phi^-(f \circ T) - \phi^-(f)$ sont positives. Mais $\mu^+(1) = \phi^+(1 \circ T) - \phi^+(1) = 0$ donc μ^+ est nulle et ϕ^+ est invariante. De même, ϕ^- est invariante. Donc par définition de $\mathcal{C}_0(X, \mathbf{R})$, pour tout $f \in \mathcal{C}_0(X, \mathbf{R})$, $\phi^+(f) = \phi^-(f) = 0$ et ϕ est nulle sur $\mathcal{C}_0(X, \mathbf{R})$. $\square$

Lemme 11. *Soit X un espace métrique compact, E un sous-espace fermé de $\mathcal{C}(X, \mathbf{R})$ muni de la norme infini, $V : \mathcal{C}(X, \mathbf{R}) \to E$ une application linéaire d'image dense, $(A_k)_{k \geq 1}$ une suite d'application linéaire de $\mathcal{C}(X, \mathbf{R})$ dans lui même et $(a_k)_{k \geq 1}$ une suite de réels positifs convergeant vers 0 telle que*

$$\forall f \in \mathcal{C}(X, \mathbf{R}), \ \|A_k V f\|_\infty \leq a_k \|f\|_\infty .$$

Soit λ une probabilité sur X et ν une probabilité sur $\mathbf{R}$. Si il existe une suite $(f_k)_{k \geq 1}$ de fonctions appartenant à E telle que $\|f_k\|_\infty$ tendent vers 0 et $A_k f_k$ converge en loi vers ν alors il existe un G_δ dense H inclus dans E tel que pour chaque $f \in H$ il existe une suite d'entiers $(n_k)_k$ tendant vers l'infini pour laquelle la suite des images de λ par $A_{n_k} f$ converge en loi vers ν quand k tend vers $+\infty$.

Dem. Soit $(\psi_p)_p$ une suite d'applications de $\mathbf{R} \to \mathbf{R}$ continues à supports compacts et dense dans l'espace des applications continues à support compact muni de la norme uniforme. Pour montrer qu'une suite (μ_k) de probabilités sur $\mathbf{R}$ converge en loi vers ν il suffit de montrer que $\mu_k(\psi_p)$ tend vers $\nu(\psi_p)$ pour tout p quand k tend vers l'infini.
1. Comme $A_k f_k$ converge en loi vers ν, pour chaque entier $p \geq 1$, il existe un entier k_p tel que

$$\forall i \leq p,\ \forall k \geq k_p \ \left| \int_{\mathbf{R}} \psi_i d\nu - \int_X \psi_i(A_k f_k) d\lambda \right| \leq \frac{1}{p}.$$

Pour $k = k_p$ on obtient,

$$\forall i \leq p,\ \left| \int_{\mathbf{R}} \psi_i d\nu - \int_X \psi_i(A_{k_p} f_{k_p}) d\lambda \right| \leq \frac{1}{p},$$

de plus, on peut supposer que la suite (k_p) est strictement croissante cela veut dire qu'a une extraction près on a

$$\forall i \leq p,\ \left| \int_{\mathbf{R}} \psi_i d\nu - \int_X \psi_i(A_p f_p) d\lambda \right| \leq \frac{1}{p}.$$

2. Comme pour chaque entier i et chaque entier p, l'application $f \in \mathcal{C}(X, \mathbf{R}) \to \psi_i(A_p f) \in \mathcal{C}(X, \mathbf{R})$ est continue, pour tout p, il existe un voisinage ouvert $\mathcal{U}_p$ de f_p dans E muni de la norme uniforme tel que

$$\forall f \in \mathcal{U}_p,\ \forall i \leq p,\ \left| \int_{\mathbf{R}} \psi_i(A_p f) d\lambda - \int_X \psi_i(A_p f_p) d\lambda \right| \leq \frac{1}{p}.$$

3. Soit $(r_k)_k$une suite de réels positifs tendant vers $+\infty$ telle que $a_k r_k$ tende vers 0. Posons pour $n \geq 1$, $H_n = \cup_{k \geq n}(\mathcal{U}_k + V(B_E(0, r_k)))$. D'une part, comme chaque $\mathcal{U}_k$ est ouvert, H_n est un ouvert. D'autre part, H_n est dense car si $f \in E$ et ε est un nombre strictement positif alors comme V est d'image dense, il existe $g \in \mathcal{C}(X, \mathbf{R})$ tel que $\|f - Vg\|_\infty \leq \varepsilon$. En choisissant $k \geq n$ tel que $\|g\|_\infty < r_k$ et $\|f_k\| \leq \varepsilon$ on a $f_k + Vg \in \mathcal{U}_k + V(B_E(0, r_k))$ et

$$\|f - (f_k + Vg)\|_\infty \leq 2\varepsilon$$

donc il existe $h \in H_n$ tel que $\|f - h\|_\infty \leq 2\varepsilon$. Finalement, $H = \cap_{n \geq 1} H_n$ est un G_δ dense de E.
4. Si f appartient à H alors pour tout entier n, il existe un entier $p_n \geq n$ tel que $f \in \mathcal{U}_{p_n} + V(B_E(0, r_{p_n}))$. Il existe donc $g_n \in \mathcal{U}_{p_n}$ et $h_n \in B_E(0, r_{p_n})$ telle que $f = g_n + V h_n$. Soit i un entier, pour n assez grand $p_n \geq i$ donc

$$\begin{aligned} \left| \int_{\mathbf{R}} \psi_i d\nu - \int_X \psi_i(A_{p_n} f) d\lambda \right| &\leq \left| \int_{\mathbf{R}} \psi_i d\nu - \int_X \psi_i(A_{p_n} f_{p_n}) d\lambda \right| + \left| \int_{\mathbf{R}} \psi_i(A_{p_n} f_{p_n}) d\nu - \int_X \psi_i(A_{p_n} f) d\lambda \right| \\ &\leq \frac{1}{p_n} + \left| \int_{\mathbf{R}} \psi_i(A_{p_n} f_{p_n}) d\nu - \int_X \psi_i(A_{p_n} g_n + A_{p_n} V h_n) d\lambda \right| \end{aligned}$$

et comme $\|A_{p_n} h_n\|_\infty \leq a_{p_n} \|h_n\|_\infty \leq a_{p_n} r_{p_n}$ tend vers 0 quand n tend vers ∞, on a grâce à l'uniforme continuité de ψ_i, $\lim_{n \to \infty} \int_X \psi_i(A_{p_n} f) d\lambda = \int_{\mathbf{R}} \psi_i d\nu$. $\square$

Démonstration du théorème. Considérons l'ensemble $\mathcal{P}_{\mathbf{Q}}$ des probabilités μ sur $\mathbf{R}$ telles que
- $\mu = \sum_{i=1}^m a_i \delta_{b_i}$ où tous les a_i et les b_i sont rationnels,

- μ est de moyenne nulle : $\sum_{i=1}^{m} a_i b_i = 0$.
L'ensemble de ces probabilités est dénombrable et on vérifie facilement qu'il est dense pour la convergence en loi dans l'ensemble des probabilités sur $\mathbf{R}$.

On va démontrer que si la probabilité $\nu = \sum_{i=1}^{m} a_i \delta_{b_i}$ appartient à $\mathcal{P}_{\mathbf{Q}}$, alors l'ensemble $\mathcal{F}_\nu$ des fonctions f de $\mathcal{C}_0(X, \mathbf{R})$ telles qu'il existe une suite d'entiers $(n_k)_k$ tendant vers l'infini pour laquelle la suite des distributions des applications

$$\frac{1}{c_{n_k}} \sum_{i=0}^{n_k-1} f \circ T^i$$

converge en loi vers ν, est un G_δ dense de $\mathcal{C}_0(X, \mathbf{R})$. Le théorème en résultera car si $f \in \cap_{v \in \mathcal{P}_{\mathbf{Q}}} \mathcal{F}_\nu$ alors toutes les probabilité ν de $\mathcal{P}_{\mathbf{Q}}$ sont des valeurs d'adhérences de la suite des distributions des $\frac{1}{c_{n_k}} \sum_{i=0}^{n_k-1} f \circ T^i$ et l'ensemble des valeurs d'adhérences d'une suite est un fermé.

Soit $\nu = \sum_{i=1}^{m} a_i \delta_{b_i}$ une probabilité appartenant à $\mathcal{P}_{\mathbf{Q}}$. En réduisant au même dénominateur on peut trouver des entiers $p_1, ..., p_m$ et $q > 0$ tels que $a_i = \frac{p_i}{q}$, on a alors $\sum_{i=1}^{m} p_i = q$ et $\sum_{i=1}^{m} b_i p_i = 0$; de plus, on peut supposer que $m \leq q$ (on peut multiplier q un entier quelconque). Soit $\varepsilon > 0$ et $n_0 > 0$ un entier. La mesure λ est apériodique donc d'après le lemme de Rokhlin-Halmous pour tout $\delta > 0$ il existe un borélien V de X et un entier n supérieur à n_0 tel que les parties $V, T^{-1}V, ..., T^{-qn+1}V$ soient deux à deux disjointes et de complémentaires de mesure inférieure à δ. Comme X est compact la mesure λ est régulière et $\lambda(V)$ est le sup des mesure des fermés F inclus dans V. On peut donc supposer que V est un fermé de X. Comme T est continue on peut trouver un ouvert ω contenant F tel que $\omega, T^{-1}\omega, ..., T^{-qn+1}\omega$ soient deux à deux disjoints. Il existe une fonction continue $\psi : X \to [0, 1]$ telle que $\psi = 1$ sur V et $\psi = 0$ sur $X \backslash \omega$. Cette fonction vérifie $\lambda(\sum_{i=0}^{qn-1} \psi \circ T^i) \geq 1 - \delta$.
Considérons les fonctions

$$\phi = b_1 \sum_{j=0}^{p_1-1} \psi \circ T^{jn} + ... + b_m \sum_{j=p_1+...+p_{m-1}}^{p_1+...+p_m} \psi \circ T^{jn},$$

et

$$f_k = \frac{c_k}{k} \sum_{j=0}^{n-k} \phi \circ T^j$$

où k est un entier ≥ 1. Par construction, la fonction ϕ prend la valeur b_i sur

$$U_i = \cup_{j=p_1+...+p_{i-1}}^{p_1+...+p_i-1} T^{-jn} V$$

dont la mesure vaut $p_i \lambda(V)$ et les U_i sont des parties deux à deux disjointes incluses dans $\Omega = \cup_{j=0}^{q-1} T^{-jn}\omega$. De plus, ϕ est de moyenne nulle pour toute probabilité $\widetilde{\lambda}$ T-invariante.

En effet,

$$\begin{aligned}
\int \phi\, d\widetilde{\lambda} &= b_1 \sum_{j=0}^{p_1-1} \int \psi \circ T^{jn} d\widetilde{\lambda} + ... + b_m \sum_{j=p_1+...+p_{m-1}}^{p_1+...+p_m} \int \psi \circ T^{jn} d\widetilde{\lambda} \\
&= b_1 \sum_{j=0}^{p_1-1} \int \psi d\widetilde{\lambda} + ... + b_m \sum_{j=p_1+...+p_{m-1}}^{p_1+...+p_m} \int \psi d\widetilde{\lambda} \\
&= \sum_{i=1}^{m} b_i p_i \int \psi d\widetilde{\lambda} = 0.
\end{aligned}$$

Donc f_k appartient à $\mathcal{C}_0(X, \mathbf{R})$. Examinons la distribution μ_k de la fonction $F_k = \frac{1}{c_k} S_k(f_k)$. On a

$$F_k = \frac{1}{k} \sum_{i=0}^{k-1} \sum_{j=0}^{n-k} \phi \circ T^{j+i} = \frac{1}{k} \sum_{l=0}^{n-1} A_l \phi \circ T^l$$

où lorsque $k \leq 2n$,

$$A_l = \left\{ \begin{array}{c} l+1 \text{ si } l \in \{0, ..., k-2\} \\ k \text{ si } l \in \{k-1, ..., n-k\} \\ n-l \text{ si } l \in \{n-k+1, ..., n-1\} \end{array} \right. .$$

Soit $i \in \{1, ..., m\}$. Comme les parties $T^{-l}\Omega$, $l = 0, ..., n-1$, sont deux à deux disjoints, sur l'ensemble

$$\cup_{l=k-1}^{n-k} T^{-l} U_i$$

la fonction F_k prend la valeur b_i. Par conséquent,

$$\begin{aligned}
\mu_k(\{a_i\}) &\geq \lambda(\cup_{l=k-1}^{n-k} T^{-l} U_i) = (n-2k)\lambda(U_i) = (n-2k) p_i \lambda(V) \\
&\geq (n-2k) p_i \frac{1-\delta}{nq} \\
&= \frac{(n-2k)(1-\delta)}{n} \times \frac{p_i}{q} = \frac{(n-2k)(1-\delta)}{n} a_i
\end{aligned}$$

et la norme de la mesure $\mu_k - \nu$ est inférieure à $1 - \frac{(n-2k)(1-\delta)}{n}$. Comme $1 - \frac{(n-2k)(1-\delta)}{n}$ tend vers 0 quand n tend vers l'infini et δ vers 0 on peut trouver une suite $(f_k)_k$ d'éléments de $\mathcal{C}_0(X, \mathbf{R})$ de limite nulle telle que la suite des distributions de $\frac{1}{c_k} S_k(f_k)$ converge vers ν. On conclut grâce au lemme précédent. $\square$

Corollaire 12. *Soit X un espace métrique compact, $T : X \to X$ un homéomorphisme et λ une probabilité T-invariante telle que T soit apériodique. Soit F l'ensemble des fonctions f appartenant à $\mathcal{C}_0(X, \mathbf{R})$ qui sont des cobords de fonctions mesurables, c-a-d les fonctions f pour lesquelles il existe $g : X \to \mathbf{R}$ mesurable telle que*

$$f = g - g \circ T$$

λ-presque sûrement. Alors l'ensemble F est maigre (inclus dans une réunion dénombrable de fermés d'intérieurs vides) dans $\mathcal{C}_0(X, \mathbf{R})$ muni de la norme uniforme.

Dem. D'après le théorème, pour montrer que F est maigre il suffit de montrer que si g est mesurable alors la mesure de Dirac en 1 δ_1, n'est pas valeur d'adhérence de la suite des distributions de $\frac{1}{\sqrt{n}}S_n(f)$ où $f = g - g \circ T$ λ-p.s.. Soit $\phi : \mathbf{R} \to \mathbf{R}$ une fonction uniformément continue bornée. On a

$$\int \phi(\frac{1}{\sqrt{n}}S_n(f))\, d\lambda = \int \phi(\frac{1}{\sqrt{n}}(g - g \circ T^n)\, d\lambda$$

car $S_n(f) = g - g \circ T^n$ λ-p.s.. Pour tout x appartenant à X, $\lim_{n\to\infty} \left| \phi(\frac{1}{\sqrt{n}}(g(x) - g \circ T^n(x)) - \phi(-\frac{1}{\sqrt{n}} g \circ T^n(x)) \right| = 0$. Par conséquent, grâce au théorème de convergence dominée on obtient

$$\lim_{n\to\infty} \int \left| \phi(\frac{1}{\sqrt{n}}(g - g \circ T^n)) - \phi(-\frac{1}{\sqrt{n}} g \circ T^n) \right| d\lambda = 0.$$

Or par invariance de la mesure λ on a

$$\int \phi(-\frac{1}{\sqrt{n}} g \circ T^n) d\lambda = \int \phi(-\frac{1}{\sqrt{n}} g) d\lambda,$$

donc grâce au théorème de convergence dominée,

$$\lim_{n\to\infty} \int \phi(\frac{1}{\sqrt{n}}S_n(f))\, d\lambda = \lim_{n\to\infty} \int \phi(-\frac{1}{\sqrt{n}} g) d\lambda = \phi(0).$$

Ce qui prouve la convergence en loi de la suite $(\frac{1}{\sqrt{n}}S_n(f))$ vers δ_0. $\square$

Le corollaire suivant est immédiat.

Corollaire 13. *Soit X un espace métrique compact et $T : X \to X$ un homéomorphisme. Si T est uniquement ergodique alors l'ensemble des cobords $f - f \circ T$ avec $f : X \to \mathbf{R}$ continue, est maigre dans $\mathcal{C}_0(X, \mathbf{R})$ muni de la convergence uniforme.*

6. Bibliographie

[Aa] : J. Aaronson, An Introduction to Infinite Ergodic Theory, A.M.S. Mathematical Surveys and Monographs 50 (1997).

[Doob] : J. L. Doob, Measure Theory, Springer, Graduate Texts in Mathematics (1994).

[Fu] : H. Furstenberg, *Strict ergodicity and transformation of the torus,* American Journal of Mathematics, Vol 83 (1961), 573-601.

[G. H] : W. H. Gottschalk, G. A. Hedlund, Topoligical Dynamics, Amer. Math. Soc. Coll. Publ., Vol. 36, Providence (1955).

[L,V] : P. Liardet, D. Volný, *Sums of continuous differentiable functions in dynamical systems,* Israel Journal of Mathematics 98 (1997), 29-60.

[Kr] : W. Krieger, *On entropy and generators of measure-preserving transformations.* Trans. Amer. Math. Soc. **149**, 453-464 (1970) erratum **168**, 519 (1972).

[Shm] : K. Schmidt, *Cocycles of Ergodic Transformation Groups,* Lect. Notes in Maths. Vol. 1, Mac Millan Co. of India (1977).