La distance de réarrangement, duale de la fonctionnelle de Bowen

Thierry Bousch*

12 mai 2010

Abstract

On the space of signed invariant measures of $A^{\mathbb{N}}$, one constructs a norm (and hence a distance) which seems to have a particular significance in dynamics. I shall present some of its properties, most particularly a duality theorem à la Kantorovich-Rubinstein, which gives an expression of this distance using couplings.

Résumé

Sur l'espace des mesures invariantes signées de $A^{\mathbb{N}}$, on construit une norme (et donc une distance) qui semble avoir une importance particulière du point de vue dynamique. Je présenterai quelques-unes de ses propriétés, et tout particulièrement un théorème de dualité à la Kantorovich-Rubinstein, qui permet d'exprimer cette distance en termes de couplages.

Titre anglais: The rearrangement distance, dual of the Bowen functional

Mots-clés: Dynamique symbolique, mesure invariante, dualité

Classification AMS (2000): 37B10, 46A20, 90C05

Table des matières

1 Introduction **2**

2 Fonctionnelle de Bowen et norme de réarrangement **2**

2.1 Fonctionnelle (et norme) de Bowen . . . 2

2.2 Norme de réarrangement . . . 5

2.3 Propriétés topologiques . . . 7

2.4 Comparaison avec la distance d'Ornstein . . . 7

3 Une propriété de dualité **9**

3.1 Enoncé du théorème . . . 10

3.2 Résultats auxiliaires . . . 11

3.3 Démonstration du théorème . . . 18

3.4 Couplages et réarrangements . . . 19

4 La norme de réarrangement sur $A^{\mathbb{Z}}$ **20**

Références **22**

*Laboratoire de Mathématique (UMR 8628 du CNRS), bât. 425/430, Université de Paris-Sud, 91405 Orsay Cedex, France. E-mail: Thierry.Bousch@math.u-psud.fr

1 Introduction

Soit A un ensemble fini non vide (un "alphabet") et $X = A^{\mathbb{N}}$ l'espace symbolique unilatère, muni de l'application de décalage $\lhd : X \to X$ définie par $\lhd(x_0, x_1, x_2, \ldots) = (x_1, x_2, \ldots)$.

On note $\mathcal{M}(X)$ l'ensemble de toutes les mesures boréliennes signées sur X, c'est-à-dire le dual de $C(X)$, et $\mathcal{M}^{(1)}(X)$, $\mathcal{M}^+(X)$, $\mathcal{M}^1(X)$ ses sous-ensembles constitués des mesures de variation totale $\leqslant 1$, positives, de probabilité, respectivement. (Dans tout cet article, le terme "mesure" désignera une mesure borélienne *signée* sur X, sauf mention contraire.)

L'ensemble des mesures $\lhd$-invariantes de $\mathcal{M}(X)$ est noté $\mathcal{M}_\lhd(X)$ et on définit similairement $\mathcal{M}_\lhd^{(1)}(X)$, $\mathcal{M}_\lhd^+(X)$, $\mathcal{M}_\lhd^1(X)$.

L'objectif de cet article est, pour l'essentiel, de construire une distance sur $\mathcal{M}_\lhd^1(X)$ qui soit "naturelle" d'un point de vue dynamique, notamment par sa topologie, et se "comporte bien" (en un sens qui sera précisé dans un article ultérieur) vis-à-vis des substitutions.

La distance de Kantorovitch (parfois appelée distance de Wasserstein, cf. [Rus, Ver]) induit une topologie naturelle (la topologie faible) sur $\mathcal{M}_\lhd^1(X)$, mais nécessite de choisir préalablement une distance sur X, et aucune d'elles n'est vraiment naturelle. La distance d'Ornstein [Orn, GNS] sur $\mathcal{M}_\lhd^1(X)$ est certainement plus naturelle d'un point de vue dynamique, mais elle définit une topologie nouvelle, intermédiaire entre les topologies faible et forte, qu'on ne sait pas décrire autrement. En effet, le principal intérêt de la distance d'Ornstein est que nombre de propriétés dynamiques de la mesure invariante (comme l'ergodicité, le mélange, l'entropie nulle...) sont *fermées* pour cette distance, alors que ce n'est évidemment pas le cas en topologie faible. En particulier, $\mathcal{M}_\lhd^1(X)$ n'est pas compact pour la distance d'Ornstein. Et, bien qu'intéressante dans l'absolu, cette distance n'est pas pertinente pour les problèmes que j'ai en tête.

2 Fonctionnelle de Bowen et norme de réarrangement

2.1 Fonctionnelle (et norme) de Bowen

On note $LC_n(X)$ l'ensemble des fonctions $X \to \mathbb{R}$ localement constantes d'ordre $\leqslant n$, c'est-à-dire qui ne dépendent que de $x_0, x_1, \ldots, x_{n-1}$. C'est un espace de dimension finie, isomorphe à $\mathbb{R}^{A^n}$. Soit $LC(n) = \bigcup_n LC_n(X)$ l'espace de toutes les fonctions localement constantes.

Pour tout entier naturel n, notons $\pi_{<n} : X \to A^n$ la fonction définie par $\pi_{<n} : x \mapsto (x_0, x_1, \ldots, x_{n-1})$, tandis que les $\pi_n : X \to A$ désignent les projections canoniques $x \mapsto x_n$. Les fonctions de $LC_n(X)$ sont donc exactement les fonctions $X \to \mathbb{R}$ qui se factorisent à travers $\pi_{<n}$. On appellera cylindres d'ordre n (ou n-cylindres) les fibres de $\pi_{<n}$.

Pour toute fonction $f : X \to \mathbb{R}$, on définit les sommes de Birkhoff $S_n f : X \to \mathbb{R}$ par la formule $S_n f = f + f\lhd + \cdots + f\lhd^{n-1}$. Cette famille de fonctions constitue un cocycle: on a $S_0 f = 0$ et $S_{n+m} f = S_n f + (S_m f)\lhd^n$ pour tous $n, m \geqslant 0$.

On dit qu'une fonction $f : X \to \mathbb{R}$ vérifie la condition de Bowen (ou que "f est Bowen") s'il existe une constante $C \geqslant 0$ vérifiant la propriété suivante: quel que soit l'entier $n \geqslant 1$ et quels que soient $x, y \in X$ tels que $\pi_{<n}(x) = \pi_{<n}(y)$, on a $|S_n f(y) - S_n f(x)| \leqslant 2C$, i.e.,

$$(x_0, \ldots, x_{n-1}) = (y_0, \ldots, y_{n-1}) \quad \Longrightarrow \quad \Big| \sum_{0 \leqslant k < n} f(\lhd^k y) - f(\lhd^k x) \Big| \leqslant 2C. \tag{2.1}$$

La plus petite constante C vérifiant cette condition est appelée *constante de Bowen* de f et notée $\mathrm{bw}(f)$. Les fonctions Bowen forment manifestement un espace vectoriel. On conviendra que $\mathrm{bw}(f) = +\infty$ pour les fonctions f qui ne sont pas Bowen.

Définition équivalente: une fonction $f : X \to \mathbb{R}$ est Bowen de constante au plus C si et seulement si, pour tout $n \geqslant 1$, il existe une fonction $f_n \in LC_n(X)$ telle que $\|S_n f - f_n\| \leqslant C$.

Cette condition de "régularité" sur les fonctions a été introduite par R. Bowen [Bow] pour les besoins du formalisme thermodynamique [HR, Rue]. Une condition un peu différente, mais de même nature, est due à P. Walters [Wa1].

Quelques propriétés immédiates sont: (i) toute fonction Bowen est bornée; (ii) toute fonction localement constante est Bowen; (iii) une fonction $f : X \to \mathbb{R}$ satisfait $\mathrm{bw}(f) = 0$ si et seulement si elle est dans $LC_1(X)$; (iv) si $f : X \to \mathbb{R}$ est une fonction bornée, son cobord $g = f\triangleleft - f$ est Bowen, avec $\mathrm{bw}(g) \leqslant \mathrm{osc}(f)$; (v) la fonctionnelle $\mathrm{bw} : \mathbb{R}^X \to [0, \infty]$ est semi-continue inférieurement pour la topologie de la convergence simple: si $f_\alpha \to f$ simplement, alors $\liminf_\alpha \mathrm{bw}(f_\alpha) \geqslant \mathrm{bw}(f)$.

La fonctionnelle de Bowen $\mathrm{bw}(\cdot)$ constitue donc une semi-norme sur l'espace $LC(X)$ des fonctions localement constantes. Définissons la *norme de Bowen*

$$\mathrm{Nbw}(f) = \mathrm{bw}(f) \vee \|f\| \tag{2.2}$$

où $\|f\| = \sup |f|$ est la norme uniforme de f.

Définition équivalente: une fonction $f : X \to \mathbb{R}$ est Bowen de norme au plus C si et seulement si, pour tout $n \geqslant 1$, il existe une fonction $f_n \in LC_n(X)$ telle que $\|S_n f - f_n\| \leqslant C$ et $\|f_n\| \leqslant (n-1)C$.

Lemme 2.1. *Soient $a, b \geqslant 0$ réels, et $f : X \to \mathbb{R}$ une fonction Bowen telle que* $\mathrm{bw}(f) \leqslant a$ *et* $\|f\| \leqslant a + b$. *Il existe une fonction $g \in LC_1(X)$ vérifiant $\|f - g\| \leqslant a$ et $\|g\| \leqslant b$.*

Démonstration. Soit u le plus grand élément de $LC_1(X)$ qui minore f, et v le plus petit élément de $LC_1(X)$ qui majore f. On a

$$-(a + b) \leqslant u \leqslant f \leqslant v \leqslant a + b$$

et d'autre part, d'après (2.1) l'oscillation de f sur tout 1-cylindre est majorée par $2a$, ce qui revient à dire que $v - u \leqslant 2a$. En particulier, on ne peut pas avoir simultanément $v(x) > a$ et $u(x) < -a$. Définissons $g : X \to \mathbb{R}$ comme la médiane des fonctions $v - a$, $u + a$ et 0, i.e.

$$g(x) = \begin{cases} v(x) - a & \text{si } v(x) > a, \\ u(x) + a & \text{si } u(x) < -a, \\ 0 & \text{sinon.} \end{cases}$$

Evidemment g est dans $LC_1(X)$, et on vérifie que dans chacun des trois cas, on a $|g(x)| \leqslant b$ et $|(f - g)(x)| \leqslant a$. ☐

Lemme 2.2. *Si une fonction $f : X \to \mathbb{R}$ est dans $LC_n(X)$ pour un certain $n \geqslant 1$, alors* $\mathrm{bw}(f) \leqslant (n-1)\|f\|$.

Démonstration. Donnons-nous un entier $m \geqslant 1$ et deux points $x, y \in X$ tels que $\pi_{<m}(x) = \pi_{<m}(y)$. Dans l'expression

$$S_m f(y) - S_m f(x) = \sum_{0 \leqslant i < m} f(\triangleleft^i y) - f(\triangleleft^i x),$$

seuls les indices $i > m - n$ contribuent effectivement; en effet, pour $i \leqslant m - n$, les n premières lettres de $\triangleleft^i x$ et $\triangleleft^i y$ coïncident et donc $f(\triangleleft^i x) = f(\triangleleft^i y)$. Par conséquent,

$$\begin{aligned} \Big| S_m f(y) - S_m f(x) \Big| &= \Big| \sum_{(m-n+1)^+ \leqslant i < m} f(\triangleleft^i y) - f(\triangleleft^i x) \Big| \\ &\leqslant \sum_{(m-n+1)^+ \leqslant i < m} \Big| f(\triangleleft^i y) - f(\triangleleft^i x) \Big| \leqslant 2(n-1)\|f\| \end{aligned}$$

ce qu'il fallait démontrer. $\square$

On peut se demander si les fonctions continues Bowen sont approchables par des fonctions localement constantes. Au sens de la norme Nbw, ce n'est pas vrai: les fonctions limites de fonctions localement constantes pour Nbw vérifient la *condition de Walters* [Wa1, Bou] et on sait que cette condition est strictement plus forte que la condition de Bowen plus la continuité [Wa4]. On peut toutefois les approcher dans un sens plus faible:

Lemme 2.3. *Soit $f : X \to \mathbb{R}$ une fonction continue Bowen. Pour tout $\epsilon > 0$, il existe une fonction localement constante $g : X \to \mathbb{R}$ telle que $\|g\| \leqslant \|f\|$, $\mathrm{bw}(g) \leqslant \mathrm{bw}(f)$, et $\|f - g\| \leqslant \epsilon$.*

Démonstration. Soit α une lettre distinguée dans l'alphabet A, et pour tout $n \geqslant 0$ soit $I_n : X \to X$ la fonction idempotente définie par

$$I_n(x_0, x_1, \ldots) = (x_0, \ldots, x_{n-1}, \alpha, \alpha, \ldots).$$

Définissons $f_n = fI_0 + fI_1 + \cdots + fI_n$. Il est évident que les fonctions $f_n/(n+1)$ sont localement constantes, convergent uniformément vers f, et que leur norme uniforme est $\leqslant \|f\|$. J'affirme que

$$\mathrm{bw}(f_n) \leqslant (n-1)\,\mathrm{bw}(f) \tag{2.3}$$

pour tout $n \geqslant 1$.

En effet, donnons-nous deux entiers $m, n \geqslant 1$ et deux points $x, y \in X$ tels que $\pi_{<m}(x) = \pi_{<m}(y)$. On peut écrire

$$\begin{aligned} S_m f_n(y) - S_m f_n(x) &= \sum_{0 \leqslant i < m} f_n(\lhd^i y) - f_n(\lhd^i x) \\ &= \sum_{\substack{0 \leqslant i < m \\ 0 \leqslant j \leqslant n}} f(I_j \lhd^i y) - f(I_j \lhd^i x) \\ &= \sum_{\substack{0 \leqslant i < m \\ 0 \leqslant j \leqslant n}} f(\lhd^i I_{i+j}\, y) - f(\lhd^i I_{i+j}\, x) \\ &= \sum_{0 \leqslant s < m+n} D(s) \end{aligned}$$

où on a posé

$$D(s) = \sum_{\substack{0 \leqslant i < m \\ 0 \leqslant s-i \leqslant n}} f(\lhd^i I_s y) - f(\lhd^i I_s x).$$

Pour $s \leqslant m$ on a $I_s x = I_s y$ et donc $D(s) = 0$. Pour les autres valeurs de s, qui satisfont $m < s < m + n$, on a

$$D(s) = \sum_{r \leqslant i < m} f(\lhd^i I_s y) - f(\lhd^i I_s x) = \sum_{0 \leqslant i < m-r} f(\lhd^i y') - f(\lhd^i x')$$

où on a posé $r = (s-n)^+$, $x' = \lhd^r I_s x$ et $y' = \lhd^r I_s y$. L'hypothèse $\pi_{<m}(x) = \pi_{<m}(y)$ entraîne $\pi_{<m}(I_s x) = \pi_{<m}(I_s y)$ et par suite $\pi_{<m-r}(x') = \pi_{<m-r}(y')$, donc

$$|D(s)| = \left|S_{m-r} f(y') - S_{m-r} f(x')\right| \leqslant 2\,\mathrm{bw}(f)$$

Il en résulte

$$|S_m f_n(y) - S_m f_n(x)| \;\leqslant\; \sum_{m < s < m+n} |D(s)| \;\leqslant\; 2(n-1)\,\mathrm{bw}(f)$$

ce qui établit l'inégalité (2.3), et le lemme. $\square$

2.2 Norme de réarrangement

La *norme de réarrangement* $\|\cdot\|_R$ sur $\mathcal{M}(X)$ est définie comme la norme duale de $\mathrm{Nbw}(\cdot)$ sur $LC(X)$: pour toute mesure μ sur X, on définit

$$\|\mu\|_R = \operatorname*{Sup}_{\substack{f\in LC(X)\\ \mathrm{Nbw}(f)\leqslant 1}} \langle f,\mu\rangle = \operatorname*{Sup}_{\substack{f\in LC(X)\\ \mathrm{Nbw}(f)\leqslant 1}} \big|\langle f,\mu\rangle\big|. \tag{2.4}$$

On obtient immédiatement l'encadrement

$$\|\pi_0\mu\| \leqslant \|\pi_{<2}\,\mu\| \leqslant \|\mu\|_R \leqslant \|\mu\|_{TV} \tag{2.5}$$

en observant que l'ensemble $\big\{f\in LC(X) : \mathrm{Nbw}(f)\leqslant 1\big\}$ contient $\big\{f\in LC_2(X) : \|f\|\leqslant 1\big\}$, et est contenu dans $\big\{f\in C(X) : \|f\|\leqslant 1\big\}$. En particulier, pour les mesures positives, la norme de réarrangement coïncide avec la norme de la variation totale.

Lemme 2.4. *Soit μ une mesure sur X. On a*

$$\operatorname*{Sup}_{\substack{f\in LC(X)\\ \mathrm{bw}(f)\leqslant a\\ \|f\|\leqslant a+b}} \langle f,\mu\rangle = a\,\|\mu\|_R + b\,\|\pi_0\mu\| \tag{2.6}$$

quels que soient les réels $a,b\geqslant 0$.

Cela entraîne en particulier

$$\operatorname*{Sup}_{\substack{f\in LC(X)\\ \mathrm{bw}(f)\leqslant 1}} \langle f,\mu\rangle = \begin{cases}\|\mu\|_R & \text{si } \pi_0\mu=0,\\ +\infty & \text{sinon.}\end{cases} \tag{2.7}$$

Démonstration. D'après le lemme 2.1, l'ensemble $\big\{f\in LC(X) : \mathrm{bw}(f)\leqslant a \text{ et } \|f\|\leqslant a+b\big\}$ peut s'écrire comme la somme de Minkowski des ensembles $\big\{h\in LC(X) : \mathrm{Nbw}(h)\leqslant a\big\}$ et $\big\{g\in LC_1(X) : \|g\|\leqslant b\big\}$. Par conséquent,

$$\begin{aligned}\operatorname*{Sup}_{\substack{f\in LC(X)\\ \mathrm{bw}(f)\leqslant a\\ \|f\|\leqslant a+b}} \langle f,\mu\rangle &= \operatorname*{Sup}_{\substack{h\in LC(X)\\ \mathrm{Nbw}(h)\leqslant a}}\ \operatorname*{Sup}_{\substack{g\in LC_1(X)\\ \|g\|\leqslant b}} \langle h+g,\mu\rangle\\ &= \operatorname*{Sup}_{\substack{h\in LC(X)\\ \mathrm{Nbw}(h)\leqslant a}} \langle h,\mu\rangle + \operatorname*{Sup}_{\substack{g\in LC_1(X)\\ \|g\|\leqslant b}} \langle g,\mu\rangle\\ &= a\,\|\mu\|_R + b\,\|\pi_0\mu\|\end{aligned}$$

ce qu'il fallait démontrer. □

A première vue, le choix de la norme (2.2) sur $LC(X)$ peut paraître artificiel. On peut imaginer beaucoup de manières de combiner la semi-norme bw avec la norme uniforme (ou autre chose) afin d'en faire une vraie norme. En fait, il y a de nombreuses raisons de choisir cette norme plutôt qu'une autre. Une de ces raisons est que beaucoup d'autres normes similaires sur $LC(X)$ ont une norme duale sur $\mathcal{M}(X)$ qui peut s'exprimer à l'aide de la norme de réarrangement, alors que le contraire n'est pas toujours vrai.

Par exemple, le lemme 2.4 revient à dire que si $LC(X)$ est muni de la norme

$$f \longmapsto \lambda^{-1}\,\|f\| \vee \mathrm{bw}(f)$$

avec $\lambda \geqslant 1$ réel, la norme duale sur $\mathcal{M}(X)$ sera donnée par

$$\mu \longmapsto \|\mu\|_R + (\lambda - 1)\,\|\pi_0\mu\|\,.$$

Un autre exemple, important pour la comparaison avec la distance d'Ornstein, est la *norme de Bowen augmentée* sur $LC(X)$, définie par

$$\mathrm{Nbwa}(f) = \mathrm{bw}(f) + \mathrm{Nbw}(f) = \max\bigl\{2\,\mathrm{bw}(f), \mathrm{bw}(f) + \|f\|\bigr\}. \tag{2.8}$$

Elle est équivalente à la norme de Bowen proprement dite, car $\mathrm{Nbw} \leqslant \mathrm{Nbwa} \leqslant 2\,\mathrm{Nbw}$. Soit $\|\cdot\|_{RR}$ la norme duale sur $\mathcal{M}(X)$, la *norme de réarrangement réduite*, définie par

$$\|\mu\|_{RR} \;=\; \operatorname*{Sup}_{\substack{f\in LC(X)\\ \mathrm{Nbwa}(f)\leqslant 1}} \langle f,\mu\rangle \;=\; \operatorname*{Sup}_{\substack{f\in LC(X)\\ \mathrm{Nbwa}(f)\leqslant 1}} \bigl|\langle f,\mu\rangle\bigr|. \tag{2.9}$$

Cette norme peut se calculer à partir de la norme de réarrangement proprement dite, de la façon suivante. Observant que

$$\mathrm{Nbwa}(f) \leqslant 1 \iff \exists a \in [0,\tfrac{1}{2}]\ \ \mathrm{bw}(f) \leqslant a \text{ et } \|f\| \leqslant 1 - a,$$

on déduit facilement du lemme 2.4 que

$$\|\mu\|_{RR} = \|\pi_0\mu\| \vee \tfrac{1}{2}\|\mu\|_R \tag{2.10}$$

pour toute mesure μ sur X. Cela entraîne en particulier

$$\|\pi_0\mu\| \leqslant \|\mu\|_{RR} \leqslant \|\mu\|_R \leqslant \|\mu\|_{TV}\,. \tag{2.11}$$

Lemme 2.5. *Soit μ une mesure invariante sur X. On a*

$$\|\pi_{<n}\mu\| \;\leqslant\; n\,\|\mu\|_R \;\leqslant\; \|\mu\|_{TV} + (n-1)\,\|\pi_{<n}\mu\| \tag{2.12}$$

quel que soit l'entier $n \geqslant 1$.

Démonstration. D'après le lemme 2.2, on a $\mathrm{Nbw}(f) \leqslant (n-1) \vee 1 \leqslant n$ pour toute fonction $f \in LC_n(X)$ telle que $\|f\| \leqslant 1$, et donc

$$\|\pi_{<n}\mu\| = \operatorname*{Sup}_{\substack{f\in LC_n(X)\\ \|f\|\leqslant 1}} \langle f,\mu\rangle \leqslant \operatorname*{Sup}_{\substack{f\in LC(X)\\ \mathrm{Nbw}(f)\leqslant n}} \langle f,\mu\rangle = n\,\|\mu\|_R$$

ce qui établit la première inégalité (qui, on le voit, n'utilise pas l'invariance de μ).
Soit $f : X \to \mathbb{R}$ une quelconque fonction localement constante vérifiant $\mathrm{Nbw}(f) \leqslant 1$. Il existe (voir §2.1) une fonction $f_n \in LC_n(X)$ telle que $|S_n f - f_n| \leqslant 1$ et $\|f_n\| \leqslant n-1$. Par conséquent,

$$\begin{aligned} n\langle f,\mu\rangle = \langle S_n f,\mu\rangle &= \langle S_n f - f_n,\mu\rangle + \langle f_n,\mu\rangle \\ &\leqslant \|S_n f - f_n\|\cdot\|\mu\|_{TV} + \|f_n\|\cdot\|\pi_{<n}\mu\| \\ &\leqslant \|\mu\|_{TV} + (n-1)\,\|\pi_{<n}\mu\| \end{aligned}$$

ce qui établit la deuxième inégalité du lemme. □

2.3 Propriétés topologiques

J'ai défini la norme de réarrangement pour n'importe quelle mesure sur X, mais c'est sa restriction aux mesures *invariantes* qui est vraiment importante. Et il est naturel de se demander quelle topologie sur $\mathcal{M}_{\triangleleft}(X)$ elle définit.

Proposition 2.6. *La topologie définie sur la boule $\mathcal{M}^{(1)}_{\triangleleft}(X)$ par la distance de réarrangement $d_R(\mu_1, \mu_2) = \|\mu_2 - \mu_1\|_R$ n'est autre que la topologie de la convergence faible des mesures.*

Ce résultat reste évidemment valable sur tout sous-ensemble de $\mathcal{M}_{\triangleleft}(X)$ *borné* pour la norme de la variation totale. Ce n'est pas vrai sur l'espace $\mathcal{M}_{\triangleleft}(X)$ tout entier, ne serait-ce que parce que la topologie faible $\sigma\big[\mathcal{M}_{\triangleleft}(X), C(X)\big]$ n'est pas normable, ni même métrisable (pour $\#A > 1$).

Démonstration. Les topologies à comparer étant métrisables, il suffit de démontrer qu'elles ont les mêmes suites convergentes. Soient $\mu_1, \mu_2, \ldots$ et μ_∞ des éléments quelconques de $\mathcal{M}^{(1)}_{\triangleleft}(X)$, c'est-à-dire des mesures invariantes de variation totale $\leqslant 1$.

Supposons d'abord $\|\mu_m - \mu_\infty\|_R \to 0$ quand $m \to \infty$. D'après le lemme 2.5, on a l'inégalité

$$\|\pi_{<n}(\mu_m - \mu_\infty)\| \leqslant n\,\|\mu_m - \mu_\infty\|_R$$

pour tous $m, n \geqslant 1$, ce qui entraîne $\lim_{m\to\infty} \|\pi_{<n}(\mu_m - \mu_\infty)\| = 0$, quel que soit $n \geqslant 1$. Mais cela revient précisément à dire que les μ_m convergent faiblement vers μ_∞.

Réciproquement, supposons que les μ_m convergent faiblement vers μ_∞, c'est-à-dire que $\lim_{m\to\infty} \|\pi_{<n}(\mu_m - \mu_\infty)\| = 0$ pour tout $n \geqslant 1$. Par la deuxième inégalité du lemme 2.5,

$$\begin{aligned} n\,\|\mu_m - \mu_\infty\|_R &\leqslant \|\mu_m - \mu_\infty\|_{TV} + (n-1)\,\|\pi_{<n}(\mu_m - \mu_\infty)\| \\ &\leqslant 2 + (n-1)\,\|\pi_{<n}(\mu_m - \mu_\infty)\| \end{aligned}$$

pour tous $m, n \geqslant 1$. Cela entraîne $\limsup_{m\to\infty} \|\mu_m - \mu_\infty\|_R \leqslant 2/n$ pour tout $n \geqslant 1$, et par conséquent $\|\mu_m - \mu_\infty\|_R \to 0$. □

2.4 Comparaison avec la distance d'Ornstein

Cette distance a été introduite en théorie ergodique par Donald Ornstein vers 1970 pour résoudre le problème de l'équivalence mesurable des décalages de Bernoulli [Orn], mais des idées très similaires ont été introduites indépendamment par Dobrouchine [Do1, Do2] pour prouver l'unicité des mesures de Gibbs à haute température; elle a ensuite trouvé des applications en théorie de l'information [GNS, Shi].

Je rappelle rapidement ici sa définition et quelques-unes de ses propriétés. Soit (A, d) un espace métrique compact non vide, et $X = A^S$ où S est un ensemble au plus dénombrable. (Si A est un ensemble à deux éléments, les élements de X peuvent être vus comme des configurations de spins, ou de bits.) La distance d'Ornstein est une distance sur $\mathcal{M}^1(X)$ définie comme suit: étant donné deux probas μ_0, μ_1 sur X, la distance d'Ornstein $d_O(\mu_0, \mu_1)$ est le plus petit réel $C \geqslant 0$ pour lequel il existe un couplage ξ de μ_0 et μ_1 (c'est-à-dire une proba sur X^2 telle que $\partial_0\xi = \mu_0$ et $\partial_1\xi = \mu_1$, où $\partial_0, \partial_1 : X^2 \to X$ sont les projections naturelles), tel que les écarts $E_s : X^2 \to \mathbb{R}$ définis par $E_s(x, y) = d(x_s, y_s)$ satisfassent $\langle E_s, \xi\rangle \leqslant C$ pour tout $s \in S$.

Cette distance se réduit évidemment à la distance de Kantorovitch si S est un singleton. Mais dans le cas général, il y a une différence essentielle avec le problème du transport optimal: ici, on doit minimiser *simultanément* toutes les moyennes $\langle E_s, \xi\rangle$ pour s décrivant S.

Si A est un ensemble fini sur lequel aucune distance n'est donnée *a priori* (ce qui est notre cas), munissons-le de la distance

$$d(a,b) = \|\delta_b - \delta_a\|_{TV} = \begin{cases} 2 & \text{si } a \neq b, \\ 0 & \text{sinon.} \end{cases} \tag{2.13}$$

Cette convention est légèrement non-standard: traditionnellement on munit A de la *moitié* de la distance ci-dessus, et donc la distance d'Ornstein traditionnelle, habituellement notée $\bar{d}$, est la moitié de la distance d'Ornstein d_O considérée dans cet article.

Comme la distance de Kantorovitch, la distance d'Ornstein provient d'une norme (sur les mesures d'intégrale nulle) et satisfait un principe de dualité analogue: la distance d'Ornstein $d_O(\mu_0, \mu_1)$ est la borne supérieure des intégrales $\langle f, \mu_1 - \mu_0 \rangle$ où $f : X \to \mathbb{R}$ est une fonction continue vérifiant

$$\forall x, y \in X \qquad |f(y) - f(x)| \leqslant \sum_{s \in S} \ell_s \, d(x_s, y_s)$$

et les $(\ell_s)_{s \in S}$ des réels positifs tels que $\sum_{s \in S} \ell_s \leqslant 1$. Je ne connais pas de référence pour cet énoncé précis, mais cette dualité est implicite dans la présentation que donne Lanford [Lan] du théorème d'unicité de Dobrouchine, qui utilise des fonctions-test f vérifiant les inégalités ci-dessus, là où Dobrouchine utilisait des couplages. Cette façon de voir a été reprise dans [Fol, Kun] ainsi que dans le livre de Georgii [Geo] (voir en particulier p.456 les notes bibliographiques pour le chapitre 8).

J'ai déjà dit que la distance d'Ornstein définissait, sur les mesures de probabilité invariantes, une topologie intermédiaire entre la topologie faible et la topologie forte. Similairement, on a:

Proposition 2.7. *Sur l'espace $\mathcal{M}^1_{\lhd}(X)$ des mesures de probabilité invariantes de X, on a l'encadrement*

$$d_{RR} \leqslant d_O \leqslant d_{TV} \tag{2.14}$$

entre la distance de réarrangement réduite $d_{RR}(\mu_0, \mu_1) = \|\mu_1 - \mu_0\|_{RR}$, la distance d'Ornstein, et la distance de la variation totale $d_{TV}(\mu_0, \mu_1) = \|\mu_1 - \mu_0\|_{TV}$.

Démonstration. La deuxième inégalité découle immédiatement de la définition de la distance d'Ornstein (et $\operatorname{diam} A \leqslant 2$). Pour la première inégalité, soient μ_0, μ_1 deux probas invariantes et $C = d_O(\mu_0, \mu_1)$ leur distance d'Ornstein; elles admettent donc un couplage ξ tel que $\langle E_s, \xi \rangle \leqslant C$ pour tous $s \in \mathbb{N}$.

Soit $n \geqslant 1$ entier, et x, y deux points de X. Notons D_n l'ensemble des indices $i \in [0, n[$ tels que $x_i \neq y_i$, et soit $\alpha = \alpha_n(x,y)$ son cardinal. Notons $e_0, \ldots, e_{\alpha-1}$ les éléments de D_n, rangés par ordre croissant, et posons $e_{-1} = -1$ et $e_\alpha = n$.

Si $f : X \to \mathbb{R}$ est une quelconque fonction localement constante, on a

$$\begin{aligned} S_n f(y) - S_n f(x) &= \sum_{i=0}^{\alpha-1} f(\lhd^{e_i} y) - f(\lhd^{e_i} x) + \sum_{i=0}^{\alpha} \sum_{e_{i-1} < k < e_i} f(\lhd^k y) - f(\lhd^k x) \\ &\leqslant 2\alpha \|f\| + 2(\alpha + 1) \operatorname{bw}(f) \leqslant 2(\alpha+1)(\|f\| + \operatorname{bw}(f)) \\ &\leqslant (2\alpha_n(x,y) + 2) \cdot \operatorname{Nbwa}(f) \end{aligned}$$

Intégrons maintenant cette inégalité par la mesure ξ sur X^2. Au premier membre, on a

$$\langle S_n f \circ \partial_1 - S_n f \circ \partial_0, \xi \rangle = \langle S_n f, \partial_1 \xi - \partial_0 \xi \rangle = \langle S_n f, \mu_1 - \mu_0 \rangle = n \langle f, \mu_1 - \mu_0 \rangle$$

Pour le second membre, on observe que $2\alpha_n(x,y) = (E_0 + \cdots + E_{n-1})(x,y)$ et par conséquent $\langle 2\alpha_n, \xi \rangle \leqslant Cn$. On obtient donc

$$n \langle f, \mu_1 - \mu_0 \rangle \leqslant (Cn + 2) \cdot \operatorname{Nbwa}(f).$$

Cette inégalité étant valable pour tout $n \geqslant 1$, il en résulte que

$$\langle f, \mu_1 - \mu_0 \rangle \leqslant C \cdot \mathrm{Nbwa}(f)$$

pour toute fonction f localement constante; et donc $\|\mu_1 - \mu_0\|_{RR} \leqslant C$. □

3 Une propriété de dualité

Notons $\lll : X^2 \to X^2$ l'opérateur de décalage sur X^2, défini par $\lll(x, y) = (\lhd x, \lhd y)$, et soient $X \xleftarrow{\partial_0} X^2 \xrightarrow{\partial_1} X$ les projections naturelles. D'autre part, notons $\Delta_1 X$ le sous-ensemble de X^2 défini par

$$\Delta_1 X = X \times_{\pi_0} X = \{(x, y) \in X^2 : \pi_0 x = \pi_0 y\}$$

Un premier résultat de "dualité" relatif à la fonctionnelle de Bowen est le suivant:

Proposition 3.1. *Pour toute mesure positive ξ sur $\Delta_1 X$, et pour toute fonction mesurable $f : X \to \mathbb{R}$ vérifiant la condition de Bowen, on a l'inégalité*

$$\left|\langle f, \partial_1 \xi - \partial_0 \xi \rangle\right| \leqslant \mathrm{bw}(f) \cdot \|\lll \xi - \xi\|_{TV}. \tag{3.1}$$

Il en résulte immédiatement que, pour toute mesure ξ positive sur $\Delta_1 X$, on a

$$\|\partial_1 \xi - \partial_0 \xi\|_R \leqslant \|\lll \xi - \xi\|_{TV} \tag{3.2}$$

par définition de la norme de réarrangement; mais la proposition ci-dessus est plus forte, car elle établit l'inégalité (3.1) sans supposer f continue (il faut toutefois qu'elle soit mesurable, afin que l'intégrale au membre de gauche soit bien définie).

Démonstration. Quitte à remplacer f par $-f$, nous pouvons supposer que l'intégrale $\langle f, \partial_1 \xi - \partial_0 \xi \rangle$ est positive. Posons $C = \mathrm{bw}(f)$.
J'affirme qu'il existe une fonction mesurable $\beta : X^2 \to [0, 2C]$, vérifiant

$$\forall (x, y) \in \Delta_1 X \qquad f(y) - f(x) \leqslant \beta(\lhd x, \lhd y) - \beta(x, y). \tag{3.3}$$

Une telle fonction, si elle existe, doit vérifier en particulier

$$S_n f(y) - S_n f(x) \leqslant \beta(\lhd^n x, \lhd^n y) - \beta(x, y) \leqslant 2C - \beta(x, y)$$

pour tous $x, y \in X$ et $n \geqslant 0$ tels que $\pi_{<n}(x) = \pi_{<n}(y)$. Cela suggère de définir β par

$$\beta(x, y) \ = \ \operatorname*{Inf}_{\substack{n \in \mathbb{N} \text{ tels que} \\ \pi_{<n}(x) = \pi_{<n}(y)}} \quad 2C - S_n f(y) + S_n f(x)$$

et on vérifie sans peine qu'elle a bien les propriétés demandées.
Il en résulte que

$$\begin{aligned}
\langle f, \partial_1 \xi - \partial_0 \xi \rangle = \langle f \partial_1 - f \partial_0, \xi \rangle &= \int_{\Delta_1 X} [f(y) - f(x)]\, d\xi(x, y) \\
&\leqslant \int_{\Delta_1 X} [\beta(\lhd x, \lhd y) - \beta(x, y)]\, d\xi(x, y) \\
&= \langle \beta \lll - \beta, \xi \rangle = \langle \beta, \lll \xi - \xi \rangle = \langle \beta - C, \lll \xi - \xi \rangle \\
&\leqslant \|\beta - C\| \cdot \|\lll \xi - \xi\|_{TV} \leqslant C \cdot \|\lll \xi - \xi\|_{TV}
\end{aligned}$$

ce qu'il fallait démontrer. □

3.1 Enoncé du théorème

L'inégalité (3.2) peut être combinée avec l'encadrement (2.5) et l'inégalité triangulaire, de la façon suivante. Donnons-nous deux mesures positives μ_0, μ_1 sur X; pour toute mesure positive ξ sur $\Delta_1 X$ telle que $\partial_0 \xi \leqslant \mu_0$ et $\partial_1 \xi \leqslant \mu_1$, l'inégalité (3.2) entraîne

$$\|\mu_1 - \mu_0\|_R \leqslant \|\mu_0\| + \|\mu_1\| - 2\|\xi\| + \|\lll \xi - \xi\|_{TV} \tag{3.4}$$

et on peut se demander s'il existe ξ réalisant l'égalité dans l'inégalité ci-dessus.

Il s'avère que c'est faux en général. Toutefois, nous allons montrer que c'est vrai sous l'hypothèse additionnelle que les mesures μ_0, μ_1 sont *invariantes*. Le théorème ci-dessous donne un résultat plus précis.

Théorème 3.2. *Soient μ_0, μ_1 deux mesures positives invariantes sur X. Il existe ξ mesure positive sur $\Delta_1 X$ vérifiant $\partial_0 \xi \leqslant \mu_0$, $\partial_1 \xi \leqslant \mu_1$, $\pi_0 \partial_0 \xi = \pi_0 \partial_1 \xi = \pi_0 \mu_0 \wedge \pi_0 \mu_1$, et*

$$\|\mu_1 - \mu_0\|_R = \|\pi_0(\mu_1 - \mu_0)\| + \|\lll \xi - \xi\|_{TV}\,. \tag{3.5}$$

Un cas particulier important est celui où μ_0, μ_1 sont des mesures de même homologie, c.à.d. $\pi_0 \mu_0 = \pi_0 \mu_1$ (et donc de même masse); dans ce cas, le théorème affirme l'existence d'un *couplage* ξ de μ_0 et μ_1, porté par $\Delta_1 X$ et tel que

$$\|\mu_1 - \mu_0\|_R = \|\lll \xi - \xi\|_{TV} \tag{3.6}$$

c'est-à-dire réalisant l'égalité dans (3.2).

Dans le cas général, la condition $\pi_0 \partial_0 \xi = \pi_0 \partial_1 \xi = \pi_0 \mu_0 \wedge \pi_0 \mu_1$ veut dire que ξ est "aussi proche que possible" d'un couplage de μ_0 et μ_1. En effet, une mesure (positive) ξ portée par $\Delta_1 X$ vérifie toujours $\pi_0 \partial_0 \xi = \pi_0 \partial_1 \xi$, puisque $\pi_0 \partial_0$ et $\pi_0 \partial_1$ coïncident sur $\Delta_1 X$; et cette valeur commune, si $\partial_0 \xi \leqslant \mu_0$ et $\partial_1 \xi \leqslant \mu_1$, doit être majorée simultanément par $\pi_0 \mu_0$ et $\pi_0 \mu_1$, et donc par leur infimum. Dans le cas d'égalité $\pi_0 \partial_0 \xi = \pi_0 \partial_1 \xi = \pi_0 \mu_0 \wedge \pi_0 \mu_1$, on a

$$\|\mu_0\| + \|\mu_1\| - 2\|\xi\| = \|\pi_0 \mu_0\| + \|\pi_0 \mu_1\| - 2\|\pi_0 \mu_0 \wedge \pi_0 \mu_1\| = \|\pi_0 \mu_1 - \pi_0 \mu_0\|$$

et donc le ξ donné par le théorème réalise bien l'égalité dans (3.4).

Une formulation équivalente du théorème 3.2 est que, quelles que soient μ_0, μ_1 mesures positives invariantes sur X, on a

$$\|\mu_1 - \mu_0\|_R = \min_{\xi \in \mathcal{R}(\mu_0, \mu_1)} c(\xi) = \min_{\xi \in \mathcal{R}'(\mu_0, \mu_1)} c(\xi) \tag{3.7}$$

où on a posé

$$\mathcal{R}(\mu_0, \mu_1) = \left\{\xi \in \mathcal{M}^+(\Delta_1 X) : \pi_0 \xi \leqslant \mu_0,\ \pi_1 \xi \leqslant \mu_1\right\}$$
$$\mathcal{R}'(\mu_0, \mu_1) = \left\{\xi \in \mathcal{R}(\mu_0, \mu_1) : \pi_0 \partial_0 \xi = \pi_0 \partial_1 \xi = \pi_0 \mu_0 \wedge \pi_0 \mu_1\right\}$$

et

$$c(\xi) = \|\mu_0\| + \|\mu_1\| - 2\|\xi\| + \|\lll \xi - \xi\|_{TV}$$

est le "coût" du réarrangement ξ. Pour $\xi \in \mathcal{R}'(\mu_0, \mu_1)$ ce coût s'écrit plus simplement

$$c(\xi) = \|\pi_0(\mu_1 - \mu_0)\| + \|\lll \xi - \xi\|_{TV}\,.$$

En particulier, si μ_0, μ_1 sont à support fini, c'est-à-dire des combinaisons linéaires (positives) d'orbites périodiques, les convexes $\mathcal{R}(\mu_0, \mu_1)$ et $\mathcal{R}'(\mu_0, \mu_1)$ sont des polytopes de dimension finie, et la formule (3.7) ramène le calcul de $\|\mu_1 - \mu_0\|_R$ à la résolution d'un programme linéaire de taille finie (ce qu'on sait faire de manière assez efficace).

Ce théorème, analogue du théorème de Kantorovitch-Rubinstein [KR1, KR2] pour les réarrangements, sera démontré plus bas. Pour le moment, donnons-en un corollaire:

Proposition 3.3. *Pour toute mesure invariante μ sur X, et toute fonction mesurable $f : X \to \mathbb{R}$ vérifiant la condition de Bowen, on a*

$$|\langle f, \mu\rangle| \;\leqslant\; \mathrm{Nbw}(f) \cdot \|\mu\|_R \tag{3.8}$$

et si de plus $\pi_0\mu = 0$, alors

$$|\langle f, \mu\rangle| \;\leqslant\; \mathrm{bw}(f) \cdot \|\mu\|_R \tag{3.9}$$

Démonstration. D'après le théorème 3.2, il existe une mesure positive ξ sur $\Delta_1 X$ et des mesures positives ρ_0, ρ_1 sur X vérifiant $\mu^- = \partial_0\xi + \rho_0$, $\mu^+ = \partial_1\xi + \rho_1$, avec $\|\rho_0\| + \|\rho_1\| = \|\pi_0\mu\|$ et $\|\mu\|_R = \|\pi_0\mu\| + \|\lll \xi - \xi\|_{TV}$. Donc

$$\begin{aligned}
|\langle f, \mu\rangle| &= |\langle f, \partial_1\xi - \partial_0\xi + \rho_1 - \rho_0\rangle| \\
&\leqslant |\langle f, \partial_1\xi - \partial_0\xi\rangle| + |\langle f, \rho_1 - \rho_0\rangle| \\
&\leqslant \mathrm{bw}(f) \cdot \|\lll \xi - \xi\|_{TV} + \|f\| \cdot \|\rho_1 - \rho_0\|_{TV} \quad \text{(cf. prop. 3.1)} \\
&\leqslant \mathrm{bw}(f)\left[\|\mu\|_R - \|\pi_0\mu\|\right] + \|f\| \cdot \|\pi_0\mu\|
\end{aligned}$$

et la proposition en découle. □

Ces formules sont à rapprocher de (2.4) et (2.7); ici, la fonction f n'est pas supposée localement constante (ou continue) mais en contrepartie, la mesure μ doit être invariante. J'ignore si ces inégalités restent valables sans cette hypothèse d'invariance.

3.2 Résultats auxiliaires

Lemme 3.4. *Soit F un sous-ensemble invariant fini, non vide, de X, et $\phi : F \to \mathbb{R}$ une fonction telle que $\mathrm{bw}(\phi) < 1$. Il existe un réel $0 \leqslant \lambda < 1$ et une fonction $\beta : F^2 \to [0, 2]$ tels que*

$$\beta(x, x) = 0, \tag{3.10}$$

$$\beta(x, z) \leqslant \beta(x, y) + \beta(y, z), \tag{3.11}$$

$$\phi(y) - \phi(x) + \beta(x, y) \leqslant \lambda\beta(\lhd x, \lhd y) \quad \text{si } x_0 = y_0, \tag{3.12}$$

pour tous $x, y, z \in F$.

Démonstration. Puisque F est fini, on peut trouver un entier $R \geqslant 1$ tel que

$$\forall x, y \in F \qquad \pi_{<R}(x) = \pi_{<R}(y) \implies x = y.$$

Notons $C = \mathrm{bw}(\phi)$, et définissons $\lambda = C^{1/R}$.

Supposons qu'il existe une fonction β avec les propriétés requises. Pour tous $n \geqslant 0$ et $x, y \in F$ tels que $\pi_{<n}(x) = \pi_{<n}(y)$, on doit avoir

$$\begin{aligned}
\phi(y) - \phi(x) &\;\leqslant\; \lambda\beta(\lhd x, \lhd y) - \beta(x, y) \\
&\;\vdots \\
\lambda^{n-1}\left[\phi(\lhd^{n-1} y) - \phi(\lhd^{n-1} x)\right] &\;\leqslant\; \lambda^n\beta(\lhd^n x, \lhd^n y) - \lambda^{n-1}\beta(\lhd^{n-1} x, \lhd^{n-1} y)
\end{aligned}$$

et donc

$$\begin{aligned}
\sum_{0 \leqslant k < n} \lambda^k\left[\phi(\lhd^k y) - \phi(\lhd^k x)\right] &\;\leqslant\; \lambda^n\beta(\lhd^n x, \lhd^n y) - \beta(x, y) \\
&\;\leqslant\; 2\lambda^n - \beta(x, y)
\end{aligned}$$

Cela suggère de définir la fonction β par

$$\beta(x,y) \;=\; \operatorname*{Inf}_{\substack{n\in\mathbb{N}\text{ t.q.}\\ \pi_{<n}(x)=\pi_{<n}(y)}} \beta_n(x,y)$$

où les fonctions $\beta_n : F^2 \to \mathbb{R}$ sont définies par

$$\beta_n(x,y) \;=\; 2\lambda^n + \sum_{0\leqslant k<n} \lambda^k \big[\phi(\lhd^k x) - \phi(\lhd^k y)\big].$$

Il est clair que cette fonction satisfait (3.10) et (3.12), et qu'elle est majorée par 2 (puisque $\beta_0 = 2$). Il reste à montrer que β est positive et qu'elle vérifie l'inégalité triangulaire (3.11). Pour établir ces deux propriétés, nous allons d'abord prouver l'inégalité suivante: quels que soient $0 \leqslant m \leqslant n$ entiers et $x, y \in F$ tels que $\pi_{<n}(x) = \pi_{<n}(y)$, on a

$$\beta_n(x,y) + 2\lambda^m \;\geqslant\; \beta_m(x,y). \tag{3.13}$$

On supposera x, y distincts (sinon c'est trivial) et donc $n < R$. Alors,

$$\begin{aligned}\beta_n(x,y) + 2\lambda^m - \beta_m(x,y) \;&=\; 2\lambda^n + \sum_{m\leqslant k<n} \lambda^k \big[\phi(\lhd^k x) - \phi(\lhd^k y)\big]\\ &=\; 2\lambda^n - \lambda^m \sum_{0\leqslant i<n-m} \lambda^i \big[\phi(\lhd^i y') - \phi(\lhd^i x')\big]\end{aligned}$$

où on a posé $x' = \lhd^m x$ et $y' = \lhd^m y$ (qui ont les mêmes $n - m$ premiers symboles). Mais, par l'inégalité d'Abel,

$$\sum_{0\leqslant i<n-m} \lambda^i \big[\phi(\lhd^i y') - \phi(\lhd^i x')\big] \;\leqslant\; \operatorname*{Sup}_{0\leqslant k\leqslant n-m} \sum_{0\leqslant i<k} \phi(\lhd^i y') - \phi(\lhd^i x') \;\leqslant\; 2C$$

et donc

$$\beta_n(x,y) + 2\lambda^m - \beta_m(x,y) \;\geqslant\; 2\lambda^m(\lambda^{n-m} - C)$$

et le membre de droite est positif, car $\lambda^{n-m} \geqslant \lambda^R = C$, et l'inégalité (3.13) est établie.

En faisant $m = 0$ dans cette inégalité, on obtient $\beta_n(x,y) \geqslant 0$ pour tout n tel que $\pi_{<n}(x) = \pi_{<n}(y)$, ce qui revient à dire que $\beta(x,y) \geqslant 0$.

Montrons enfin l'inégalité triangulaire (3.11). On peut supposer x, y, z distincts, sinon c'est trivial. Il existe alors des entiers naturels n, m tels que $\beta(x,y) = \beta_n(x,y)$ avec $\pi_{<n}(x) = \pi_{<n}(y)$, et $\beta(y,z) = \beta_m(y,z)$ avec $\pi_{<m}(y) = \pi_{<m}(z)$. Je supposerai que n est le plus grand des deux; l'autre cas se traiterait identiquement. On note d'abord que $\pi_{<m}(x) = \pi_{<m}(z)$, et donc

$$\begin{aligned}\beta(x,z) &\leqslant \beta_m(x,z)\\ &= -2\lambda^m + \beta_m(x,y) + \beta_m(y,z)\\ &\leqslant \beta_n(x,y) + \beta_m(y,z)\\ &= \beta(x,y) + \beta(y,z)\end{aligned}$$

ce qu'il fallait démontrer. □

Lemme 3.5. *Soit F un sous-ensemble invariant fini, non vide, de X, et $\phi : F \to \mathbb{R}$ une fonction telle que* $\mathrm{Nbw}(\phi) < 1$. *Il existe une fonction continue $\Phi : X \to \mathbb{R}$ telle que* $\mathrm{Nbw}(\Phi) \leqslant 1$ *et qui prolonge ϕ.*

Démonstration. Notons

$$\overline{F} = F \sqcup \{\perp\}$$

l'ensemble obtenu en adjoignant à F un élément spécial, $\perp$. La preuve repose sur une représentation de ϕ par un système itéré de fonctions "topicales" de $\mathbb{R}^{\overline{F}}$ dans lui-même, indicé par les éléments de A; on pourra se référer à [BM] pour les principaux concepts et notations.

Soit $\mathcal{T}$ le sous-ensemble de $\mathbb{R}^{\overline{F}}$ constitué des points $\mathbf{x} = (x_i)_{i\in\overline{F}}$ vérifiant

$$\forall i \in F \quad x_\perp \leqslant x_i \leqslant x_\perp + 2.$$

Dans la terminologie de [BM], $\mathcal{T}$ est un "tube" dans $\mathbb{R}^{\overline{F}}$: il est non vide, invariant par les translations de $\mathbb{R}\vec{\mathbf{u}}$, et le quotient $\tilde{\mathcal{T}} = \mathcal{T}/\mathbb{R}\vec{\mathbf{u}}$ est compact. En outre, c'est un sous-treillis de $\mathbb{R}^{\overline{F}}$, c'est-à-dire qu'il est stable par les opérations $\vee$ et $\wedge$.

Notons $h : \mathbb{R}^{\overline{F}} \to \mathbb{R}$ la fonction coordonnée $\mathbf{x} \mapsto x_\perp$, et soit

$$\begin{aligned}\mathcal{T}_0 &= \{\mathbf{x} \in \mathcal{T} : h(\mathbf{x}) = 0\}\\ &= \{\mathbf{x} \in \mathbb{R}^{\overline{F}} : x_\perp = 0,\ 0 \leqslant x_i \leqslant 2 \ \ \forall i \in F\}\end{aligned}$$

(qui est compact, et homéomorphe à $\tilde{\mathcal{T}}$).

Soient λ et β comme dans le lemme 3.4, et considérons la fonction $B : F \to \mathbb{R}^{\overline{F}}$ dont les fonctions coordonnées sont données par

$$B_\perp(x) = 0, \quad B_i(x) = \beta(i,x) \tag{3.14}$$

pour tous $x, i \in F$. Il est manifeste que B prend ses valeurs dans $\mathcal{T}_0$, et on vérifie facilement que

$$\forall x, y \in F \qquad \lceil B(x), B(y)\rceil = \beta(x,y). \tag{3.15}$$

Considérons maintenant les fonctions topicales U et V_y (pour y décrivant F) de $\mathbb{R}^{\overline{F}}$ dans lui-même, définies par

$$\begin{aligned}U(\mathbf{p}) &= (p_\perp - 1)\vec{\mathbf{u}},\\ V_y(\mathbf{p}) &= \big[\lambda\lfloor B(\lhd y), \mathbf{p}\rfloor + (1-\lambda)p_\perp + \phi(y)\big]\vec{\mathbf{u}} + B(y).\end{aligned}$$

On a évidemment

$$\forall y \in F \qquad V_y\big[B(\lhd y)\big] = \phi(y)\vec{\mathbf{u}} + B(y).$$

Définissons maintenant les fonctions topicales T_a (pour a décrivant A) de $\mathbb{R}^{\overline{F}}$ dans lui-même, par

$$T_a(\mathbf{p}) \ = \ U(\mathbf{p}) \vee \operatorname*{Sup}_{\substack{y\in F\\ y_0 = a}} V_y(\mathbf{p}) \tag{3.16}$$

(en particulier $T_a = U$ si aucun mot de F ne commence par a). Manifestement les fonctions U et V_y envoient le tube $\mathcal{T}$ dans lui-même, et donc il en est de même pour les fonctions T_a (puisque $\mathcal{T}$ est stable par les opérations du treillis). J'affirme que ces fonctions satisfont

$$\forall a \in A \quad \forall x \in F \qquad T_a\big[B(\lhd x)\big] = \phi(x)\vec{\mathbf{u}} + B(x) \quad \text{si } x_0 = a. \tag{3.17}$$

L'inégalité $\geqslant$ est évidente, car $T_a \geqslant V_x$. Dans l'autre sens, il faut démontrer que $\phi(x)\vec{\mathbf{u}} + B(x)$ majore $U\big[B(\lhd x)\big]$ ainsi que $V_y\big[B(\lhd x)\big]$ pour tous les y tels que $y_0 = a$. Premièrement, on a

$U\big[B(\lhd x)\big] = -\vec{\mathbf{u}}$, qui est bien $\leqslant \phi(x)\vec{\mathbf{u}} + B(x)$, car $\phi(x) \geqslant -1$ et $B(x) \geqslant 0$. Deuxièmement, pour tout $y \in F$ tel que $y_0 = a$, on a

$$\begin{aligned} V_y\big[B(\lhd x)\big] &= \big[\lambda\lfloor B(\lhd y), B(\lhd x)\rfloor + \phi(y)\big]\vec{\mathbf{u}} + B(y) \\ &= \big[-\lambda\beta(\lhd x, \lhd y) + \phi(y)\big]\vec{\mathbf{u}} + B(y) \\ &\leqslant \big[-\beta(x,y) + \phi(x)\big]\vec{\mathbf{u}} + B(y) \\ &= \big[-\lceil B(x), B(y)\rceil + \phi(x)\big]\vec{\mathbf{u}} + B(y) \\ &\leqslant \phi(x)\vec{\mathbf{u}} + B(x), \end{aligned}$$

ce qu'il fallait démontrer.
D'autre part, j'affirme que les T_a vérifient

$$\forall a \in A \quad \forall \mathbf{p} \in \mathbb{R}^{\overline{F}} \qquad h(\mathbf{p}) - 1 \ \leqslant\ h(T_a\mathbf{p}) \ \leqslant\ h(\mathbf{p}) + 1 \tag{3.18}$$

La première inégalité est évidente, car $T_a \geqslant U$. Pour la deuxième inégalité, notons que pour tout $y \in F$,

$$\begin{aligned} hV_y(\mathbf{p}) &= \lambda\lfloor B(\lhd y), \mathbf{p}\rfloor + (1-\lambda)h(\mathbf{p}) + \phi(y) + hB(y) \\ &\leqslant \lambda\big[h(\mathbf{p}) - hB(\lhd y)\big] + (1-\lambda)h(\mathbf{p}) + \phi(y) \\ &= h(\mathbf{p}) + \phi(y) \\ &\leqslant h(\mathbf{p}) + 1 \end{aligned}$$

car $\phi(y) \leqslant 1$, et les T_a vérifieront donc l'inégalité analogue.
La dernière étape de la preuve consiste à construire une fonction $E : X \to \mathcal{T}_0$ qui prolonge B, et une fonction $\Phi : X \to \mathbb{R}$ qui prolonge ϕ, et telles que

$$\forall a \in A \quad \forall x \in X \qquad T_a\big[E(\lhd x)\big] = \Phi(x)\vec{\mathbf{u}} + E(x) \quad \text{si } x_0 = a. \tag{3.19}$$

Si une telle fonction E existe, la fonction quotient $\tilde{E} : X \to \tilde{\mathcal{T}}$ doit vérifier $\tilde{T}_a : \tilde{E}(x) \mapsto \tilde{E}(ax)$ pour tous $x \in X$, $a \in A$. Or, je me suis arrangé pour que les fonctions $\tilde{T}_a$ soient contractantes sur $\tilde{\mathcal{T}}$, pour la distance $\tilde{d}$ définie par $\tilde{d}(\mathbf{p},\mathbf{q}) = \frac{1}{2}\big(\lceil\mathbf{p},\mathbf{q}\rceil + \lceil\mathbf{q},\mathbf{p}\rceil\big)$. En effet, les fonctions V_y vérifient

$$\lceil V_y(\mathbf{p}), V_y(\mathbf{q})\rceil \leqslant \lambda\lceil\mathbf{p},\mathbf{q}\rceil + (1-\lambda)(q_\perp - p_\perp)$$

et par ailleurs

$$\lceil U(\mathbf{p}), U(\mathbf{q})\rceil = q_\perp - p_\perp \leqslant \lambda\lceil\mathbf{p},\mathbf{q}\rceil + (1-\lambda)(q_\perp - p_\perp)$$

donc les T_a vérifient de même

$$\forall a \in A \quad \forall \mathbf{p},\mathbf{q} \in \mathbb{R}^{\overline{F}} \qquad \lceil T_a(\mathbf{p}), T_a(\mathbf{q})\rceil \leqslant \lambda\lceil\mathbf{p},\mathbf{q}\rceil + (1-\lambda)(q_\perp - p_\perp)$$

et par conséquent

$$\forall a \in A \quad \forall \mathbf{p},\mathbf{q} \in \mathbb{R}^{\overline{F}} \qquad \tilde{d}\big[T_a(\mathbf{p}), T_a(\mathbf{q})\big] \leqslant \lambda\, \tilde{d}(\mathbf{p},\mathbf{q}). \tag{3.20}$$

Soit $\mathcal{O}$ l'application λ-lipschitzienne de $C(X, \tilde{\mathcal{T}})$ dans lui-même, définie par

$$(\mathcal{O}\eta)(x) = \tilde{T}_{x_0}\big[\eta(\lhd x)\big] \tag{3.21}$$

et soit $\mathcal{K}$ le sous-ensemble fermé (non vide) de $C(X, \tilde{\mathcal{T}})$ constitué des fonctions qui prolongent $\tilde{B}$. Il est stable par $\mathcal{O}$ en vertu de (3.17), et contient donc un point fixe que je noterai $\tilde{E}$. Evidemment, son relèvement $E : X \to \mathcal{T}_0$ est continu et prolonge B. L'invariance de $\tilde{E}$ par

$\mathcal{O}$ signifie qu'il existe une fonction continue $\Phi : X \to \mathbb{R}$ vérifiant (3.19). En rapprochant cette équation de (3.17), on voit que Φ est un prolongement de ϕ. En écrivant

$$\Phi(x) = h\,T_{x_0}E(\lhd x) \tag{3.22}$$

et en comparant cette formule avec (3.18), on voit que $|\Phi| \leqslant 1$. Plus généralement, on a

$$\Phi(x) + \Phi(\lhd x) + \cdots + \Phi(\lhd^{n-1}x) = h\,T_{x_0}T_{x_1}\cdots T_{x_{n-1}}E(\lhd^n x)$$

ce qui a la conséquence suivante: si x, y sont deux éléments de X qui commencent par les mêmes n lettres $a_0, \ldots, a_{n-1}$, on peut écrire $S_n\Phi(x) = hW(\mathbf{p})$ et $S_n\Phi(y) = hW(\mathbf{q})$, avec $W = T_{a_0}\cdots T_{a_{n-1}}$, $\mathbf{p} = E(\lhd^n x)$ et $\mathbf{q} = E(\lhd^n y)$, donc

$$S_n\Phi(y) - S_n\Phi(x) = hW(\mathbf{q}) - hW(\mathbf{p}) \leqslant \lceil W(\mathbf{p}), W(\mathbf{q})\rceil \leqslant \lceil \mathbf{p}, \mathbf{q}\rceil \leqslant 2$$

car W est topicale et $\mathbf{p}, \mathbf{q} \in \mathcal{T}_0$.

Ceci montre que $\mathrm{bw}(\Phi) \leqslant 1$, et termine la preuve du lemme. □

Lemme 3.6. *Soit F un sous-ensemble fini invariant, non vide, de X, un réel $\alpha \geqslant 0$, et $\phi : F \to \mathbb{R}$ une fonction telle que, pour tous $x, y \in F$ et $n \geqslant 0$ vérifiant $\pi_{<n}(x) = \pi_{<n}(y)$,*

$$|S_n\phi(y) - S_n\phi(x)| \;\leqslant\; 2(1 + \alpha n). \tag{3.23}$$

Il existe alors une fonction $\beta : F^2 \to [0, 2]$ vérifiant

$$\beta(x, x) = 0 \tag{3.24}$$
$$\beta(x, z) \leqslant \beta(x, y) + \beta(y, z) \tag{3.25}$$
$$\phi(y) - \phi(x) + \beta(x, y) \leqslant 2\alpha + \beta(\lhd x, \lhd y) \quad si\ x_0 = y_0 \tag{3.26}$$

pour tous $x, y, z \in F$.

Démonstration. S'il existe une fonction β avec les propriétés requises, pour tous $n \geqslant 0$ et $x, y \in F$ tels que $\pi_{<n}(x) = \pi_{<n}(y)$, on doit avoir

$$\begin{aligned} \phi(y) - \phi(x) &\leqslant 2\alpha + \beta(\lhd x, \lhd y) - \beta(x, y) \\ &\vdots \\ \phi(\lhd^{n-1}y) - \phi(\lhd^{n-1}x) &\leqslant 2\alpha + \beta(\lhd^n x, \lhd^n y) - \beta(\lhd^{n-1}x, \lhd^{n-1}y) \end{aligned}$$

et donc

$$\begin{aligned} S_n\phi(y) - S_n\phi(x) &\leqslant 2\alpha n + \beta(\lhd^n x, \lhd^n y) - \beta(x, y) \\ &\leqslant 2(1 + \alpha n) - \beta(x, y) \end{aligned}$$

Cela suggère de définir la fonction β par

$$\beta(x, y) \;=\; \operatorname*{Inf}_{\substack{n\in\mathbb{N}\text{ t.q.}\\ \pi_{<n}(x)=\pi_{<n}(y)}} \beta_n(x, y)$$

pour tous x, y distincts (et 0 sinon), où les fonctions $\beta_n : F^2 \to \mathbb{R}$ sont définies par

$$\beta_n(x, y) \;=\; 2(1 + \alpha n) - S_n\phi(y) + S_n\phi(x).$$

Il est évident que cette fonction vérifie $0 \leqslant \beta \leqslant 2$ ainsi que les conditions (3.24) et (3.26). Seule l'inégalité triangulaire (3.25) requiert une démonstration.

Comme précédemment, nous l'obtiendrons comme conséquence d'une autre inégalité: quels que soient $0 \leqslant m \leqslant n$ entiers et $x, y \in F$ tels que $\pi_{<n}(x) = \pi_{<n}(y)$, on a

$$\beta_n(x,y) + 2 \ \leqslant \ \beta_m(x,y). \tag{3.27}$$

En effet, on a

$$\begin{aligned}\beta_n(x,y) + 2 - \beta_m(x,y) &= 2 + 2\alpha(n-m) + \sum_{m \leqslant k < n} \phi(\lhd^k x) - \phi(\lhd^k y) \\ &= 2 + 2\alpha(n-m) - \sum_{0 \leqslant i < n-m} \phi(\lhd^i y') - \phi(\lhd^i x')\end{aligned}$$

où on a posé $x' = \lhd^m x$ et $y' = \lhd^m y$ (qui ont les mêmes $n-m$ premiers symboles), et cette dernière expression est visiblement positive.

De cette inégalité, on déduit l'inégalité triangulaire (3.25) par un argument similaire à celui donné dans la preuve du lemme 3.4. $\square$

Lemme 3.7. *Soit F un sous-ensemble fini invariant, non vide, de X, un réel $\alpha \geqslant 0$, et $\phi : F \to \mathbb{R}$ une fonction telle que, pour tous $x, y \in F$ et $n \geqslant 0$ vérifiant $\pi_{<n}(x) = \pi_{<n}(y)$,*

$$|S_n\phi(y) - S_n\phi(x)| \ \leqslant \ 2(1 + \alpha n), \tag{3.28}$$

et d'autre part $\|\phi\| \leqslant 1 + \alpha$.

Il existe alors des fonctions $\Phi, g, h : F \to \mathbb{R}$ telles que

$$\phi \ = \ \Phi + g + h\lhd - h \tag{3.29}$$

avec $\mathrm{Nbw}(\Phi) \leqslant 1$ *et* $\|g\| \leqslant \alpha$.

Démonstration. Soient $\overline{F}, \mathcal{T}, \mathcal{T}_0$ et h comme dans la preuve du lemme 3.5. A partir de la fonction β donnée par le lemme 3.6, définissons la fonction $B : F \to \mathcal{T}_0$ par les formules (3.14); ici encore, elle vérifie l'identité (3.15).

Soient U et V_y (pour y décrivant F) les fonctions topicales de $\mathbb{R}^{\overline{F}}$ dans lui-même définies par

$$\begin{aligned}&U(\mathbf{p}) = (p_\perp - 1)\vec{\mathbf{u}} \\ &V_y(\mathbf{p}) = \big[\, \lfloor B(\lhd y), \mathbf{p} \rfloor - \alpha + \phi(y) \big]\vec{\mathbf{u}} + B(y).\end{aligned}$$

On a évidemment

$$\forall y \in F \qquad V_y\big[B(\lhd y)\big] = [\phi(y) - \alpha]\vec{\mathbf{u}} + B(y).$$

Définissons maintenant les fonctions topicales T_a (pour a décrivant A) de $\mathbb{R}^{\overline{F}}$ dans lui-même par la formule (3.16). Tout comme U et les V_y, ces fonctions envoient le tube $\mathcal{T}$ dans lui-même. J'affirme que

$$\begin{aligned}&\forall a \in A \quad \forall x \in F \\ &\qquad [\phi(x) - \alpha]\vec{\mathbf{u}} + B(x) \leqslant T_a\big[B(\lhd x)\big] \leqslant [\phi(x) + \alpha]\vec{\mathbf{u}} + B(x) \quad \text{si } x_0 = a.\end{aligned} \tag{3.30}$$

La première inégalité est évidente, car $T_a \geqslant V_x$. Pour la deuxième inégalité, il faut montrer que $[\phi(x) + \alpha]\vec{\mathbf{u}} + B(x)$ majore $U\big[B(\lhd x)\big]$ ainsi que $V_y\big[B(\lhd x)\big]$ pour tout y tel que $y_0 = a$. Premièrement, on a $U\big[B(\lhd x)\big] = -\vec{\mathbf{u}}$, qui est bien $\leqslant [\phi(x) + \alpha]\vec{\mathbf{u}} + B(x)$, car $\phi(x) \geqslant -1 - \alpha$ et $B(x) \geqslant 0$. Deuxièmement, pour tout y tel que $y_0 = a$, on a

$$\begin{aligned}V_y\big[B(\lhd x)\big] &= \big[\, \lfloor B(\lhd y), B(\lhd x) \rfloor - \alpha + \phi(y) \big]\vec{\mathbf{u}} + B(y) \\ &= \big[-\beta(\lhd x, \lhd y) - \alpha + \phi(y)\big]\vec{\mathbf{u}} + B(y) \\ &\leqslant \big[-\beta(x,y) + \alpha + \phi(x)\big]\vec{\mathbf{u}} + B(y) \\ &= \big[-\lceil B(x), B(y) \rceil + \alpha + \phi(x)\big]\vec{\mathbf{u}} + B(y) \\ &\leqslant [\phi(x) + \alpha]\vec{\mathbf{u}} + B(x)\end{aligned}$$

ce qu'il fallait démontrer.

J'affirme d'autre part que les T_a vérifient l'encadrement (3.18). La première inégalité est évidente, car $T_a \geqslant U$. Pour la deuxième inégalité, notons que pour tout $y \in F$,

$$\begin{aligned} hV_y(\mathbf{p}) &= \lfloor B(\lhd y), \mathbf{p} \rfloor + \phi(y) - \alpha + hB(y) \\ &\leqslant h(\mathbf{p}) - hB(\lhd y) + \phi(y) - \alpha \\ &= h(\mathbf{p}) + \phi(y) - \alpha \\ &\leqslant h(\mathbf{p}) + 1 \end{aligned}$$

car $\phi(y) \leqslant 1 + \alpha$, et les T_a vérifieront donc l'inégalité analogue.

La dernière étape de la preuve consiste à construire des fonctions $E : F \to \mathcal{T}_0$ et $\Phi : X \to \mathbb{R}$ telles que

$$\forall a \in A \quad \forall x \in X \qquad T_a\big[E(\lhd x)\big] = \Phi(x)\vec{\mathbf{u}} + E(x) \quad \text{si } x_0 = a. \tag{3.31}$$

Si elles existent, la fonction quotient $\tilde{E} : F \to \tilde{\mathcal{T}}$ doit vérifier $\tilde{T}_{x_0} : \tilde{E}(\lhd x) \mapsto \tilde{E}(x)$ pour tout x. Cela équivaut à dire que $\tilde{E}$ est un point fixe de l'opérateur $\mathcal{O}$ de $C(F, \tilde{\mathcal{T}})$ dans lui-même, défini par la formule (3.21). Mais $C(F, \tilde{\mathcal{T}})$ est homéomorphe à $\mathcal{T}_0^F \simeq [0,2]^{F^2}$, qui est une boule, et donc $\mathcal{O}$ admet au moins un point fixe, par le théorème de Brouwer. Soit $\tilde{E}$ un tel point fixe, et $E : F \to \mathcal{T}_0$ son relèvement. Il existe alors une et une seule fonction $\Phi : X \to \mathbb{R}$ vérifiant (3.31), donnée comme précedemment par la formule (3.22). De là on établit, comme dans la preuve du lemme 3.5, que $|\Phi| \leqslant 1$ et $\mathrm{bw}(\Phi) \leqslant 1$.

Il reste à comprendre maintenant le lien entre ϕ et Φ. Soit x un élément quelconque de F, et n sa période pour l'application de décalage. Des formules (3.31) et (3.30) il resulte qu'on a d'une part

$$W\big[E(x)\big] = S_n\Phi(x)\vec{\mathbf{u}} + E(x)$$

où $W = T_{x_0}T_{x_1}\cdots T_{x_{n-1}}$, et d'autre part

$$[S_n\phi(x) - \alpha n]\vec{\mathbf{u}} + B(x) \leqslant W\big[B(x)\big] \leqslant [S_n\phi(x) + \alpha n]\vec{\mathbf{u}} + B(x).$$

Il en résulte que

$$\begin{aligned} \lfloor E(x), B(x) \rfloor &\leqslant \lfloor WE(x), WB(x) \rfloor \\ &\leqslant \lfloor S_n\Phi(x)\vec{\mathbf{u}} + E(x), [S_n\phi(x) + \alpha n]\vec{\mathbf{u}} + B(x) \rfloor \\ &= \lfloor E(x), B(x) \rfloor + S_n\phi(x) - S_n\Phi(x) + \alpha n \end{aligned}$$

et d'autre part

$$\begin{aligned} \lfloor B(x), E(x) \rfloor &\leqslant \lfloor WB(x), WE(x) \rfloor \\ &\leqslant \lfloor [S_n\phi(x) - \alpha n]\vec{\mathbf{u}} + B(x), S_n\Phi(x)\vec{\mathbf{u}} + E(x) \rfloor \\ &= \lfloor B(x), E(x) \rfloor + S_n\Phi(x) - S_n\phi(x) + \alpha n \end{aligned}$$

ce qui montre que

$$|S_n\phi(x) - S_n\Phi(x)| \;\leqslant\; \alpha n$$

autrement dit, sur chaque orbite de F, les moyennes de ϕ et Φ diffèrent d'au plus α. Comme on le sait bien, cela revient à dire que la fonction $\phi - \Phi$ est cohomologue (sur F) à une fonction g telle que $|g| \leqslant \alpha$, ce qui termine la preuve du lemme. $\square$

3.3 Démonstration du théorème

Nous allons en fait démontrer l'énoncé suivant, qui est *équivalent* au théorème 3.2, comme le lecteur le vérifiera.

Proposition 3.8. *Pour toute mesure invariante μ sur X, on peut écrire*

$$\mu = \partial_1\xi - \partial_0\xi + \rho$$

où ξ est une mesure positive sur $\Delta_1 X$ et ρ une mesure sur X, telles que

$$\begin{aligned} \|\mu\|_{TV} &= 2\,\|\xi\|_{TV} + \|\rho\|_{TV} \\ \|\mu\|_R &= \|\lll\xi - \xi\|_{TV} + \|\rho\|_{TV} \end{aligned}$$

Démonstration. Pour des raisons de densité, on voit qu'il suffit de démontrer la proposition dans le cas où μ est à support fini. On supposera donc que μ est portée par un certain ensemble fini F, invariant et non vide.

Par le lemme 3.5 (combiné avec le lemme 2.3), la norme de réarrangement de μ peut s'écrire

$$\|\mu\|_R \;=\; \operatorname*{Sup}_{\substack{f:F\to\mathbb{R}\\ \mathrm{Nbw}(f)\leqslant 1}} \langle f,\mu\rangle \tag{3.32}$$

Posons $C = \|\mu\|_{TV} + \|\mu\|_R$, et soit $\mathcal{K}$ le sous-ensemble de $\mathcal{M}(F)$ défini par

$$\begin{aligned} \mathcal{K} = \big\{\partial_1\xi - \partial_0\xi + \rho : \xi \in \mathcal{M}^+(\Delta_1 F), \rho \in \mathcal{M}(F), \\ 2\,\|\xi\|_{TV} + \|\lll\xi - \xi\|_{TV} + 2\,\|\rho\|_{TV} \leqslant C\big\} \end{aligned} \tag{3.33}$$

Il est évident que $\mathcal{K}$ est compact, convexe, et contient 0.

J'affirme que $\mathcal{K}$ contient μ. Raisonnons par l'absurde, en supposant que ce n'est pas le cas. Il existerait alors une forme linéaire ϕ sur $\mathcal{M}(F)$, c'est-à-dire une fonction $\phi : F \to \mathbb{R}$, et une constante réelle c telles que

$$\forall \nu \in \mathcal{K} \qquad \langle \phi, \nu\rangle \leqslant c \tag{3.34}$$

$$\langle \phi, \mu\rangle > c \tag{3.35}$$

et bien sûr, $C > 0$ puisque $\mu \neq 0$.

Comme $\mathcal{K}$ contient toutes les mesures ρ telles que $\|\rho\|_{TV} \leqslant C/2$, il vient

$$\forall \rho \in \mathcal{M}(F) \qquad \|\rho\|_{TV} \leqslant C/2 \implies \langle \phi, \nu\rangle \leqslant c$$

ce qui équivaut à dire que

$$\|\phi\| \leqslant 2c/C. \tag{3.36}$$

D'autre part, pour tous $x, y \in F$ et $n \geqslant 0$ tels que $\pi_{<n}(x) = \pi_{<n}(y)$, la mesure

$$\xi = \frac{C}{2(1+n)} \sum_{0\leqslant i<n} \delta_{(\lhd^i x, \lhd^i y)}$$

est une mesure positive sur $\Delta_1 F$, avec $\|\xi\|_{TV} = Cn/2(1+n)$ et $\|\lll\xi - \xi\|_{TV} \leqslant C/(1+n)$, donc

$$2\,\|\xi\|_{TV} + \|\lll\xi - \xi\|_{TV} \leqslant C$$

et $\nu = \partial_1\xi - \partial_0\xi$ est dans $\mathcal{K}$. Par (3.34), on a

$$\sum_{0 \leqslant i < n} \phi(\lhd^i y) - \phi(\lhd^i x) \leqslant \frac{2c}{C}(1+n) \tag{3.37}$$

et cette inégalité est valable quels que soient $x, y \in F$ et $n \geqslant 0$ tels que $\pi_{<n}(x) = \pi_{<n}(y)$.
D'après le lemme 3.7, les inégalités (3.36) et (3.37) entraînent l'existence de trois fonctions $f, g, h : F \to \mathbb{R}$ telles que $\phi = f + g + h\lhd - h$, avec $\mathrm{Nbw}(f) \leqslant c/C$ et $\|g\| \leqslant c/C$. Par conséquent,

$$\begin{aligned}\langle \phi, \mu\rangle &= \langle f, \mu\rangle + \langle g, \mu\rangle + \langle h\lhd - h, \mu\rangle \\ &\leqslant \mathrm{Nbw}(f)\cdot \|\mu\|_R + \|g\|\cdot\|\mu\|_{TV} + \|h\|\cdot\|\lhd\mu - \mu\|_{TV} \\ &\leqslant \frac{c}{C}\|\mu\|_R + \frac{c}{C}\|\mu\|_{TV} + 0 = c\end{aligned}$$

ce qui contredit (3.35).
Nous avons donc établi que $\mu \in \mathcal{K}$, c'est-à-dire qu'il existe une mesure positive ξ sur $\Delta_1 F$ et une mesure ρ sur F telles que $\mu = \partial_1\xi - \partial_0\xi + \rho$, avec

$$2\|\xi\|_{TV} + \|\lll\xi - \xi\|_{TV} + 2\|\rho\|_{TV} \leqslant \|\mu\|_R + \|\mu\|_{TV}$$

et en combinant ceci avec les deux inégalités évidentes

$$\begin{aligned}\|\mu\|_{TV} &\leqslant 2\|\xi\|_{TV} + \|\rho\|_{TV} \\ \|\mu\|_R &\leqslant \|\lll\xi - \xi\|_{TV} + \|\rho\|_{TV}\end{aligned}$$

on voit que les trois inégalités ci-dessus sont en réalité des *égalités*, ce qui termine la preuve de la proposition. □

3.4 Couplages et réarrangements

Considérons un cas particulier du théorème de dualité, où μ_0, μ_1 sont deux orbites périodiques distinctes, vues comme mesures de probabilité invariantes, de même homologie (i.e. $\pi_0\mu_0 = \pi_0\mu_1$), avec la propriété que chaque 1-cylindre de X contient exactement un point de chacune des deux orbites. En notant L le cardinal de A, on peut écrire $\mu_0 = L^{-1}\sum_{a\in A}\delta_{\Phi_0(a)}$, où $\Phi_0 : A \to X$ est définie par

$$\Phi_0(a) = \big(a, S_0(a), S_0^2(a), \ldots\big)$$

et S_0 une permutation circulaire de A; et de même $\mu_1 = L^{-1}\sum_{a\in A}\delta_{\Phi_1(a)}$, avec

$$\Phi_1(a) = \big(a, S_1(a), S_1^2(a), \ldots\big)$$

et S_1 une autre permutation circulaire de A. Les mesures μ_0, μ_1 n'ayant qu'un seul atome dans chaque 1-cylindre de X, elles ont un *unique* couplage porté par $\Delta_1 X$, à savoir

$$\xi = \frac{1}{L}\sum_{a\in A}\delta_{(\Phi_0(a), \Phi_1(a))}$$

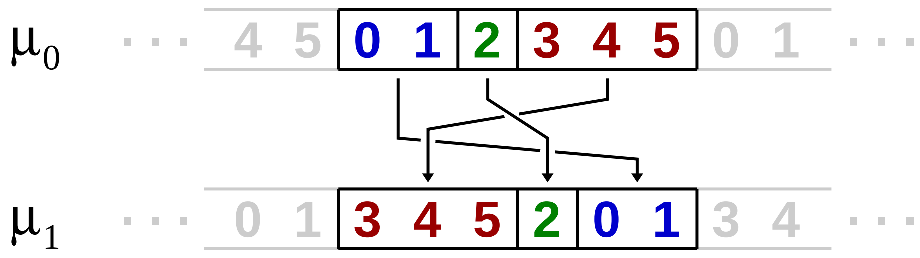
dont on calcule facilement le défaut d'invariance: on obtient

$$\|\mu_1 - \mu_0\|_R = \|\lll\xi - \xi\|_{TV} = \frac{2k}{L}$$

où k est le nombre de lettres $a \in A$ telles que $S_0(a) \neq S_1(a)$. Cet entier peut prendre n'importe quelle valeur entre 2 et L.

Si on considère les orbites μ_0, μ_1 comme des "génomes circulaires", ce nombre k est exactement le nombre de "breakpoints" entre ces deux génomes, c'est-à-dire le nombre de couples (a, b) de "gènes" (i.e., de lettres) qui sont consécutifs dans l'un des génomes et pas dans l'autre; voir par exemple [F+], sections 2.6.1, 9.1.1 et 9.1.4.

Une autre manière de voir ce nombre k, plus dynamique, est la suivante. Considérons le premier "génome", c'est-à-dire le graphe de S_0, qui est constitué d'un unique cycle, et découpons-le en ses breakpoints, autrement dit, enlevons les arêtes qui ne sont pas dans le graphe de S_1. Comme on a enlevé k arêtes, avec $k \geqslant 2$, on se retrouve avec k composantes connexes, qui sont des intervalles (éventuellement réduits à des singletons). Ces intervalles peuvent ensuite être "recollés" pour obtenir le second génome (i.e. le graphe de S_1), en ajoutant les k arêtes du graphe de S_1 qui n'étaient pas déjà présentes dans le graphe de S_0. D'autre part, il est bien clair que k est le nombre minimum de coupures (ou, ce qui revient au même, de fragments) nécessaires pour un tel réarrangement.



Par exemple, prenons $A = \{0, 1, 2, 3, 4, 5\}$ et soient μ_0, μ_1 les orbites 6-périodiques correspondant aux permutations circulaires $S_0 = (0, 1, 2, 3, 4, 5)$ et $S_1 = (3, 4, 5, 2, 0, 1)$. Il y a 3 breakpoints (les lettres $1, 2, 5$ ont des images distinctes par S_0 et S_1), et donc 3 fragments qu'on peut réarranger comme sur la figure, et la distance de réarrangement vaut

$$\|\mu_1 - \mu_0\|_R = \frac{2 \cdot 3}{6} = 1$$

c'est-à-dire la moitié de la distance de la variation totale, ce qui veut dire qu'il y a un breakpoint tous les deux caractères en moyenne.

4 La norme de réarrangement sur $A^{\mathbb{Z}}$

On peut définir la condition de Bowen (et les fonctionnelles bw, Nbw) pour les fonctions réelles sur $\hat{X} = A^{\mathbb{Z}}$ de la même manière que pour les fonctions sur $X = A^{\mathbb{N}}$, et avec essentiellement les mêmes propriétés. Il en résulte qu'on peut définir la norme de réarrangement pour les mesures $\hat{\mu}$ sur $A^{\mathbb{Z}}$, en posant

$$\|\hat{\mu}\|_R = \operatorname*{Sup}_{\substack{f \in LC(\hat{X}) \\ \mathrm{Nbw}(f) \leqslant 1}} \langle f, \hat{\mu} \rangle = \operatorname*{Sup}_{\substack{f \in LC(\hat{X}) \\ \mathrm{Nbw}(f) \leqslant 1}} \left| \langle f, \hat{\mu} \rangle \right| \tag{4.1}$$

et tout ce qui a été dit précédemment sur la norme de réarrangement "unilatère" (sur $\mathcal{M}(X)$) reste valable essentiellement sans modification pour sa forme "bilatère" (sur $\mathcal{M}(\hat{X})$), simplement en changeant X en $\hat{X}$ dans les énoncés (et les preuves). En particulier, la proposition 2.6 et le théorème de dualité (théorème 3.2) restent valables dans leur forme bilatère.

Une des raisons pour considérer les mesures (invariantes ou non) sur $A^{\mathbb{Z}}$ au lieu de $A^{\mathbb{N}}$ est la présence d'une symétrie supplémentaire: l'involution J de $A^{\mathbb{Z}}$ définie par

$$J : (x_i)_{i \in \mathbb{Z}} \longmapsto (x_{-i})_{i \in \mathbb{Z}}$$

est un anti-automorphisme du décalage bilatère $(A^{\mathbb{Z}}, \lhd)$: elle conjugue $\lhd$ en son inverse. En particulier, $\mathcal{M}_{\lhd}(A^{\mathbb{Z}})$ est globalement invariant par l'action de J.

On a évidemment $\mathrm{bw}(fJ) = \mathrm{bw}(f)$ pour toute fonction $f : \hat{X} \to \mathbb{R}$, et par conséquent

$$\|J\hat{\mu}\|_R = \|\hat{\mu}\|_R \tag{4.2}$$

pour toute mesure $\hat{\mu}$ sur $\hat{X}$.

Soit $\Pi_+ : \hat{X} \to X$ l'application naturelle, consistant à "oublier" toutes les coordonnées < 0. On a évidemment $\mathrm{bw}(\phi\Pi_+) = \mathrm{bw}(\phi)$ pour toute fonction $\phi : X \to \mathbb{R}$: une fonction sur X peut être vue comme fonction sur $\hat{X}$, avec la même constante de Bowen. Par conséquent,

$$\|\Pi_+\hat{\mu}\|_R \leqslant \|\hat{\mu}\|_R \tag{4.3}$$

pour toute mesure $\hat{\mu}$ sur $\hat{X}$.

Mais on sait d'autre part que Π_+ induit un isomorphisme (d'espaces vectoriels) entre $\mathcal{M}_{\lhd}(\hat{X})$ et $\mathcal{M}_{\lhd}(X)$, et il est courant d'identifier une mesure $\lhd$-invariante sur $\hat{X}$ avec la mesure invariante correspondante sur X. Les définitions unilatère et bilatère de la norme de réarrangement sont compatibles avec cette identification, par le résultat suivant:

Proposition 4.1. *Pour toute mesure invariante $\hat{\mu}$ sur $\hat{X}$, on a*

$$\|\Pi_+\hat{\mu}\|_R = \|\hat{\mu}\|_R\,. \tag{4.4}$$

Démonstration. Il s'agit de montrer que, pour toute fonction $f : \hat{X} \to \mathbb{R}$ localement constante et telle que $\mathrm{Nbw}(f) < 1$, on a

$$\langle f, \hat{\mu}\rangle \leqslant \|\mu\|_R \tag{4.5}$$

où on a posé $\mu = \Pi_+(\hat{\mu})$.

Pour des raisons de densité, il suffit de traiter le cas où $\hat{\mu}$ est à support fini. Soit donc $\hat{F}$ un sous-ensemble fini, invariant, non vide de $\hat{X}$ qui porte $\hat{\mu}$; c'est une union finie d'orbites périodiques de $\hat{X}$, et $F = \Pi_+(\hat{F})$ est une union finie d'orbites périodiques de X, avec $\Pi_+ : \hat{F} \to F$ bijective.

Soit $\phi : F \to \mathbb{R}$ l'unique fonction telle que

$$\forall x \in \hat{F} \qquad \phi(\Pi_+ x) = f(x),$$

autrement dit, ϕ est la restriction à $\hat{F}$ de f, mais vue comme fonction sur F. Cela entraîne évidemment $\mathrm{Nbw}(\phi) \leqslant \mathrm{Nbw}(f)$, et par le lemme 3.5, il existe une fonction continue $\Phi : X \to \mathbb{R}$ telle que $\mathrm{Nbw}(\Phi) \leqslant 1$ et qui prolonge ϕ. Comme $\Phi\Pi_+$ et f coïncident $\hat{\mu}$-presque partout, on a

$$\langle f, \hat{\mu}\rangle = \langle \Phi\Pi_+, \hat{\mu}\rangle = \langle \Phi, \Pi_+\hat{\mu}\rangle \leqslant \mathrm{Nbw}(\Phi) \cdot \|\mu\|_R \leqslant \|\mu\|_R$$

ce qu'il fallait démontrer. $\square$

Références

[Bou] T. Bousch, *La condition de Walters*, Ann. Sci. Ec. Norm. Sup. **34** (2001), 287–311

[BM] T. Bousch & J. Mairesse, *Asymptotic height optimization for topical IFS, Tetris heaps, and the finiteness conjecture*, J. Amer. Math. Soc. **15** (2002), 77–111

[Bow] R. Bowen, *Some systems with unique equilibrium states*, Math. Syst. Theory **8** (1974), 193–202

[Do1] R. L. Dobrushin, *Description of a random field by means of conditional probabilities and conditions for its regularity*, Teor. Verojatnost. i Primenen **13** (1968), 201–229. Traduction anglaise dans: Theor. Probab. Appl. **13** (1968), 197–224

[Do2] R. L. Dobrushin, *Definition of a system of random variables by means of conditional distributions*, Teor. Verojatnost. i Primenen **15** (1970), 469–497. Traduction anglaise dans: Theor. Probab. Appl. **15** (1970), 458–486

[Dud] R. M. Dudley, *Probabilities and metrics. Convergence of laws on metric spaces, with a view to statistical testing*, Lecture Notes Series **45**, Matematisk Institut, Aarhus Universitet (1976)

[F+] G. Fertin, A. Labarre, I. Rusu, E. Tannier & S. Vialette, *Combinatorics of genome rearrangements*, MIT Press (2009)

[Fol] H. Föllmer, *A covariance estimate for Gibbs measures*, J. Funct. Anal. **46** (1982), 387–395

[Geo] H.-O. Georgii, *Gibbs measures and phase transitions*, de Gruyter studies in Mathematics **9** (1988)

[GNS] R. M. Gray, D. L. Neuhoff & P. C. Shields, *A generalization of Ornstein's $\bar{d}$ distance with applications to information theory*, Ann. Prob. **3** (1975), 315–328

[HR] N. T. A. Haydn & D. Ruelle, *Equivalence of Gibbs and equilibrium states for homeomorphisms satisfying expansiveness and specification*, Comm. Math. Phys. **148** (1992), 155–167

[KR1] L. V. Kantorovich & G. S. Rubinshtein, *On a functional space and certain extremum problems*, Dokl. Akad. Nauk SSSR (N.S.) **115** (1957), 1058–1061

[KR2] L. V. Kantorovich & G. S. Rubinshtein, *On a space of totally additive functions*, Vest. Leningrad Univ. **13** (1958), 52–59

[Kun] H. Künsch, *Decay of correlations under Dobrushin's uniqueness condition and its applications*, Comm. Math. Phys. **84** (1982), 207–222

[Lan] O. E. Lanford, *Entropy and equilibrium states in classical statistical mechanics*, 1–113 dans: A. Lenard (dir.), *Statistical mechanics and mathematical problems*, Battelle Seattle 1971 Rencontres, Lecture notes in Physics **20**, Springer (1973)

[Orn] D. S. Ornstein, *Ergodic theory, randomness, and dynamical systems*, Yale Mathematical Monographs **5**, Yale University Press (1974)

[Rue] D. Ruelle, *Thermodynamic formalism for maps satisfying positive expansiveness and specification*, Nonlinearity **5** (1992), 1223–1236

[Rus] L. Rüschendorf, *Wasserstein-metric* (1998), article pour M. Hazewinkel (dir.), Encyclopaedia of Mathematics, Supplement I, II, III, Kluwer Academic Publishers (1997–2001). Disponible sur le Web, en:
http://www.stochastik.uni-freiburg.de/~rueschendorf/papers/wasserstein.pdf

[Shi] P. C. Shields, *The interactions between ergodic theory and information theory*, IEEE Trans. Inform. Theory **44** (1998), 2079–2093

[Ver] A. Vershik, *The Kantorovich metric: the initial history and little-known applications*, arXiv:math/0503035 (2005)

[Wa1] P. Walters, *Invariant measures and equilibrium states for some mappings which expand distances*, Trans. Amer. Math. Soc. **236** (1978), 127–153

[Wa2] P. Walters, *Convergence of the Ruelle operator for a function satisfying Bowen's condition*, Trans. Amer. Math. Soc. **353** (2001), 327–347

[Wa3] P. Walters, *Regularity conditions and Bernoulli properties of equilibrium states and g-measures*, J. London Math. Soc. (2) **71** (2005), 379–396

[Wa4] P. Walters, *A natural space of functions for the Ruelle operator theorem*, Ergod. Th. Dynam. Sys. **27** (2007), 1323–1348