Principe Variationnel des Entropies

Adrien Ellis, Danylo Khilko

Sous la direction de Maxence Novel

Résumé

En général, la théorie des systèmes dynamiques discrets étudie l'évolution d'un système donné par un espace de phase X et une transformation T. On a besoin de distinguer les systèmes "simples" des systèmes "compliqués", ce que l'on fait d'une certaine manière à travers la notion d'entropie. L'entropie est un nombre réel positif qui caractérise la complexité d'un système. Nous considérons deux cadres différents. Dans le premier cas nous avons un espace mesurable $(X, \mathcal{A}, \mu)$ où $\mathcal{A}$ est une tribu, μ est une mesure de probabilité et T préserve la mesure. Dans le second cas nous avons un espace métrique X et une transformation continue $T : X \to X$. Nous définissons respectivement dans chacun de ces contextes l'entropie métrique $h_\mu(T)$ et l'entropie topologique $h(T)$. Le but de cet exposé est de prouver le lien entre ces deux définitions d'entropies, donné par le principe variationnel : $h(T) = \sup h_\mu(T)$, où le supremum est pris sur toutes les mesures μ appropriées.

1 Introduction

Tout au long de ce mémoire, nous noterons X l'espace de phase étudié, qui sera soit un espace métrique compact muni d'une distance d, soit un espace de probabilité muni d'une tribu $\mathcal{A}$ et d'une mesure de probabilité μ. On notera T une transformation de X dans lui-même : le concept d'entropie peut être interprété comme une quantification de la capacité de T à mélanger les points et les parties de X. On commencera par définir l'entropie métrique dans le cadre d'un espace de probabilité puis l'on introduira l'entropie topologique dans le cas où X est un espace métrique. On impose des hypothèses élémentaires de régularité à T, que l'on exige mesurable lorsque X est un espace de probabilité, continue lorsque X est un espace métrique, et à la fois mesurable et continue dans la partie finale où l'on exprime un lien entre les deux définitions de l'entropie. De plus, dans le cas de l'entropie métrique, on demande à ce que T préserve μ au sens précis de la définition ci-dessous :

Définition 1.1. On dit que μ préserve T, ou que μ est T-invariante si

$$\forall A \in \mathcal{A}, \quad \mu(T^{-1}A) = \mu(A).$$

1.1 Dynamique symbolique

Nous illustrerons les concepts développés dans ce document en utilisant le modèle jouet de la dynamique symbolique, qui aide à construire de nombreux exemples de systèmes dynamiques.

Définition 1.2. On définit l'espace de phases Σ du *décalage* sur l'*alphabet* $\{0,1\}$ par

$$\Sigma = \{x = \{x_i\}_{i=0}^{\infty} : x_i \in \{0,1\} \text{ pour tout } i \in \mathbb{Z}_+\},$$

où l'on munit $\{0,1\}$ de la topologie discréte et Σ de la topologie produit. Les éléments de l'alphabet $\{0,1\}$ sont appelés *symboles*.

On définit sur Σ une transformation continue dont l'on étudiera l'entropie :

Définition 1.3. l'*application de décalage* $\sigma : \Sigma \to \Sigma$ est une surjection continue donnée par

$$x = \{x_i\}_{i=0}^{\infty} \mapsto \sigma x = \{x_{i+1}\}_{i=0}^{\infty},$$

autrement dit $\sigma(x)$ est une suite obtenue en supprimant le premier élément de x.

On note $x = \{x_i\}_{i=0}^{\infty} = x_0x_1x_2x_3\dots$ un élément x de Σ.
La définition suivante introduit de la terminologie de base relative au décalage :

Définition 1.4.

1. Un *bloc* w sur Σ est une suite finie de symboles et sa *longueur* est le nombre de ses symboles (notée $|w|$). Un n-*bloc* est un bloc de longueur n. En général nous sommes interessés seulement par des blocs w avec $|w| \geq 1$ et nous signifions par Σ^* l'ensemble de tous les blocs sur Σ.
2. Le bloc w est un *sous-bloc* du bloc $v = v_1 \dots v_m$ avec $v_1, \dots, v_m \in A$ s'il existe $1 \leq i \leq j \leq m$ tels que $w = v_i \dots v_j$.
3. La concaténation de deux blocs $u = a_1 \dots a_k$ et $v = b_1 \dots b_l$ est le bloc $uv = a_1 \dots a_k b_1 \dots b_l$. On ecrit u^n pour la concaténation de $n \geq 1$ copies du bloc u, et u^∞ pour la suite $uuu \cdots \in \Sigma$.
4. On dénote $x_{[i,j]}$ le bloc $x_i x_{i+1} \dots x_j$, où $0 \leq i \leq j$ et $x = \{x_k\}_{k=0}^{\infty} \in \Sigma$.
5. Le *cylindre* du n-bloc $w \in \Sigma^*$ est l'ensemble $C[w] = \{x \in X : x_{[0,n-1]} = w\}$.

Remarque 1.1. La collection de tous les cylindres forme la base de la topologie de Σ.

On munit ensuite Σ d'une distance d :

Définition 1.5. Pour $x \neq y \in \Sigma$ on pose $d(x,y) = 1/(n+1)$ où $n = \max\{k : x_{[0,k-1]} = y_{[0,k-1]}\}$ et $d(x,y) = 0$ si $x = y \in \Sigma$. Il est facile de vérifier que d est une métrique sur Σ et que d engendre la topologie définie plus tôt. Donc Σ est un espace métrique.

On peut munir Σ de nombreuses façons d'une mesure de probabilité borélienne préservée par σ. On en présente deux :

1. Soit δ_{0^∞} la mesure de Dirac au point 0^∞. On voit que si A est borélienne, $\mu(A) = 1_A(0^\infty) = 1_{\sigma^{-1}A}(0^\infty) = \mu(\sigma^{-1}A)$ parce que $0^\infty \in A$ ssi $0^\infty \in \sigma^{-1}A$. On conclut que δ_{0^∞} est prşervée par σ.
2. Soient $p_0, p_1 \in [0,1]$ telles que $p_0 + p_1 = 1$. Soit $w = a_0 \dots a_k$, $k \geq 0$. Pour cylindre $C[w]$ on pose $l(C[w]) = \prod_{i=0}^{k} p_{a_i}$.

$$\lambda^*(A) = \inf\{\sum_{k=1}^{\infty} l(C[w_k]),\ \text{où } C[w_k] \text{ est une suite de cylindres avec } A \subset \bigcup_{k=1}^{\infty} C[w_k]\}.$$

λ^* est la mesure extérieure et $\lambda^*(C[w]) = l(C[w])$. Les cylindres engendrent la tribu borélienne de Σ, donc λ^* engendre une mesure λ sur la tribu borélienne de Σ.

On vérifie que σ préserve λ. Par la définition de λ^* il suffit de le faire pour les cylindres :

$$\lambda(\sigma^{-1}C[w]) = \lambda(C[0w] \cup C[1w]) = \lambda(C[0w]) + \lambda(C[1w]) = p_0\lambda(C[w]) + p_1\lambda(C[w]) = \lambda(C[w]).$$

On peut interpréter cette dernière mesure en termes probabilistes : soit X une variable aléatoire tel que $\mathbb{P}(X = 0) = p_0$, $\mathbb{P}(X = 1) = p_1$. Σ peut être vu comme l'espace de probabilité des suites de tirages de X. On dit que λ est engendrée par p_0 et p_1.

2 Entropie métrique

2.1 La notion d'entropie métrique

Soit $(X, \mathcal{A}, \mu)$ un espace probabilisé.

Définition 2.1. Nous disons qu'une famille finie $\xi = \{C_1, \dots, C_k\}$ de sous-ensembles mesurables de X est une partition si :

1. $\mu(\cup_{i=1}^{k} C_i) = 1$.
2. $\mu(C_i \cap C_j) = 0$ pour tout $i \neq j$.

Nous soulignons que l'on considére seulement les partitions finies, sur lesquelles ont défini une relation d'équivalence :

Définition 2.2. On dit que deux partitions finies ξ et η sont équivalentes si

$$\forall A \in \xi,\ B \in \eta, \quad \mu(A \cap B) = \mu(A) = \mu(B) \quad \text{ou} \quad \mu(A \cap B) = 0$$

Remarque 2.1. Pout toute partition ξ, on peut trouver une partition équivalente η dont les éléments sont deux à deux disjoints. En effet si $\xi = \{C_1, \dots, C_n\}$ on définit $\eta = \{C'_1, \dots, C'_n\}$ par :

$$C'_i = C_i \setminus \left(\bigcup_{i<j} C_j\right) \quad \text{et} \quad C'_n = C_n \cup \left(\bigcup_{i<j} C_j\right)^c$$

Définition 2.3. L'entropie de la partition mesurable ξ est donnée par

$$H_\mu(\xi) = -\sum_{C\in\xi} \mu(C)\log(\mu(C)),$$

avec la convention $0 \cdot \log 0 = 0$.

Remarque 2.2. Deux partitions équivalentes ont la même entropie. On supposera donc librement que les partitions mesurables étudiées sont des partions au sens classique.

On définit la fonction $\psi : [0,1] \to \mathbb{R}$ par

$$\psi(x) = x\log x,\ \ x > 0$$
$$\psi(0) = 0.$$

Evidemment, $H_\mu(\xi) = -\sum_{C\in\xi} \psi(\mu(C))$.

Nous avons aussi $\psi''(x) = 1/x$ pour tout $x > 0$ donc ψ est convexe.Par l'inégalité de Jensen, si $a_1, \dots, a_k$ sont des reéls positifs tel que $a_1 + \cdots + a_k = 1$, alors

$$\psi(a_1x_1 + \cdots + a_kx_k) \leq a_1\psi(x_1) + \cdots + a_k\psi(x_k).$$

Remarque 2.3. En particulier, si k est égale au cardinal de ξ

$$H_\mu(\xi) = -k\sum_{C\in\xi} 1/k \cdot \psi(\mu(C)) \leq k \cdot \psi(1/k(\sum_{C\in\xi} \mu(C))) = -k\psi(1/k) = \log k.$$

Certaines partitions peuvent être plus fine que autres.

Définition 2.4. Soient ξ et η deux partitions. Nous disons que ξ est plus fine que η (noté $\eta \preccurlyeq \xi$) si pour tout $D \in \xi$ il existe $C \in \eta$ tel que $\mu(D \setminus C) = 0$. Autrement dit, ξ est plus fine que η si chaque élément de η est une réunion d'éléments de ξ modulo les ensembles de mesure 0.

Propriété 2.1. *Soient ξ et η deux partitions. Si $\xi \preccurlyeq \eta$ alors $H_\mu(\eta) \geq H_\mu(\xi)$.*

Démonstration. Comme $\xi \preccurlyeq \eta$, on a

$$H_\mu(\eta) = -\sum_{D\in\eta} \psi(\mu(D)) = -\sum_{C\in\xi}\sum_{D\subset C} \psi(\mu(D))$$

Clairement pour tout $a_1, \dots, a_m \in \mathbb{R}_+$

$$(a_1 + \cdots + a_m)\log(a_1 + \cdots + a_m) \geq a_1\log a_1 + \cdots + a_m\log a_m.$$

Donc, puisque $\sum_{D\subset C} \mu(D) = \mu(C)$,

$$\psi(\mu(C)) = \psi\left(\sum_{D\subset C} \mu(D)\right) \geq \sum_{D\subset C} \psi(\mu(D)).$$

Finalement,

$$H_\mu(\xi) = -\sum_{C\in\xi} \psi(\mu(C)) \leq -\sum_{C\in\xi}\sum_{D\subset C} \psi(\mu(D)) = H_\mu(\eta).$$

□

Définition 2.5. Soient ξ et η deux partitions. Nous definissons leur partition jointe comme

$$\xi \vee \eta = \{C \cap D : C \in \xi, D \in \eta\}.$$

Propriété 2.2. *Si ξ et η sont deux partitions,*

$$H_\mu(\xi \vee \eta) \leq H_\mu(\xi) + H_\mu(\eta).$$

Démonstration.

$$\begin{aligned}
H_\mu(\xi \vee \eta) &= -\sum_{D\in\eta, C\in\xi} \mu(C \cap D) \log \mu(C \cap D) \\
&= -\sum_{D\in\eta} \mu(D) \sum_{C\in\xi} \frac{\mu(C \cap D)}{\mu(D)} \log \mu(C \cap D) \\
&= -\sum_{D\in\eta} \mu(D) \sum_{C\in\xi} \frac{\mu(C \cap D)}{\mu(D)} \left(\log \frac{\mu(C \cap D)}{\mu(D)} - \log \mu(D)\right) \\
&= -\sum_{D\in\eta} \mu(D) \sum_{C\in\xi} \frac{\mu(C \cap D)}{\mu(D)} \log \frac{\mu(C \cap D)}{\mu(D)} + H_\mu(\eta) \\
&= -\sum_{C\in\xi} \sum_{D\in\eta} \mu(D) \psi\left(\frac{\mu(C \cap D)}{\mu(D)}\right) + H_\mu(\eta) \\
&\leq -\sum_{C\in\xi} \psi\left(\sum_{D\in\eta} \mu(D) \frac{\mu(C \cap D)}{\mu(D)}\right) + H_\mu(\eta) \quad \text{par l'inégalité de Jensen} \\
&\leq H_\mu(\xi) + H_\mu(\eta)
\end{aligned}$$

□

Maintenant, pour une partition ξ donnée nous allons considérer la partition $T^{-1}\xi = \{T^{-1}C : C \in \xi\}$.

Propriété 2.3. *Soit ξ une partition. Alors*

$$H_\mu(\xi) = H_\mu(T^{-1}\xi).$$

Démonstration. Puisque μ est T-invariant,

$$H_\mu(\xi) = -\sum_{C\in\xi} \psi(\mu(C)) = -\sum_{C\in\xi} \psi(\mu(T^{-1}C)) = H_\mu(T^{-1}\xi).$$

□

Nous avons aussi besoin du résultat général suivant.

Lemme 2.1. *Si $(a_n)_{n=0}^{+\infty}$ est une suite de nombres reéls positifs tel que pour tout n, $m \in \mathbb{N}$, $a_m + a_n \geq a_{m+n}$ (suite sous-additive), alors*

$$\lim_{n\to+\infty} \frac{a_n}{n} = \inf\{\frac{a_n}{n} : n \in \mathbb{N}\} \geq 0$$

et la limite de $\left(\frac{a_n}{n}\right)_{n\in\mathbb{N}}$ existe.

Démonstration. On fixe $k \in \mathbb{N}$. Puis on écrit $n = qk + r$ pour un $q \in \mathbb{N}$ et $r \in \{0, \ldots, k-1\}$. On a

$$\frac{a_n}{n} \leq \frac{a_{kq} + a_r}{kq + r} \leq \frac{qa_k + a_r}{kq + r}.$$

Donc si $n \to +\infty$ nous obtenons

$$\limsup_{n \to +\infty} \frac{a_n}{n} \leq \frac{a_k}{k}.$$

Finalement,

$$\limsup_{n \to +\infty} \frac{a_n}{n} \leq \inf\{\frac{a_k}{k}\} \leq \liminf_{n \to +\infty} \frac{a_n}{n},$$

et la limite existe. □

Théorème 2.1. *Si $T : X \to X$ est une application mesurable qui préserve la mesure μ et ξ est une partition mesurable de X, alors*

$$\lim_{n \to +\infty} \frac{1}{n} H_\mu(\xi_n) = \inf_{n \in \mathbb{N}} \frac{1}{n} H_\mu(\xi_n),$$

où

$$\xi_n = \bigvee_{i=0}^{n-1} T^{-k}\xi.$$

Démonstration. Par les lemmes précédents

$$H_\mu(\xi_{n+m}) = H_\mu(\xi_n \vee T^{-n}\xi_m) \leq H_\mu(\xi_n) + H_\mu(T^{-n}\xi_m) = H_\mu(\xi_n) + H_\mu(\xi_m).$$

Donc $H_\mu(\xi_n)$ satisfait les conditions du lemme. Donc on obtient l'existence de

$$\lim_{n \to +\infty} \frac{1}{n} H_\mu(\xi_n).$$

□

Ce théorème confirme que la définition suivante est valide.

Définition 2.6. Soit $T : X \to X$ une application mesurable preservant la mesure. Nous définissons l'entropie métrique de T par rapport à μ et à la partition ξ par

$$h_\mu(T, \xi) = \lim_{n \to +\infty} \frac{1}{n} H_\mu(\xi_n)$$

où ξ_n est définie comme précédement. Nous definissons l'entropie de T par rapport à μ par

$$h_\mu(T) = \sup_\xi h_\mu(T, \xi)$$

.

On a la propriété importante suivante sur l'entropie des puissances de T.

Propriété 2.4. $h_\mu(T^k) = kh_\mu(T)$ *pour tout* $k \in \mathbb{N}^*$.

Démonstration. On a

$$\frac{1}{n}H_\mu(\bigvee_{l=0}^{n-1} T^{-kl}\xi_k) = \frac{1}{n}H_\mu(\bigvee_{i=0}^{nk-1} T^{-i}\xi).$$

Par $n \to \infty$

$$h_\mu(T^k, \xi_k) = kh_\mu(T, \xi).$$

Donc

$$h_\mu(T^k) \geq \sup_\xi h_\mu(T^k, \xi_k) = \sup_\xi kh_\mu(T, \xi) = kh_\mu(T).$$

En revanche, $\bigvee_{i=0}^{n-1} T^{-ki}\xi$ est moins fine que $\bigvee_{i=0}^{kn-1} T^{-i}\xi$, donc

$$h_\mu(T^{-k}, \xi) \leq \lim_{n\to\infty} \frac{1}{n}H_\mu(\bigvee_{i=0}^{kn-1} T^{-i}\xi) = kh_\mu(T, \xi),$$

d'où

$$h_\mu(T^{-k}) \leq kh_\mu(T).$$

□

2.2 Entropie conditionnelle

Nous introduisons maintenant le concept d'entropie conditionelle d'une partition par rapport à une autre, qui peut être interprété comme la quantité d'information apportée par une partition étant donné l'information correspondant à une autre partition.

Définition 2.7. Soit ξ et ζ deux partitions mesurables de X. On définit l'entropie conditionelle $H(\xi|\zeta)$ de ξ par rapport à ζ en posant :

$$H(\xi|\zeta) = -\sum_{C\in\xi,\, D\in\zeta} \mu(C\cap D)\log\frac{\mu(C\cap D)}{\mu(D)}.$$

Nous exposons maintenant une deuxième caractérisation de l'entropie métrique qui utilise l'entropie conditionelle et qui nous sera utile par la suite.

Théorème 2.2. *Soit $T : X \to X$ une application mesurable continue préservant une mesure μ sur X et ξ une partition mesurable sur X, alors :*

$$h_\mu(T, \xi) = \lim_{n\to\infty} H_\mu\left(\xi|\bigvee_{i=1}^{n} T^{-i}\xi\right).$$

Démonstration. On commence par établir le fait que l'entropie conditionelle $H_\mu(\xi|\zeta)$ est décroissante en ζ vis-à-vis de l'ordre partiel donné par le raffinement, formalisé dans le lemme suivant :

Lemme 2.2. *Si η est un raffinement de ζ alors pour toute partition ξ on a :*

$$H_\mu(\xi|\eta) \leq H_\mu(\xi|\zeta).$$

Démonstration du lemme.

$$\begin{aligned}
H_\mu(\xi|\eta) &= -\sum_{C\in\xi,\, D\in\eta} \mu(C\cap D)\log\frac{\mu(C\cap D)}{\mu(D)} \\
&= -\sum_{C\in\xi,\, E\in\zeta}\ \sum_{D\in\eta,\, D\subset E} \mu(C\cap D)\log\frac{\mu(C\cap D)}{\mu(D)} \\
&= -\sum_{C\in\xi,\, E\in\zeta} \mu(E) \sum_{D\in\eta,\, D\subset E} \frac{\mu(D)}{\mu(E)}\frac{\mu(C\cap D)}{\mu(D)}\log\frac{\mu(C\cap D)}{\mu(D)} \\
&= -\sum_{C\in\xi,\, E\in\zeta} \mu(E) \sum_{D\in\eta,\, D\subset E} \frac{\mu(D)}{\mu(E)}\psi\left(\frac{\mu(C\cap D)}{\mu(D)}\right).
\end{aligned}$$

On note ensuite que $\sum_{D\in\eta,\, D\subset E}\frac{\mu(D)}{\mu(E)} = 1$ ce qui nous permet d'invoquer la convexité de ψ afin d'utiliser l'inégalité de Jensen :

$$\begin{aligned}
H_\mu(\xi|\eta) &\leq -\sum_{C\in\xi,\, E\in\zeta} \mu(E)\psi\left(\sum_{D\in\eta,\, D\subset E}\frac{\mu(D)}{\mu(E)}\frac{\mu(C\cap D)}{\mu(D)}\right) \\
&= -\sum_{C\in\xi,\, E\in\zeta} \mu(E)\psi\left(\frac{\mu(C\cap E)}{\mu(E)}\right) \\
&= H_\mu(\xi|\zeta).
\end{aligned}$$

□

Ce lemme démontre l'existence de la limite du théorème. En effet, $\bigvee_{i=1}^{n+1} T^{-i}\xi$ est un raffinement de $\bigvee_{i=1}^{n} T^{-i}\xi$ donc la suite dont on veut trouver la limite est décroissante et positive, et donc convergente.

On veut maintenant faire le lien avec l'entropie des ξ_n définis dans le théorème 2.1 du chapitre précédent. Il est donc naturel de s'interroger sur les liens existants entre l'entropie de mesures jointes et l'entropie conditionelle. Le résultat clé pour la démonstration du théorème est donné dans le lemme 2.3.

Lemme 2.3.

$$H_\mu(\xi\vee\zeta|\eta) = H_\mu(\xi|\zeta\vee\eta) + H_\mu(\zeta|\eta).$$

Démonstration du lemme.

$$\begin{aligned}
H_\mu(\xi\vee\zeta|\eta) &= -\sum_{C\in\xi, E\in\zeta, D\in\eta} \mu(C\cap E\cap D)\log\frac{\mu(C\cap E\cap D)}{\mu(D)} \\
&= -\sum_{C\in\xi, E\in\zeta, D\in\eta} \mu(C\cap E\cap D)\left(\log\frac{\mu(C\cap E\cap D)}{\mu(E\cap D)} + \log\frac{\mu(E\cap D)}{\mu(D)}\right) \\
&= -\sum_{C\in\xi, B\in\zeta\vee\eta} \mu(C\cap B)\log\frac{\mu(C\cap B)}{\mu(B)} - \sum_{E\in\zeta, D\in\eta}\left(\sum_{C\in\xi}\mu(C\cap E\cap D)\right)\log\frac{\mu(E\cap D)}{\mu(D)} \\
&= H_\mu(\xi|\zeta\vee\eta) + H_\mu(\zeta|\eta).
\end{aligned}$$

□

Une conséquence facile de ce lemme s'obtient en prenant pour η la partition triviale $\{X\}$, ce qui donne :

Corollaire 2.1.

$$H_\mu(\xi \vee \zeta) = H_\mu(\xi|\zeta) + H_\mu(\zeta).$$

Par récurrence, on peut maintenant démontrer

$$H_\mu(\xi_n) = H_\mu(\xi) + \sum_{k=1}^{n-1} H_\mu \left(\xi | \bigvee_{i=1}^{k} T^{-i}\xi \right)$$

Cette equation est trvialement correcte pour $n = 0$ et $n = 1$, et en la supposant vraie pour n on a :

$$\begin{aligned}
H_\mu(\xi_{n+1}) &= H_\mu \left(\xi \vee \left(\bigvee_{i=1}^{n} T^{-i}\xi \right) \right) \\
&= H_\mu \left(\xi | \bigvee_{i=1}^{n} T^{-i}\xi \right) + H_\mu \left(\bigvee_{i=1}^{n} T^{-i}\xi \right) \\
&= H_\mu \left(\xi | \bigvee_{i=1}^{n} T^{-i}\xi \right) + H_\mu \left(T^{-1}\xi_n \right) \\
&= H_\mu \left(\xi | \bigvee_{i=1}^{n} T^{-i}\xi \right) + H_\mu(\xi) + \sum_{k=1}^{n-1} H_\mu \left(\xi | \bigvee_{i=1}^{k} T^{-i}\xi \right),
\end{aligned}$$

ce qui termine la récurrence. On déduit facilement de cette égalité et du théorème de Césaro que

$$h_\mu(T, \xi) = \lim_{n\to\infty} \frac{1}{n} H_\mu(\xi_n) = \lim_{n\to\infty} H_\mu \left(\xi | \bigvee_{i=1}^{n} T^{-i}\xi \right).$$

□

Remarque 2.4.

1. Le lemme 2.2 s'interprète aisément en terme de l'information encodée dans les partitions : la partition ξ ne peut pas apporter plus d'information à η qu'elle n'en apporte à la partition moins fine ζ.
2. De même, le lemme 2.3 peut se lire de manière intuitive comme suit : l'information que ξ et ζ apportent à η est l'information que ζ apporte à η plus l'information que ξ apporte à ζ et η.

Nous allons aussi utiliser le concept d'entropie conditionelle pour montrer que l'entropie de T est la limite de l'entropie de n'importe quelle suite de partitions engendrant la tribu $\mathcal{A}$ en un sens précis détaillé ci-dessous :

Définition 2.8. Notons $\mathcal{A}(\xi)$ la tribu engendrée par les éléments de ξ et, si $(\zeta^n)_{n\in I}$ est une famille de partitions indexée sur I, nous utiliserons la notation $\bigvee_{n\in I} \mathcal{A}(\zeta^n)$ pour désigner la plus petite tribu contenant les éléments des ζ^n.

Théorème 2.3. *Soit $T : X \to X$ une application mesurable continue préservant une mesure μ sur X et $(\xi^n)_{n\in\mathbb{N}}$ une suite de partitions telle que ξ^{n+1} est un raffinement de ξ^n et $\bigvee_{n\in\mathbb{N}} \mathcal{A}(\xi^n) = \mathcal{A}$. Alors :*

$$h_\mu(T) = \lim_{n\to\infty} h_\mu(\xi^n, T) = \sup_{n\in\mathbb{N}} h_\mu(\xi^n, T).$$

Démonstration. Nous commençons par remarquer que la séquence $(h_\mu(\xi^n, T))_{n\in\mathbb{N}}$ est croissante. En effet, par la Propriété 2.1 on a pour tout m, $n \in \mathbb{N}$

$$H_\mu\left(\bigvee_{i=0}^{n} T^{-i}\xi^m\right) \leq H_\mu\left(\bigvee_{i=0}^{n} T^{-i}\xi^{m+1}\right).$$

On en déduit que la limite de l'énoncé existe.
On montre maintenant que l'on peut borner la différence entre l'entropie de deux partitions en utilisant l'entropie conditionelle dans l'optique de borner l'entropie de n'importe quelle partition par la limite dont on a établi l'existence plus haut.

Lemme 2.4.

$$h_\mu(T, \xi) \leq h_\mu(T, \zeta) + H_\mu(\xi|\zeta)$$

Démonstration du lemme. On commence par utiliser le fait que $\xi_n \vee \zeta_n$ est un raffinement de ξ_n et le corollaire 2.1 pour dire que

$$H_\mu(\xi_n) \leq H_\mu(\xi_n \vee \zeta_n) = H_\mu(\zeta_n) + H_\mu(\xi_n|\zeta_n).$$

On note aussi la conséquence suivante des lemmes 2.2 et 2.3 :

$$H_\mu(\xi \vee \zeta|\eta) \leq H_\mu(\xi|\eta) + H_\mu(\zeta|\eta).$$

En utilisant cette inégalité $n-1$ fois sur le terme $H_\mu(\xi_n|\zeta_n)$ plus haut on obtient

$$\begin{aligned}
H_\mu(\xi_n) &\leq H_\mu(\zeta_n) + \sum_{k=0}^{n-1} H_\mu(T^{-k}\xi|\zeta_n) \\
&\leq H_\mu(\zeta_n) + \sum_{k=0}^{n-1} H_\mu(T^{-k}\xi|T^{-k}\zeta) \quad (\zeta_n \text{ est un raffinement de } T^{-k}\zeta) \\
&\leq H_\mu(\zeta_n) + nH_\mu(\xi|\zeta) \quad (\text{par l'invariance de } \mu \text{ sous } T).
\end{aligned}$$

Le résultat recherché suit naturellement en divisant par n et en faisant tendre n vers l'infini. □

Pour achever la démonstration du théorème, on voit qu'il suffit de montrer que $H_\mu(\xi|\xi^n) \to 0$ quand $n \to \infty$ pour toute partition ξ. En effet, en posant $\xi = \eta$ et $\zeta = \xi^n$ dans l'inégalité du lemme ci-dessus on obtient

$$h_\mu(T, \eta) \leq h_\mu(T, \xi^n) + H_\mu(\eta|\xi^n) \quad \forall \eta.$$

En faisant tendre n vers l'infini dans l'inégalité obtenue grâce au lemme 2.4, on déduit que $\lim_{n\to\infty} h_\mu(\xi^n, T)$ est une borne supérieure pour l'entropie de toute partition, et on conclut que le théorème est vrai. C'est exactement ce que nous donne le lemme 2.6, pour lequel nous auront besoin du lemme suivant :

Lemme 2.5. *Soit* $\eta = \{C_i \ : \ i = 1, \dots, k\}$ *une partition mesurable, et* $\zeta^n = \{D_i^n \ : \ i = 0, \dots, k\}$ *une suite de partitions mesurables telle que* $\forall n \in \mathbb{N}, \ i = 1, \dots, k,$ *on a* $D_i^n \subset C_i$ *et* $\lim\limits_{n \to \infty} \mu(C_i \setminus D_i^n) = 0$. *Alors*

$$\lim_{n \to \infty} H_\mu(\eta | \zeta^n) = 0.$$

Démonstration du lemme. On a évidemment

$$C_i \cap D_i^n = D_i^n \quad \text{et} \quad C_i \cap D_j^n = \emptyset \quad \text{pour } i \neq j, \ i \geq 1, \ j \geq 1.$$

On en déduit

$$\begin{aligned} H_\mu(\eta|\zeta_n) &= -\sum_{E \in \eta, B \in \zeta_n} \mu(E \cap B) \log \frac{\mu(E \cap B)}{\mu(B)} \\ &= -\sum_{i=1}^{k} \mu(D_i^n) \log \frac{\mu(D_i^n)}{\mu(D_i^n)} - \sum_{E \in \eta} \mu(E \cap D_0^n) \log \frac{\mu(E \cap D_0^n)}{\mu(D_0^n)} \\ &= -\sum_{E \in \eta} \mu(E \cap D_0^n) \log \frac{\mu(E \cap D_0^n)}{\mu(D_0^n)} \\ &\leq \sum_{E \in \eta} \mu(E \cap D_0^n) \log \mu(E \cap D_0^n) \to 0 \quad \text{quand} \quad n \to \infty. \end{aligned}$$

□

Lemme 2.6. *Soit* η *une partition de* X

$$\lim_{n \to \infty} H_\mu(\eta | \xi^n) = 0.$$

Démonstration du lemme. Les conditions $\bigvee_{n \in \mathbb{N}} \mathcal{A}(\xi^n) = \mathcal{A}$ et $\xi^n \preccurlyeq \xi^{n+1}$ nous permettent d'affirmer que l'on peut trouver pour chaque C_k une suite d'ensembles $D_k^n \in \mathcal{A}(\xi^n)$ contenues dans C_k telle que $\lim_{n \to \infty} \mu(C_k \setminus D_k^n) = 0$. On pose $D_0^n = X \setminus \bigcup_{i=1}^k D_i^n$ pour définir des partitions ζ^n contenues dans $\mathcal{A}(\xi^n)$. On peut alors invoquer le lemme précédent pour dire

$$H_\mu(\eta | \zeta^n) \to 0 \quad \text{quand} \quad n \to \infty.$$

Le fait que ξ^n soit un raffinement de ζ_n implique $H_\mu(\eta|\xi^n) \leq H_\mu(\eta|\zeta^n)$ par le lemme 2.2, ce qui achève la démonstration du lemme, et donc du théorème. □

□

2.3 Générateurs

Cette section à pour but de montrer que dans certains cas le supremum de l'entropie des partition par rapport à T est atteint sur une partition finie, ce qui permet de calculer l'entropie de T en calculant simplement l'entropie d'une partition par rapport à T. On commence par caractériser une famille de partitions, appelées générateurs.

Définition 2.9. On dit qu'une partition ξ est un générateur (unilatéral) de la tribu $\mathcal{A}$ (par rapport à T) si $\bigvee_{n\in\mathbb{N}} \mathcal{A}(T^{-n}\xi) = \mathcal{A}$. Dans le cas où T est presque partout inversible, on dit que ξ est un générateur bilatéral de $\mathcal{A}$ si $\bigvee_{n\in\mathbb{Z}} \mathcal{A}(T^{-n}\xi) = \mathcal{A}$.

Théorème 2.4 (Kolmogorov-Sinai)**.** *Soit $T : X \to X$ une application continue mesurable pour une tribu $\mathcal{A}$ et préservant une mesure μ sur X. Alors :*

1. *Si ξ est un générateur unilatéral, $h_\mu(T) = h_\mu(T,\xi)$.*
2. *Si T est inversible presque partout et que ξ est un générateur bilatéral, $h_\mu(T) = h_\mu(T,\xi)$.*

Démonstration. On commence par la première affirmation en prenant : ξ un générateur unilatéral. On suit ensuite de très près la démonstration du théorème 2.3. On trouve une majoration de l'entropie de n'importe quelle partition η grâce au lemme 2.4 :

$$h_\mu(T,\eta) \leq h_\mu(T,\xi_n) + H_\mu(\eta|\xi_n).$$

Par ailleurs, on sait calculer l'entropie de ξ_n

$$\begin{aligned} h_\mu(T,\xi_n) &= \lim_{m\to\infty} \frac{1}{m} H_\mu\Big(\bigvee_{i=0}^{m-1} T^{-i}\xi_n\Big) = \lim_{m\to\infty} \frac{1}{m} H_\mu(\xi_{m+n}) \\ &= h_\mu(T,\xi). \end{aligned}$$

On utilise le lemme 2.6 et le fait que

$$\bigvee_{n\in\mathbb{N}} \mathcal{A}(T^{-n}\xi) = \bigvee_{n\in\mathbb{N}} \mathcal{A}(\xi_n) = \mathcal{A}$$

pour obtenir

$$h_\mu(T,\eta) \leq h_\mu(T,\xi) + \lim_{n\to\infty} H_\mu(\eta|\xi_n) = h_\mu(T,\xi)$$

et ainsi conclure :

$$h_\mu(T) = \sup_\eta h_\mu(T,\eta) \leq h_\mu(T,\xi) \leq h_\mu(T).$$

Supposons à présent que ξ soit un générateur bilatéral et que T soit inversible presque partout pour démontrer la deuxième affirmation du théorème. On adopte la même stratégie que plus haut, mais on pose

$$h_\mu(T,\eta) \leq h_\mu\left(T, \bigvee_{i=-n}^{n} T^{-i}\xi\right) + H_\mu\left(\eta| \bigvee_{i=-n}^{n} T^{-i}\xi\right)$$

comme inégalité de départ. Par l'invariance de μ sous T et ce qui à été fait dans la première partie de la démonstration, on a :

$$h_\mu\left(T, \bigvee_{i=-n}^{n} T^{-i}\xi\right) = h_\mu\left(T, T^{-n} \bigvee_{i=-n}^{n} T^{-i}\xi\right) = h_\mu(T,\xi_{2n+1}) = h_\mu(T,\xi).$$

On a encore

$$\mathcal{A} = \bigvee_{n\in\mathbb{N}} \mathcal{A}(T^{-n}\xi) = \bigvee_{n\in\mathbb{N}} \mathcal{A}\left(\bigvee_{i=-n}^{n} T^{-i}\xi\right)$$

et donc
$$\lim_{n\to\infty} H_\mu\left(\eta|\bigvee_{i=-n}^{n} T^{-i}\xi\right) = 0$$
ce qui implique l'inégalité $h_\mu(T,\eta) \le h_\mu(T,\xi)$ dont découle la fin du théorème. $\square$

2.4 Exemples

Exemple 2.1. *Soit* (Σ,σ) *le décalage et* δ_{0^∞} *la mesure de Dirac. Alors* $h_{\delta_{0^\infty}}(\sigma) = 0$.

Démonstration. On prend la partition ξ. Comme ξ_n est une partition mesurable $\sum_{A\in\xi_n}\delta_{0^\infty}(A) = 1$ d'où l'on déduit qu'il y a seulement un A de ξ_n tel que $\delta_{0^\infty}(A) = 1$. Tous les autres éléments $B \in \xi_n$ sont de mesure 0. Alors
$$H_{\delta_{0^\infty}}(\xi_n) = \psi(1) + \sum_{C\in\xi_n,\ 0^\infty\notin C} \psi(0) = 0.$$
Donc $h_{\delta_{0^\infty}} = 0$. $\square$

Exemple 2.2. *Soit* (Σ,σ) *le décalage et soit* λ *la mesure engendrée par* p_0 *et* p_1*. Alors* $h_\lambda(\sigma) = -p_0\log p_0 - p_1\log p_1$.

Démonstration. La partition mesurable $\xi = \{C[0], C[1]\}$ est un generateur. En effet,
$$\bigvee_{j=0}^{n}\sigma^{-i}\xi = \{C[w]\ :\ |w| = n+1\}$$
et les cylindres engendrent la tribu borélienne. Alors, d'après le Théorème 2.4 on obtient $h_\lambda(\sigma) = h_\lambda(\sigma,\xi)$, mais

$$\begin{aligned}
h_\lambda(\sigma,\xi) &= \lim_{n\to\infty}\frac{1}{n}H_\lambda\left(\bigvee_{j=0}^{n-1}\sigma^{-j}\xi\right)\\
&= \lim_{n\to\infty} -\frac{1}{n}\sum_{w:|w|=n}\lambda(C[w])\log\lambda(C[w])\\
&= \lim_{n\to\infty} -\frac{1}{n}\sum_{(i_0,\dots,i_{n-1})\in\{0,1\}^n}\lambda(C[i_0 i_1\dots i_{n-1}])\log\lambda(C[i_0 i_1\dots i_{n-1}])\\
&= \lim_{n\to\infty} -\frac{1}{n}\sum_{(i_0,\dots,i_{n-1})\in\{0,1\}^n} p_{i_0}\dots p_{i_{n-1}}(\log p_{i_0} + \dots\log p_{i_{n-1}})\\
&= \lim_{n\to\infty} -\frac{1}{n}\left(\sum_{(i_0,\dots,i_{n-1})\in\{0,1\}^n} p_{i_0}\dots p_{i_{n-1}}\log p_{i_0} + \dots + \sum_{(i_0,\dots,i_{n-1})\in\{0,1\}^n} p_{i_0}\dots p_{i_{n-1}}\log p_{i_{n-1}}\right)\\
&= \lim_{n\to\infty} -\frac{1}{n}n\sum_{(i_0,\dots,i_{n-1})\in\{0,1\}^n} p_{i_0}\dots p_{i_{n-1}}\log p_{i_0}\\
&= -\lim_{n\to\infty}\left(p_0\log p_0\sum_{(i_1,\dots,i_{n-1})\in\{0,1\}^{n-1}} p_{i_1}\dots p_{i_{n-1}} + p_1\log p_1\sum_{(i_1,\dots,i_{n-1})\in\{0,1\}^{n-1}} p_{i_1}\dots p_{i_{n-1}}\right)\\
&= -p_0\log p_0 - p_1\log p_1,
\end{aligned}$$

parce que

$$1 = (p_0 + p_1)^{n-1} = \sum_{(i_1,\ldots,i_{n-1})\in\{0,1\}^{n-1}} p_{i_1} \ldots p_{i_{n-1}}.$$

□

3 Entropie topologique

L'entropie topologique est un autre invariant topologique introduit en 1965 par Adler, Konheim et McAndrew [1], puis avec une définition alternative pour les espaces métriques par Bowen et Dinaburg. L'objet de ce chapitre est de montrer l'équivalence de ces 2 définitions dans les espaces métriques, et on exposera dans le chapitre suivant le lien avec l'entropie définit précédemment sur un espace mesurable.

3.1 L'entropie topologique par les recouvrements ouverts

Définition 3.1 (Entropie d'un recouvrement ouvert)**.** Soit α un recouvrement ouvert d'un espace topologique compact X. On note $N(\alpha)$ la cardinalité minimale d'un sous-recouvrement fini de α, et on définit l'entropie $H(\alpha)$ de α par

$$H(\alpha) = \log(N(\alpha)).$$

De manière analogue à ce qui a été fait pour les partitions on introduit les concepts de raffinement et de recouvrement joint pour des recouvrements α, β :

Définition 3.2.

1. Le recouvrement joint $\alpha \vee \beta$ est le recouvrement formé par les parties $A \cap B$ où $A \in \alpha$ et $B \in \beta$
2. On dit que α est un raffinement de β, noté $\beta < \alpha$, si tout élément de α est une partie d'un élément de β
3. pour $T : X \to X$ continue, on définit le recouvrement ouvert $T^{-1}\alpha = \{T^{-1}A : A \in \alpha\}$

Les propriétés suivantes découlent directement de ces définitions :

Propriété 3.1.

1. $H(\alpha) \geq 0$ *et* $H(\alpha) = 0 \iff X \in \alpha$
2. $\alpha < \beta \implies H(\alpha) \leq H(\beta)$
3. $H(\alpha \vee \beta) \leq H(\alpha) + H(\beta)$
4. $T^{-1}(\alpha \vee \beta) = T^{-1}\alpha \vee T^{-1}\beta$ *et* $\alpha < \beta \implies T^{-1}\alpha \leq T^{-1}\beta$
5. $H(T^{-1}\alpha) \leq H(\alpha)$ *avec égalité si* T *est surjective*

Définition 3.3. Soit α un recouvrement ouvert d'un espace topologique compact X et $T : X \to X$ une application continue, alors on définit l'entropie topologique de T par rapport à α par la limite ci-dessous dont l'existence est donnée dans le Théorème 3.1 :

$$h(T, \alpha) = \lim_{n\to\infty} \frac{1}{n} H\left(\bigvee_{j=0}^{n-1} T^{-j}\alpha\right)$$

Théorème 3.1. *L'entropie d'un recouvrement ouvert par rapport à une application continue est bien définie.*

Démonstration. Par le Lemme 2.1 il suffit de montrer que la suite

$$a_n = H\left(\bigvee_{j=0}^{n-1} T^{-j}\alpha\right)$$

est sous-additive. Or :

$$\begin{aligned}
a_{n+k} &= H\left(\bigvee_{j=0}^{n+k-1} T^{-j}\alpha\right) \\
&= H\left(\left(\bigvee_{j=0}^{n-1} T^{-j}\alpha\right) \vee T^{-n}\left(\bigvee_{j=0}^{k-1} T^{-j}\alpha\right)\right) \text{ par le point 4 de Propriété 3.1} \\
&\leq H\left(\bigvee_{j=0}^{n-1} T^{-j}\alpha\right) + H\left(T^{-n}\left(\bigvee_{j=0}^{k-1} T^{-j}\alpha\right)\right) \text{ par le point 3 de Propriété 3.1} \\
&\leq H\left(\bigvee_{j=0}^{n-1} T^{-j}\alpha\right) + H\left(\bigvee_{j=0}^{k-1} T^{-j}\alpha\right) = a_n + a_k \text{ par le point 5 de Propriété 3.1.}
\end{aligned}$$

□

Remarque 3.1.

1. Une conséquence de 1. dans la Propriété 3.1 est que $h(T, \alpha) \geq 0$.
2. Par 2. dans Propriété 3.1 on obtient directement $\alpha < \beta \implies h(T, \alpha) \leq h(T, \beta)$.
3. Les points 3 et 5 nous donnent quant à eux

$$H\left(\bigvee_{j=0}^{n-1} T^{-j}\alpha\right) \leq nH(\alpha)$$

ce qui nous permet de conclure $H\left(\bigvee_{j=0}^{n-1} T^{-j}\alpha\right) \leq H(\alpha)$.

Définition 3.4. On définit par $h(T) = \sup_\alpha h(T, \alpha)$ l'entropie topologique de T.

3.2 La définition de Bowen de l'entropie

Nous exposons dans cette partie une définition proposée par Dianburg et Bowen de l'entropie d'une application uniformément continue (notée T dans le reste du chapitre) d'un espace métrique compact (X, d) dans lui-même. Nous utiliserons par la suite la métrique d_n dépendant de d et, implicitement, de T et définie par

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

Définition 3.5. Soit n un entier, $\epsilon > 0$. Une partie $F \subset X$ est (n, ϵ)-couvrante pour X par rapport à T si

$$\forall x \in X \, \exists y \in F \quad \text{tel que} \quad d_n(x, y) < \epsilon$$

Définition 3.6. Notons $r_n(\epsilon)$ la cardinalité minimale de tout ensemble (n, ϵ)-couvrant de X par rapport à T.

Définition 3.7. Notons $r(\epsilon, T) = \limsup_{n \to \infty} (1/n) \log(r_n(\epsilon))$

Définition 3.8. On définit l'entropie topologique $h(T)$ de T par

$$h(T) = \lim_{\epsilon \to 0} r(\epsilon, T).$$

Cette limite existe bien, car $r(\epsilon, T)$ est clairement décroissante en ϵ

Remarque 3.2. Dans le cas plus général où X n'est pas forcément compact, on peut étendre notre définition de l'entropie en définissant l'entropie $h(K, T)$ de T sur un compact $K \subset X$ comme précédement, puis en posant :

$$h(T) = \sup_K h(K, T)$$

où le supremum est sur tous les compacts de X.

Définition 3.9. Un ensemble $E \subset X$ est (n, ϵ)-séparé si $\forall x, y \in E$, $d_n(x, y) > \epsilon$. La cardinalité maximale d'un ensemble (n, ϵ)-séparé est notée $s_n(\epsilon)$, et de manière analogue à ce qui a été fait précédemment, nous posons $s(\epsilon, T) = \limsup_{n \to \infty} (1/n) \log s_n(\epsilon)$

Théorème 3.2. $h(T) = \lim_{\epsilon \to 0} s(\epsilon, T)$

Démonstration. On montre d'abord que $r_n(\epsilon) \leq s_n(\epsilon) \leq r_n(\epsilon/2)$. En effet, un ensemble (n, ϵ)-séparé de cardinalité maximale est clairement (n, ϵ)-couvrant. Prenons maintenant des ensembles E, $F \subset X$ respectivement (n, ϵ)-séparé de cardinalité maximale, et (n, ϵ)-couvrant de cardinalité minimale. $\forall x \in E$ notons $\phi(x)$ l'élément de F tel que $d_n(x, \phi(x)) < \epsilon/2$. Par l'inégalité du triangle, ϕ est injective ce qui nous donne la deuxième inégalité. On déduit de ce qui précède que $r(\epsilon, T) \leq s(\epsilon, T) \leq r(\epsilon/2, T)$ et en faisant tendre ϵ vers 0 on obtient $h(T) = \lim_{\epsilon \to 0} s(\epsilon, T)$, ce qui achève la démonstration □

Notre but est maintenant de démontrer que les deux définitions de l'entropie topologique coïncident. Pour éviter toute confusion nous noterons provisoirement $h^*(T, \alpha)$ l'entropie topologique d'un recouvrement ouvert α et $h^*(T) = \sup_\alpha h^*(T, \alpha)$ l'entropie de T telle qu'elle est définie dans la définition 3.4.

Théorème 3.3. *Soit (X, d) un espace métrique compact, et $\{\alpha_n\}_{n\in\mathbb{N}}$ une suite de recouvrements ouverts de X telle que* $\operatorname{diam}(\alpha_n) = \sup_{A\in\alpha_n} \operatorname{diam}(A) \underset{n\to\infty}{\to} 0$. *Alors on a*

$$h^*(T) = \lim_{n\to\infty} h^*(T, \alpha_n).$$

Démonstration. Supposons que $h^*(T) < \infty$, et soit $\epsilon > 0$. Il existe un recouvrement ouvert γ tel que $h^*(\gamma) > h^*(T) - \epsilon$. Si δ est un nombre de Lebesgue pour γ, il existe $N \in \mathbb{N}$ tel que pour tout $n \geq N$, $\operatorname{diam}(\alpha_n) < \delta$ et donc $\gamma < \alpha_n$. Le point 2 de la propriété 3.1 nous permet de dire que $h^*(T, \alpha_n) \geq h^*(\gamma)$, ce qui implique que $h^*(T) \geq h^*(T, \alpha_n) > h^*(T) - \epsilon$, ce qu'il nous fallait pour conclure. Dans le cas où $h^*(T) = \infty$ on peut choisir γ telle que $h^*(T, \gamma)$ soit arbitrairement grand pour montrer comme ci-dessus que $\lim_{n\to\infty} h^*(T, \alpha_n) = \infty$ $\square$

Théorème 3.4. *Soit $T : X \to X$ une transformation continue de l'espace métrique compact (X, d). Alors*

$$h(T) = h^*(T).$$

Démonstration. Soit α un recouvrement avec un nombre de Lebesgue δ, et soit F un ensemble $(n, \delta/2)$-couvrant de cardinalité minimale. Alors

$$X = \bigcup_{x\in F} \bigcap_{i=0}^{n-1} T^{-i}\overline{B}(T^i x, \delta/2).$$

Comme les $\overline{B}(T^i x, \delta/2)$ sont contenus dans des éléments de α, les ensembles $\bigcap_{i=0}^{n-1} T^{-i}\overline{B}(T^i x, \delta/2)$ sont contenus dans des éléments de $\bigvee_{i=0}^{n-1} T^{-i}\alpha$ et donc

$$N\left(\bigvee_{i=0}^{n-1} T^{-i}\alpha\right) \leq r_n(\delta/2, T) \leq s_n(\delta/2, T).$$

Supposons maintenant que $\operatorname{diam} \alpha \leq \epsilon$, et considérons un ensemble F (n, ϵ)-séparé de taille $s_n(\epsilon, T)$. Chaque élément de $\bigvee_{i=0}^{n-1} T^{-i}\alpha$ a un diamètre borné par ϵ dans la distance d_n et ne peut donc contenir au plus qu'un seul élément de F ce qui montre

$$r_n(\epsilon, T) \leq s_n(\epsilon, T) \leq N\left(\bigvee_{i=0}^{n-1} T^{-i}\alpha\right).$$

En notant α_ϵ le recouvrement composé de toutes les boules ouvertes de rayon 2ϵ et γ_ϵ n'importe quel recouvrement de diamètre $\epsilon/2$ ce qui précède nous donne l'inégalité suivante :

$$N\left(\bigvee_{i=0}^{n-1} T^{-i}\alpha_\epsilon\right) \leq r_n(\epsilon, T) \leq s_n(\epsilon, T) \leq N\left(\bigvee_{i=0}^{n-1} T^{-i}\gamma_\epsilon\right)$$

et on a donc, faisant tendre n vers l'infini :

$$h^*(T,\alpha_\epsilon) \leq r(\epsilon,X,T) \leq s(\epsilon,X,T) \leq h^*(T,\gamma_\epsilon)$$

puis, d'après théorème 3.3, quand on fait tendre ϵ à 0 on obtient

$$h^*(T) \leq h(T) \leq h^*(T)$$

□

3.3 Propriétés

On a une propriété similaire à la Propriété 2.4.

Lemme 3.1. *$h(T^k) = kh(T)$, pour tout $k \in \mathbb{N}^*$.*

Démonstration. Clairement,

$$\max_{0\leq i\leq n-1} d(T^{ik}x, T^{ik}y) \leq \max_{0\leq i\leq kn-1} d(T^i x, T^i y),$$

alors $s_{nk}(\epsilon,T) \geq s_n(\epsilon,T^k)$, d'où

$$h(T^k) = \sup_{\epsilon>0} \limsup_{n\to\infty} \frac{1}{n} s_n(\epsilon,T^k) \leq \sup_{\epsilon>0} \limsup_{n\to\infty} k\frac{1}{nk} s_{nk}(\epsilon,T) \leq kh(T).$$

D'un autre côté, par la continuité uniforme de T^i pour $\epsilon > 0$ il existe $\delta(\epsilon)$ tel que si $d(x,y) < \delta(\epsilon)$ alors $d(T^i x, T^i y) < \delta(\epsilon)$ pour tout $0 \leq i \leq k-1$. On écrit $m = kq - r$, où q et r sont entiers et $0 \leq r \leq k-1$. Donc, si $\max\limits_{0\leq i\leq q-1} d(T^{ik}x, T^{ik}y) \leq \delta(\epsilon)$, on a

$$\max_{0\leq i\leq m-1} d(T^i x, T^i y) \leq \max_{0\leq i\leq kq-1} d(T^i x, T^i y) < \epsilon,$$

d'où $s_q(\delta(\epsilon),T^k) \geq s_m(\epsilon,T)$. Alors,

$$h(T^k) = \sup_{\epsilon>0} \limsup_{q\to\infty} \frac{1}{q} s_q(\epsilon,T^k) = \sup_{\epsilon>0} \limsup_{n\to\infty} \frac{1}{q} s_q(\delta(\epsilon),T^k) \geq \sup_{\epsilon>0} \limsup_{m\to\infty} \frac{m}{q}\cdot\frac{1}{m} s_m(\epsilon,T) = kh(T),$$

parce que $\lim_{m\to\infty} \frac{m}{q} = \lim_{q\to\infty} \frac{kq-r}{q} = k$. □

Définition 3.10. Soient (X,T) et (Y,S) deux systèmes dynamiques. On dit que (X,T) est un facteur de (Y,S) si il existe une surjection continue π telle que $\pi \circ T = S \circ \pi$. Si π est un homomorphisme, on dit que (X,T) et (Y,S) sont conjugués.

Lemme 3.2. *Si (X,T) est un facteur de (Y,S) avec X,Y des compacts, alors $h(T) \leq h(S)$.*

Démonstration. Comme $\pi : Y \to X$ est uniformément continue, pour tout $\epsilon > 0$ il existe $\delta(\epsilon) > 0$ tel que si $d_Y(x,y) < \delta(\epsilon)$ on a $d_X(\pi(x),\pi(y)) < \epsilon$. Alors, si $d_Y(S^i x, S^i y) < \delta(\epsilon)$,

$$d_X(T^i\pi(x), T^i\pi(x)) = d_X(\pi(S^i x), \pi(S^i y)) < \epsilon$$

donc $d_{n,Y}(x,y) < \delta(\epsilon)$ implique $d_{n,X}(x,y) < \epsilon$.

Soit $A(n,\epsilon,X)$ un ensemble (n,ϵ)-separé de cardinalité maximale de (X,T). Comme π est une surjection, on peut prendre $B \subset Y$ tel que $|B| = |A(n,\epsilon,X)|$ et $\pi(B) = A(n,\epsilon,X)$. En fait, B est $(n,\delta(\epsilon))$-separé : en effet, pour $x,y \in B$, $x \neq y$, si $d_{n,Y}(x,y) < \delta(\epsilon)$, alors $d_{n,X}(\pi(x),\pi(y)) < \epsilon$, une contradiction. Donc $s_n(\epsilon,T,X) \leq s_n(\delta(\epsilon),S,Y)$ et on conclut

$$h(T) = \sup_{\epsilon>0} \limsup_{n\to\infty} \frac{1}{n} s_n(\epsilon,T,X) \leq \sup_{\epsilon>0} \limsup_{n\to\infty} \frac{1}{n} s_n(\delta(\epsilon),S,Y) \leq h(S)$$

□

On obtient une propriété importante : l'entropie topologique est un *invariant topologique.*

Corollaire 3.1. *Si (X,T) et (Y,S) sont conjugés avec X,Y compacts, alors $h(T) = h(S)$.*

Démonstration. Dans ce cas, (X,T) est un facteur de (Y,S) et vice versa. Alors $h(T) \leq h(S) \leq h(T)$, d'où $h(T) = h(S)$. □

3.4 Exemples

On montre ici des exemples de calculs d'entropie topologique.

L'exemple suivant montre que les applications "simples" ont une petite entropie.

Exemple 3.1. *Soit X un espace métrique compact et* $\mathrm{Id} : X \to X$ *l'identité. Alors,* $h(\mathrm{Id}) = 0$.

Démonstration. On voit que $d_n(x,y) = \max_{0\leq i\leq n-1} d(\mathrm{Id}^i x, \mathrm{Id}^i y) = d(x,y)$. Alors $s(n,\epsilon)$ ne dépend pas de n. Donc

$$h(\mathrm{Id}) = \lim_{\epsilon\to 0} \limsup_{n\to\infty} \frac{1}{n} \log s(n,\epsilon) = \lim_{\epsilon\to 0} 0 = 0.$$

□

En revanche, les systèmes plus compliqués comme le décalage sont d'entropie strictement positive.

Exemple 3.2. *Soit (Σ,σ) le décalage décrit dans l'introduction. Alors $h(\sigma) = \log 2$.*

Démonstration. On prend $\epsilon < 1$ fixé. Soit $k \in \mathbb{N}$ tel que $1/k \geq \epsilon > 1/(k+1)$ et $n \in \mathbb{N}$. On démontre que $s(n,\epsilon) = 2^{n+k}$. Nous supposons qu'il existe un ensemble (n,ϵ)-séparé S tel que $|S| > 2^{n+k}$. Alors il existe deux points $x \neq y$ dans S tels que $x_{[0,n+k-1]} = y_{[0,n+k-1]}$. Donc pour $0 \leq i \leq n-1$, $d(\sigma^i x, \sigma^i y) \leq 1/(n+k-i) \leq 1/(k+1) < \epsilon$. On conclut que $d_n(x,y) < \epsilon$, mais ç'est une contradiction avec la condition que S est (n,ϵ)-separé.

Maintenant, on pose $S = \{a_0 a_1 \dots a_{n+k-1} 0^\infty \ : \ a_i \in \{0,1\}$ pour $0 \leq i \leq n+k-1\}$. Si $x \neq y \in S$ il existe $0 \leq i \leq n+k-1$ minimal tel que $x_i \neq y_i$. Alors si $i \leq n-1$ $d(\sigma^i x, \sigma^i y) = 1 > \epsilon$. Sinon, $d(\sigma^{n-1}x, \sigma^{n-1}y) = 1/(i-(n-1)) \geq 1/(n+k-1-(n-1)) = 1/k \geq \epsilon$. Donc $d_n(x,y) \geq \epsilon$ et S est (n,ϵ)-separé. Clairement, $|S| = 2^{n+k}$.

Finalement,

$$h(\sigma) = \lim_{\epsilon\to 0} \limsup_{n\to\infty} \frac{1}{n} \log s(n,\epsilon) = \lim_{\epsilon\to 0} \limsup_{n\to\infty} \frac{n+k}{n} \log 2 = \lim_{\epsilon\to 0} \log 2 = \log 2.$$

□

4 Principe variationnel

Dans cette section, on suppose que la tribu $\mathcal{A}$ sur X est borélienne.

On commence par revenir brièvement au système du décalage. Dans les exemples des chapitres précédents, on a vu que l'entropie métrique du décalage n'est pas plus grande que l'entropie topologique. Effectivement, dans l'Exemple 2.1 $h_{\delta_0\infty}(\sigma) = 0 < \log 2 = h(\sigma)$; si λ est la mesure definie pour p_0, p_1 alors d'après l'Exemple 2.2, $h_\lambda(\sigma) = -p_0 \log p_0 - p_1 \log p_1$. En utilisant 2.1, on obtient

$$\begin{aligned} h_\lambda(\sigma) &= -p_0 \log p_0 - p_1 \log p_1 \\ &\leq -2\big(\frac{1}{2} \log \frac{p_0 + p_1}{2}\big) = -\log(1/2) = \log 2 \\ &= h(\sigma), \end{aligned}$$

donc l'entropie metrique de mesures de cette famille n'est jamais supérieure à l'entropie topologique et on a égalité pour $p_0 = p_1 = 1/2$.

Ce qui précède illustre le théorème principal suivant, qui est l'objectif de ce mémoire.

Théorème 4.1 (Principe variationnel)**.** *Soit X un espace métrique compact, et $T : X \to X$ une application continue. Alors*

$$h(T) = \sup\{h_\mu(T) : \mu \ \ \textit{est une mesure de probabilité } T\textit{-invariante}\}.$$

Naturellement, la preuve consiste en deux parties : d'abord nous montrons que pour toute mesure μ T-invariante, $h(T) \geq h_\mu(T)$; puis nous montrons qu'il est possible d'approcher $h(T)$ par $h_{\mu_n}(T)$ où μ_n sont des mesures appropriées.

Lemme 4.1. *Soit μ une mesure T-invariante. Alors $h(T) \geq h_\mu(T)$.*

Démonstration. Soit $\eta = \{C_1, \ldots, C_k\}$ une partition mesurable et $\delta > 0$. D'après les remarques 2.1 et 2.2, on peut supposer les C_i deux à deux disjoints. Par la régularité des mesures boréliennes, il existe des ensembles compacts $D_i \subset C_i$ pour $i \in \{1, \ldots, k\}$ tels que $\mu(C_i \backslash D_i) < \delta$. En effet, soit $C \subset X$ mesurable, donc $\mu(X \setminus C) = \mu^*(X \setminus C)$, donc il existe O ouvert tel que $X \setminus C \subset O$ et $\mu(O) - \mu(X \setminus C) < \delta$. Alors $X \setminus O$ est compact, $X \setminus O \subset C$ et $\mu(C) - \mu(X \setminus O) < \delta$.

Maintenent, on considère $\beta = \{D_0, D_1, \ldots, D_k\}$ où $D_0 = X \setminus \bigcup_{i=1}^k D_i$. Clairement, β est une partition mesurable. En choisissant δ assez petit, on peut supposer $H_\mu(\eta|\beta) < 1$ d'après le lemme 2.5. Par le lemme 2.4

$$h_\mu(T^m, \eta) \leq h_\mu(T^m, \beta) + H_\mu(\eta|\beta) < h_\mu(T^m, \beta) + 1,$$

pour tous m entier. Comme les $D_1, \ldots, D_n$ sont disjoints, $\mathcal{U} = \{D_0 \cup D_1, \ldots, D_0 \cup D_n\}$ est un recouvrement ouvert. On considère aussi

$$\mathcal{U}_{mn} = \{\bigcap_{i=0}^{n-1} T^{-im} U_i : \ U_0, U_1, \ldots U_{n-1} \in \mathcal{U}\}$$

et

$$\beta_{mn} = \bigvee_{i=0}^{n-1} T^{-im}\beta.$$

On a

$$\text{card } \ \beta_{mn} \leq 2^n \text{ card } \ \mathcal{U}_{mn}.$$

En effet, chaque élement de $\mathcal{U}_{mn}$ est de la forme

$$(D_0 \cup D_{i_0}) \cap T^{-m}(D_0 \cup D_{i_1}) \cdots \cap T^{-m(n-1)}(D_0 \cup D_{i_{n-1}}).$$

Donc si cet élément n'est pas vide il y a au plus 2^n élements non-vides de β_{mn} de forme $B_0 \cap T^{-m}B_1 \cap \cdots \cap T^{-m(n-1)}B_{n-1}$, où $B_j = D_0$ ou D_{i_j}. Inversement, si $B_0 \cap T^{-m}B_1 \cap \cdots \cap T^{-m(n-1)}B_{n-1} \in \beta_{mn}$ n'est pas vide il est facile de construire
$U_0 \cap T^{-m}U_1 \cap \cdots \cap T^{-m(n-1)}U_{n-1} \in \mathcal{U}_{mn}$ non-vide : si $B_j \neq D_0$ on prend $U_j = B_j \cup D_0$, si $B_j = D_0$ on prend U_j arbitrairement dans $\mathcal{U}$.

D'après la Remarque 2.3 nous obtenons

$$H_\mu(\beta_{mn}) \leq \log \text{card } \ \beta_{mn} \leq n \log 2 + \log \text{card } \ \mathcal{U}_{mn}.$$

Alors

$$\begin{aligned} h_\mu(T^m, \eta) &\leq h_\mu(T^m, \beta) + 1 = \lim_{n\to+\infty} \frac{1}{n} H_\mu(\beta_{mn}) + 1 \\ &\leq \limsup_{n\to+\infty} \frac{1}{n} \log \text{card } \ \mathcal{U}_{mn} + \log 2 + 1 \\ &= h(T^m, \mathcal{U}) + \log 2 + 1. \end{aligned}$$

Ainsi

$$h_\mu(T^m, \eta) \leq h(T^m) + \log 2 + 1,$$

et

$$h_\mu(T^m) \leq h(T^m) + \log 2 + 1.$$

Finalement, comme d'après Propriété 2.4 et le lemme 3.1 $h_\mu(T^m) = mh_\mu(T)$ et $h(T^m) = mh(T)$, on a

$$h_\mu(T) \leq h(T) + \frac{1}{m}(\log 2 + 1).$$

Alors

$$h_\mu(T) \leq h(T).$$

□

Avant la second partie de la preuve on a besoin de définir l'espace métrique des mesures de probabilité sur X.

Soit $\mathcal{M}(X)$ l'ensemble des mesures de probabilité de X définies sur la tribu borélienne. Comme X est compact et métrique, l'espace $\mathcal{C}(X)$ est separé. Donc il existe une suite de fonctions $(\phi_n)_{n\in\mathbb{N}}$ dense dans $\{\phi \in \mathcal{C}(X) : ||\phi||_\infty \leq 1\}$. Pour un couple de mesures μ, ν on définit

$$d(\mu,\nu) = \sum_{n=0}^{\infty} \frac{1}{2^n} \left| \int_X \phi_n \, d\mu - \int_X \phi_n \, d\nu \right|.$$

d est bien définie et vérifie les axiomes d'une distance. On a donc muni $\mathcal{M}(X)$ d'une structure d'espace métrique. Cet espace se caractérise par la propriété suivant :

Propriété 4.1. *Soient $(\mu_n)_{n\in\mathbb{N}}$ un suite de mesures de $\mathcal{M}(X)$ et $\mu \in \mathcal{M}(X)$. Alors $\mu_n \to \mu$, $n \to \infty$ ssi*

$$\forall \phi \in \mathcal{C}(X), \quad \lim_{n\to\infty} \int_X \phi \, d\mu_n = \int_X \phi \, d\mu.$$

De plus :

Théorème 4.2. *$\mathcal{M}(X)$ est compact.*

Avant de continuer nous avons besoin du concept et du lemme suivant.

Définition 4.1. Le frontière ∂A d'une partie A est définie par $\partial A = \overline{A} \cap \overline{X \setminus A}$.

Lemme 4.2. *Etant donné une suite de mesures $\mu_n \to \mu$, $n \to \infty$ où l'on sait que μ est T-invariante. Soit ξ un partition mesurable de X tel que pour chaque $A \in \xi$ $\mu(\partial A) = 0$. Alors*

1. *pour tout $A \in \xi_m$ $\mu(\partial A) = 0$;*
2. *pour tout $A \in \xi_m$ $\mu_n(A) \to \mu(A), n \to \infty$.*

Démonstration.

1. Chaque $A \in \xi$ s'écrit comme $A = A_0 \cap \cdots \cap T^{-m+1}A_{m-1}$ où $A_0, \ldots, A_{m-1} \in \xi$.

$$\begin{aligned}
\overline{A} \cap \overline{X \setminus A} &= \overline{A_0 \cap \cdots \cap T^{-m+1}A_{m-1}} \cap \overline{X \setminus (A_0 \cap \cdots \cap T^{-m+1}A_{m-1})} \\
&\subset \overline{A_0} \cap \cdots \cap \overline{T^{-m+1}A_{m-1}} \cap (\overline{X \setminus A_0} \cup \cdots \cup \overline{X \setminus T^{-m+1}A_{m-1}}) \\
&\subset \overline{A_0} \cap \cdots \cap T^{-m+1}\overline{A_{m-1}} \cap (\overline{X \setminus A_0} \cup \cdots \cup T^{-m+1}\overline{X \setminus A_{m-1}}) \\
&\subset \bigcup_{i=0}^{m-1} T^{-i}(\overline{A_i} \cap \overline{X \setminus A_i}).
\end{aligned}$$

Donc

$$\mu(\overline{A} \cap \overline{X \setminus A}) \leq \sum_{i=0}^{m-1} \mu(T^{-i}(\overline{A_i} \cap \overline{X \setminus A_i})) = \sum_{i=0}^{m-1} \mu(\overline{A_i} \cap \overline{X \setminus A_i}) = 0.$$

2. Soient $\phi_k : X \to \mathbb{R}_+$ une suite décroissante de fonctions convergant à $\chi_{\overline{A}}$ (par exemple, $\phi_k(x) = exp(-kd(x, \overline{A}))$).

Alors d'après le lemme de Fatou

$$\limsup_{n\to\infty} \mu_n(\overline{A}) \leq \limsup_{n\to\infty} \int_X \phi_k \, d\mu_n = \int_X \phi_k \, d\mu \to \mu(\overline{A}),$$

quand $k \to \infty$.

Comme $\mu(\partial A) = 0$, $\mu(A) = \mu(\overline{A})$, donc

$$\limsup_{n\to\infty} \mu_n(A) \leq \mu(\overline{A}) = \mu(A).$$

De manière similaire, comme $\partial A = \partial(X \setminus A)$,

$$\limsup_{n\to\infty} \mu_n(X \setminus A) \leq \mu(X \setminus A),$$

ce qui donne

$$\liminf_{n\to\infty} \mu_n(A) \geq \mu(A).$$

Les deux inégalités impliquent finalement

$$\lim_{n\to\infty} \mu_n(A) = \mu(A).$$

□

Lemme 4.3. $h(T) \leq \sup\{h_\mu(T) : \mu$ *est une mesure de probabilité* T*-invariante*$\}$.

Démonstration. On fixe $\epsilon > 0$. Soit E_n un ensemble (n,ϵ)-separé maximal. Nous definissons les mesures

$$\nu_n = \frac{1}{\text{card } E_n} \sum_{x \in E_n} \delta_x, \qquad \mu_n = \frac{1}{n} \sum_{i=0}^{n-1} T_*^i \nu_n,$$

où $T_*^i \mu(A) = \mu(T^{-i}A)$.

Soit k_n une sous-suite telle que

$$h(T,\epsilon) = \limsup_{n\to\infty} \frac{1}{n} \log \text{card } E_n = \lim_{n\to\infty} \frac{1}{k_n} \log \text{card } E_{k_n}.$$

Soit μ un point d'accumulation de la suite $(\mu_{k_n})_{n\in\mathbb{N}}$ dans $\mathcal{M}(X)$, dont l'existence est garantie par le Théorème 4.2. Il existe donc une sous-suite $(m_n)_{n\in\mathbb{N}}$ de $(k_n)_{n\in\mathbb{N}}$ telle que $\mu_{m_n} \to \mu$, $n \to \infty$. La mesure μ est T-invariante parce que

$$\begin{aligned} T_*\mu &= \lim_{n\to\infty} \frac{1}{m_n} \sum_{i=0}^{m_n-1} T_*^{i+1} \nu_{m_n} \\ &= \lim_{n\to\infty} \frac{1}{m_n} \Big(\sum_{i=0}^{m_n-1} T_*^i \nu_{m_n} - \nu_{m_n} + T_*^{m_n} \nu_{m_n} \Big) = \mu. \end{aligned}$$

On construit maintenant une partition ξ satisfaisant les hypothèses du lemme 4.2, afin de pouvoir calculer son entropie sous μ. Pour chaque $x \in X$ les valeures de $r > 0$ pour lesquelles $\mu(\partial B(x,r)) > 0$ sont au plus dénombrables. On peut donc prendre $r(x) < \epsilon/2$ tel que $\mu(\partial B(x,r(x))) > 0$. La famille d'ouverts $\{B(x,r(x))\}$ est alors un recouvrement de X, dont il est possible d'extraire un sous-recouvrement fini $\{B_1,\ldots,B_k\}$. Nous definissons $\xi = \{C_1,\ldots,C_k\}$ par

$$C_1 = \overline{B_1}, \quad C_i = \overline{B_i} \setminus \bigcup_{j=0}^{i-1} \overline{B_j}, \quad \text{si} \quad i > 1.$$

Nous avons par ailleurs diam $C_i < \epsilon$, $\mu(\partial C_i) = 0$ parce que $\partial C_i \subset \cup_{j=1}^k \partial B_j$. Dans chaque élement de $\xi_n = \bigvee_{i=0}^{n-1} T^i \xi$ il y a au plus un point de E_n parce que si $\{x, y\} \subset C_{i_0} \cap \cdots \cap T^{-n+1} C_i i_{n-1}$ pour $i_o, \ldots, i_{n-1} \in \{1, \ldots, k\}$, alors $d(T^j x, T^j y) \leq$ diam $C_{i_j} < \epsilon$, et on parvient à la contradiction $d_n(x, y) < \epsilon$. Donc d'après le lemme 4.2 il y a exactement (card E_n) éléments A de ξ_n avec $\nu_n(A) = 1/\text{card } E_n$ et les autres éléments de ξ_n étant de mesure ν_n nulle. On peut donc calculer H_{ν_n} :

$$H_{\nu_n}(\xi_n) = \text{card } E_n \cdot (-(1/\text{card } E_n) \log \text{card } (1/E_n)) = \log \text{card } E_n.$$

On prend n, m entiers et on écrit $n = qm + r$, où $q \geq 0$, $0 \leq r < m$.

Par la convexité de la fonction ψ on a

$$\begin{aligned}
H_{\mu_n}(\xi_m) &= -\sum_{A \in \xi_m} \psi(\mu_n(A)) = -\sum_{A \in \xi_m} \psi(\frac{1}{n} \sum_{l=0}^{n-1} \nu_n(T^{-l}A)) \\
&\geq -\frac{1}{n} \sum_{A \in \xi_m} \sum_{l=0}^{n-1} \psi(\nu_n(T^{-l}A) \\
&= \frac{1}{n} \sum_{l=0}^{n-1} H_{\nu_n}(T^{-l}\xi_m) = \frac{1}{n} \sum_{l=0}^{mq+r-1} H_{\nu_n}(T^{-l}\xi_m) \\
&\geq \frac{1}{n} \sum_{l=0}^{mq-1} H_{\nu_n}(T^{-l}\xi_m) = \frac{1}{n} \sum_{i=0}^{m-1} \sum_{j=0}^{q-1} H_{\nu_n}(T^{-jm-i}\xi_m).
\end{aligned}$$

On détermine maintenant $\sum_{j=0}^{q-1} H_{\nu_n}(T^{-jm-i}\xi_m)$.

Clairement, ξ_{qm+m} est plus fine que $\xi_{qm+r} = \xi_n$. Donc $H_{\nu_n}(\xi_n) \leq H_{\nu_n}(\xi_{qm+m})$. On a aussi

$$\xi_{qm+m} = \bigvee_{j=0}^{q-1} T^{-jm-i}\xi_m \vee \bigvee_{j=qm+i}^{qm+m-1} T^{-j}\xi \vee \bigvee_{j=0}^{i-1} T^{-j}\xi.$$

Comme card $\bigvee_{j=qm+i}^{qm+m-1} T^{-j}\xi \vee \bigvee_{j=0}^{i-1} \xi \leq (\text{card } \xi)^{2m}$, d'après la Remarque 1

$$H_{\nu_n}(\bigvee_{j=qm+i}^{qm+m-1} T^{-j}\xi \vee \bigvee_{j=0}^{i-1} \xi) \leq 2m \log \text{card } \xi.$$

Donc

$$H_{\nu_n}(\xi_{qm+m}) \leq \sum_{j=0}^{q-1} H_{\nu_n}(T^{-jm-i}\xi_m) + 2m \log \text{card } \xi.$$

Finalement,

$$H_{\mu_n}(\xi_m) \geq \frac{1}{n} \sum_{i=0}^{m-1} \sum_{j=0}^{q-1} H_{\nu_n}(T^{-jm-i}\xi_m) \geq \frac{1}{n}(mH_{\nu_n}(\xi_n) - 2m^2 \log \text{card } \xi),$$

$$\frac{m}{n} \log \text{card } E_n = \frac{m}{n} H_{\nu_n}(\xi_n) \leq H_{\mu_n}(\xi_m) + \frac{2m^2}{n} \log \text{card } \xi.$$

On remplace n par m_n

$$\frac{1}{m_n}\log\operatorname{card}\, E_{m_n} \leq \frac{1}{m}H_{\mu_{m_n}}(\xi_m) + \frac{2m}{m_n}\log\operatorname{card}\,\xi.$$

On passe à limite

$$h(T,\epsilon) \leq \lim_{n\to\infty}\frac{1}{m}H_{\mu_{m_n}}(\xi_m).$$

D'après le lemme 4.2 pour chaque $A \in \xi_m$ $\lim_{n\to\infty}\mu_{m_n}(A) = \mu(A)$, donc

$$\lim_{n\to\infty}\frac{1}{m}H_{\mu_{m_n}}(\xi_m) = \frac{1}{m}H_\mu(\xi_m).$$

Finalement, en faisant tendre m vers l'infini on obtient :

$$\begin{aligned} h(T,\epsilon) &\leq \lim_{m\to\infty}\frac{1}{m}H_\mu(\xi_m) = h_\mu(T,\xi) \leq h_\mu(T) \\ &\leq \sup\{h_\mu(T) : \mu \text{ est une mesure de probabilité } T\text{-invariante}\}, \end{aligned}$$

et par passage à la limite quand ϵ tend vers 0 on a le résultat final

$$h(T) \leq \sup\{h_\mu(T) : \mu \text{ est la mesure de probabilité } T\text{-invariant}\}.$$

$\square$

Références

[1] R. L. Adler, A. G. Konheim, M. H. McAndrew. *Topological entropy*, 1965

[2] Peter Walters. *Topological entropy*, Springer 1982

[3] Luis Barreira. *Ergodic Theory, Hyperbolic Dynamics and Dimension Theory*, Springer 2012

[4] Michael Brin, Garrett Stuck. *Dynamical Systems*, Cambridge University Press 2002