Séminaire de probabilités (Strasbourg)

Laurent Miclo

Remarques sur l'hypercontractivité et l'évolution de l'entropie pour des chaînes de Markov finies

Séminaire de probabilités (Strasbourg), tome 31 (1997), p. 136-167
<http://www.numdam.org/item?id=SPS_1997__31__136_0>

Remarques sur l'hypercontractivité et l'évolution de l'entropie pour des chaînes de Markov finies

Laurent Miclo
Université Paul Sabatier de Toulouse

Summary : We will show how, in the discrete times setting of finite (irreducible and aperiodic) Markov chains, one can still use some logarithmic-Sobolev inequalities to study hypercontractivity and the evolution of entropy. As an application, we will give a new simple proof of a criterion of Hwang and Sheu for the strong ergodicity in law of the generalised simulated annealing algorithms in discrete times.

Résumé : Nous allons montrer comment, dans le cadre du temps discret des chaînes de Markov finies irréductibles et apériodiques, on peut encore utiliser certaines inégalités de Sobolev-logarithmiques pour obtenir des résultats sur l'hypercontractivité et l'évolution de l'entropie. On illustrera ces techniques de semi-groupes en retrouvant simplement un critère de Hwang et Sheu pour l'ergodicité en loi forte des algorithmes de recuit généralisés à temps discret.

Abbreviated title : Hypercontractivité des chaînes de Markov finies.

American Mathematical Society 1991 subject classifications : Primary 60J10 ; secondary 47A50 and 47N30.

Key words and phrases : Irreducible and aperiodique finite Markov chains, logarithmic-Sobolev inequalities, hypercontractivity, entropy evolution, strong ergodicity in law for generalised simulated annealing.

1 Introduction

Les inégalités de Sobolev-logarithmiques furent introduites par Gross dans [8] pour traiter notamment de l'hypercontractivité des processus d'Ornstein-Uhlenbeck en dimension infinie. Mais elles se sont vite révélées intéressantes pour d'autres types de processus (cf. par exemple Rothaus [18] et Holley et Stroock [9]) et jusque dans un cadre général de théorie des semi-groupes (voir Bakry [1] et les références qui y sont données). Même dans le contexte relativement plus simple (en apparence!) des processus de Markov finis homogènes, elles permirent de faire des progrès dans la compréhension des vitesses de convergence vers l'équilibre (cf. Diaconis et Saloff-Coste [4], nous reprendrons d'ailleurs ici les notations de cet article). Cependant leur domaine d'application semblait restreint aux cas de processus indicés par un temps continu, car l'intérêt de ces inégalités était souvent de permettre une majoration des dérivées temporelles de certaines quantités naturellement associées aux semi-groupes. Notre but ici est de montrer comment on peut également les utiliser pour étudier des chaînes de Markov finies à temps discret. Précisons tout de suite que de supposer l'espace des états finis est un peu frustrant, mais nous a été imposé par les inégalités de Sobolev-logarithmiques présentées dans la section suivante, qui ne sont pertinentes que dans ce cadre fini (les autres calculs étant sinon souvent valables dans une plus grande généralité).

Soit donc S un ensemble fini (non réduit à un singleton), muni d'un noyau $p = (p(x,y))_{x,y\in S}$ de probabilités de transitions que l'on supposera irréductible : pour tous $x, y \in S$, il existe $n \in \mathbb{N}$ et une suite finie $x = x_0, x_1, \cdots, x_n = y$ d'éléments de S telle que pour tout $1 \leq i \leq n$, $p(x_{i-1}, x_i) > 0$ (une telle suite sera appelée un chemin de longueur n, de points de départ x et d'arrivée y), i.e. il n'y a qu'une classe de récurrence et pas de points transients. Il existe donc une unique probabilité invariante μ pour p, qui est caractérisée par

$$\forall\, x \in S, \qquad \sum_{y\in S} \mu(y)p(y,x) = \sum_{y\in S} \mu(x)p(x,y) = \mu(x)$$

et elle charge tous les points.

Comme d'habitude, on associe à p un opérateur P qui agit d'une part sur les fonctions réelles définies sur S (leur ensemble sera désormais noté $\mathcal{F}(S)$) par

$$\forall\, f \in \mathcal{F}(S),\ \forall\, x \in S, \qquad Pf(x) = \sum_{y\in S} p(x,y)f(y)$$

et d'autre part sur les probabilités sur S (dont l'ensemble sera désigné par $\mathcal{P}(S)$) par

$$\forall\, m \in \mathcal{P}(S),\ \forall\, x \in S, \qquad mP(x) = \sum_{y\in S} m(y)p(y,x)$$

l'invariance de μ s'écrivant alors $\mu P = \mu$.

Ces opérations ont clairement les interprétations probabilistes suivantes : si $X = (X_n)_{n\in\mathbb{N}}$ est une chaîne de Markov sur S dont les probabilités de transitions sont données par p et qui est issue de $x \in S$, alors pour tout $f \in \mathcal{F}(S)$, $Pf(x) = \mathbb{E}_x[f(X_1)]$. Si par contre on suppose X de loi initiale $m \in \mathcal{P}(S)$, alors mP est la loi de X_1.

Pour $n \in \mathbb{N}$, on désignera aussi par P^n les itérés n fois de P (auxquelles correspondent les matrices p^n), en convenant que $P^0 = I$, où I est l'application identité, agissant suivant les cas sur $\mathcal{F}(S)$ ou sur $\mathcal{P}(S)$.

La première question que l'on se pose est de savoir si le semi-groupe $(P^n)_{n\in\mathbb{N}}$ est hypercontractif: si $2 \leq q_0 < +\infty$ est donné, existe-t-il des $q > q_0$ tels que pour tout $f \in \mathcal{F}(S)$,

$$(1) \qquad \|Pf\|_q \leq \|f\|_{q_0}$$

où pour tout $1 \leq r \leq \infty$, $\|\cdot\|_r$ désignera la norme usuelle de $L^r(S,\mu)$.

La seconde question concerne l'évolution de l'entropie le long des itérées de P agissant sur $\mathcal{P}(S)$, et plus particulièrement de décrire quantitativement sa décroissance, pour pouvoir dans les situations apériodiques donner sa vitesse de convergence vers 0 en temps grand. Rappelons que l'entropie par rapport à μ d'une mesure $m \in \mathcal{P}(S)$ est définie par

$$\mathrm{Ent}(m) = \sum_{x\in S} \ln\left(\frac{m(x)}{\mu(x)}\right) m(x)$$

(par la suite, dans le cas de processus inhomogènes, plusieurs mesures invariantes instantanées apparaîtront et cette expression sera alors plutôt notée $\mathrm{Ent}(m|\mu)$ pour éviter les confusions), et que c'est une quantité qui mesure d'une certaine manière un écart à la probabilité μ, car on a par exemple

$$(2) \qquad \|m-\mu\|_{vt} \leq \sqrt{2}\sqrt{\mathrm{Ent}(m)}$$

où $\|\cdot\|_{vt}$ représente la variation totale (cf. Stroock [19] formule (1.12)).

Dans la section suivante, on introduira certaines inégalités de Sobolev-logarithmiques dont les constantes associées permettront d'apporter des réponses aux problèmes précédents, respectivement dans les sections 3 et 4. Enfin dans une dernière section, on illustrera ces techniques en retrouvant, relativement facilement, un critère d'ergodicité en loi des algorithmes de recuit généralisés finis à temps discret.

2 Inégalités de Sobolev-logarithmiques

Les inégalités de Sobolev-logarithmiques consistent à comparer une certaine fonctionnelle $\mathcal{L}$, qui ressemble à l'entropie mais qui est définie sur $\mathcal{F}(S)$ par

$$\forall\, f \in \mathcal{F}(S), \qquad \mathcal{L}(f) = \int f^2 \ln\left(\frac{f^2}{\|f\|_2^2}\right) d\mu$$

(cette quantité sera aussi notée $\mathcal{L}_\mu(f)$ quand la probabilité μ ne sera plus sous-entendue), à une forme de Dirichlet. Quand les processus étudiés sont à temps continu, la forme intervenant est celle associée à l'opposé du générateur (ou de manière équivalente, à l'opposé du symétrisé additif de cet opérateur dans $L^2(\mu)$), mais quand le temps est discret, c'est suivant les cas celle associée à $I - P^*P$ ou à $I - PP^*$ qui apparaît naturellement (voir aussi Fill [5] et Diaconis et Saloff-Coste [3], où P^*P est appelé le symétrisé multiplicatif pour le distinguer du symétrisé additif $(P + P^*)/2$).

On a noté ci-dessus P^* l'adjoint de P dans $L^2(\mu)$, et pour tout $f \in \mathcal{F}(S)$, soit

$$\mathcal{E}(f,f) = \int (I - P^*P)(f)f\, d\mu = \int f^2 - (Pf)^2\, d\mu = \|f\|_2^2 - \|Pf\|_2^2$$

(quand on voudra préciser les opérateurs intervenant, cette expression sera aussi notée $\mathcal{E}_{I-P^*P}(f,f)$). Du fait que P est une contraction dans $L^2(\mu)$, il est clair que $\mathcal{E}(f,f) \geq 0$. On calcule immédiatement que P^* agit sur $\mathcal{F}(S)$ par

$$\forall\, f \in \mathcal{F}(S),\ \forall\, x \in S, \qquad P^*f(x) = \sum_{y\in S} p^*(x,y)f(y)$$

avec pour tous $x, y \in S$, $p^*(x,y) = \mu(y)p(y,x)/\mu(x)$.

Le fait que μ est invariante pour P implique que P^* est un noyau markovien

$$\forall\, x \in S, \qquad \sum_{y\in S} p^*(x,y) \;=\; \frac{1}{\mu(x)} \sum_{y\in S} \mu(y)p(y,x) = 1$$

qui de plus admet aussi μ pour probabilité invariante

$$\forall\, x \in S, \qquad \sum_{y\in S} \mu(x)p^*(x,y) = \sum_{y\in S} \mu(y)p(y,x) = \mu(x)$$

Notons qu'en posant pour $x, y \in S$,

$$q(x,y) = (p^*p)(x,y) = \mu(x)^{-1} \sum_{z\in S} \mu(z)p(z,x)p(z,y)$$

on peut expliciter un peu plus la forme de Dirichlet,

$$(3) \qquad \forall\, f \in \mathcal{F}(S), \qquad \mathcal{E}(f,f) = \frac{1}{2} \sum_{x,y\in S} \mu(x)q(x,y)(f(y)-f(x))^2$$

Définition :

On dit que $I - P^*P$ satisfait une inégalité de Sobolev-logarithmique s'il existe une constante $\alpha > 0$ telle que pour tout $f \in \mathcal{F}(S)$,

$$\mathcal{L}(f) \leq \alpha^{-1}\mathcal{E}(f,f)$$

et $\alpha(I - P^*P)$ désignera alors la plus grande constante (appelée constante de Sobolev-logarithmique) telle que toutes ces inégalités soient satisfaites.

Remarque :

Une telle inégalité implique que pour tout $f \in \mathcal{F}(S)$, $\mathcal{L}(f) \leq \alpha^{-1} \|f\|_2^2$, ce qui à son tour impose que S est fini (du moins que μ est une combinaison convexe d'un nombre fini de masses de Dirac, ce qui nous ramène à ce cas), comme on en s'en rend compte en considérant sinon des indicatrices d'ensembles de mesure de plus en plus petites pour μ (voir aussi les contre-exemples de la fin de cette section).

En fait dans le contexte des processus de Markov finis, il est bien connu qu'une inégalité de Sobolev-logarithmique du type précédent est équivalente à l'irréductiblité de p^*p: en effet pour la condition nécessaire, il suffit de remarquer que si on pose $f_\epsilon = \mathbf{I} + \epsilon h$, avec $\mathbf{I}$ la fonction prenant toujours la valeur 1 et $h \in \mathcal{F}(S)$ fixé, on a $\mathcal{E}(f_\epsilon, f_\epsilon) = \epsilon^2\mathcal{E}(h,h)$ et pour ϵ tendant vers 0, $\mathcal{L}(f_\epsilon) \sim 2\epsilon^2\mu((f_\epsilon - \mu(f_\epsilon))^2)$, d'où l'existence d'un trou spectral (d'au moins $2\alpha(I - P^*P)$) pour l'opérateur symétrique $I - P^*P$ (pour plus de détails, on renvoit à Rothaus [18] ou à Diaconis et Saloff-Coste

[4]), ce qui est aussi équivalent à l'irréductibilité de p^*p (ce dernier point se montrant à partir de l'expression (3), voir par exemple la preuve de l'inégalité de Poincaré faite par Holley et Stroock [10]). Quant à la condition suffisante, puisque $\mathcal{L}$ et $\mathcal{E}$ sont tous deux homogènes d'ordre 2, il suffit de voir que

$$\inf_{f\in\mathcal{F}(S)\,/\,\|f\|_2=1,\,f\notin\mathrm{Vect}(\mathbf{I})} \frac{\mathcal{E}(f,f)}{\mathcal{L}(f)} > 0$$

et même que, puisque pour tout $f \in \mathcal{F}(S)$, $\mathcal{E}(f,f) \geq \mathcal{E}(|f|,|f|)$,

$$\inf_{f\in\mathcal{F}(S)\,/\,\|f\|_2=1,\,f\neq\mathbf{I},\,f\geq 0} \frac{\mathcal{E}(f,f)}{\mathcal{L}(f)} > 0$$

mais ceci découle, outre du trou spectral et d'un développement du type précédent au voisinage de $\mathbf{I}$, du fait que l'application $f \mapsto \mathcal{E}(f,f)/\mathcal{L}(f)$ est continue (on aura noté que par l'inégalité de Jensen et la stricte convexité de $\mathbb{R}_+ \ni t \mapsto t\ln(t)$, $\mathcal{L}(f)$ est nul si et seulement si f^2 est constante) sur le compact formé de l'ensemble $\{f \in \mathcal{F}(S)\,/\; \|f\|_2 = 1, f \neq \mathbf{I}, f \geq 0\}$ privé de son intersection avec une petite boule (pour $\|\cdot\|_2$) autour de $\mathbf{I}$, elle y atteint donc son minimum qui ne peut être nul car $\mathcal{E}(f,f) = 0$ équivaut de par la formule (3) et l'irréductibilité de p^*p à f constant.

Pour ceci et pour des estimations de $\alpha(I - P^*P)$ (qui constituent le point crucial dans les applications, voir par exemple la dernière section) on renvoit aussi à Diaconis et Saloff-Coste [4].

Notons que le trou spectral de $I - P^*P$ est toujours majoré par 1, ce qui montre que d'une manière générale, on a $\alpha(I - P^*P) \leq 1/2$.

Le résultat très simple suivant va nous permettre de voir quelles sont les situations pour lesquelles des inégalités de Sobolev-logarithmiques peuvent être utiles. On y suppose seulement que le noyau p admet μ pour mesure invariante. L'adjoint P^* est alors toujours bien défini dans $L^2(\mu)$, même si la matrice p^* n'est plus unique, il suffit de poser $p^*(x,y) = \mu(y)p(y,x)/\mu(x)$ si $\mu(x) > 0$ et de prendre pour $p^*(x,\cdot)$ une probabilité quelconque sur S si $\mu(x) = 0$ (pour permettre au noyau p^* de rester markovien et d'admettre également μ pour mesure invariante). Le résultat ci-dessous ne dépend pas du choix éventuel de p^*.

Proposition 1

On a équivalence entre
(i) Il existe un $k \geq 1$ tel que $p^{k*}p^k$ est irréductible.
(ii) p est irréductible et apériodique.

Démonstration :

Supposons (ii), il est bien connu qu'alors il existe un $k \geq 1$ tel que pour tous $x, y \in S$, $p^k(x,y) > 0$, or il est clair à partir des égalités

$$p^{k*} = (p^k)^* = (p^*)^k$$

que ceci implique aussi que pour tous $x, y \in S$, $p^{k*}(y,x) > 0$, puis que

$$\forall\; x,y \in S, \qquad (p^{k*}p^k)(x,y) > 0$$

et donc notamment l'irréductibilité de $p^{k*}p^k$.

Réciproquement, supposons d'abord que p admette une classe de récurrence $C \neq S$. Il existe une telle classe satisfaisant pour tout $x \in C$, $\mu(x) > 0$. Or pour $x \in C$ et $y \in S$, l'inégalité $p^*(x,y) > 0$ implique, puisque $\mu(x) > 0$, d'une part que $\mu(y) > 0$, et donc notamment que y ne peut pas être un point transient, et d'autre part que $p(y,x) > 0$. Ces deux informations montrent que $y \in C$ et on aboutit à la même conclusion si $p(x,y) > 0$. Ainsi pour tous $k \geq 1$, $x \in C$ et $y \in S$, l'inégalité $(p^{k*}p^k)(x,y) > 0$ impose que y appartient aussi à la classe C et il en découle que $p^{k*}p^k$ a une classe de récurrence dans C, ce qui est incompatible avec l'irréductibilité de $p^{k*}p^k$. Supposons maintenant que p irréductible admette plusieurs classes de périodicité, disons $C_0, \cdots, C_{d-1}$ ($d > 1$ est alors la période, et on suppose que $C_0, \cdots, C_{d-1}$ sont rangés consécutivement). En considérant leurs indices comme des éléments de $\mathbb{Z}/d\mathbb{Z}$, pour tous $k \geq 1$, $i \in \mathbb{Z}/d\mathbb{Z}$, $x \in C_i$ et $y \in S$, on a l'implication $p^k(x,y) > 0 \Rightarrow y \in C_{i+k}$ (respectivement $p^{k*}(x,y) > 0 \Rightarrow y \in C_{i-k}$), ce qui prouve que $p^{k*}p^k(x,y) > 0$ assure que x et y sont dans une même classe de périodicité et donc que $p^{k*}p^k$ ne peut être irréductible.

$\square$

Si p est irréductible et apériodique, notons

$$k(p) = \min\{k \in \mathbb{N}^* \,/\, p^{k*}p^k \text{ est irréductible}\} \tag{4}$$

Pour appliquer les résultats des deux sections suivantes, qui seront satisfaits sous l'hypothèse d'existence d'inégalités de Sobolev-logarithmiques, il faudra au moins remplacer la matrice p par $p^{k(p)}$, et même souvent il sera intéressant de considérer plutôt p^k avec un certain $k > k(p)$ pour accroître fortement la constante de Sobolev-logarithmique (voir l'exemple de la remarque (b) de la fin de la section 5). L'argument précédent montre que

$$k(p) \leq \min\{k \geq 1 \,/\, \forall\, x,y \in S,\ p^k(x,y) > 0\}$$

mais on peut avoir une inégalité stricte comme le montre l'exemple donné par la matrice

$$p = \frac{1}{2}\begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 0 & 2 & 0 \end{pmatrix}$$

pour laquelle $k(p) = 1$. Par ailleurs pour tout $n \in \mathbb{N}$, $n \geq 2$, la matrice p définie sur $\mathbb{Z}/(n\mathbb{Z})$ par

$$\forall\, i,j \in \mathbb{Z}/(n\mathbb{Z}), \qquad p(i,j) = \begin{cases} 1 & , \text{si } i \neq 0 \text{ et } j = i+1 \\ 1/2 & , \text{si } i = 0 \text{ et } j = 1 \\ 1/2 & , \text{si } i = 0 = j \\ 0 & , \text{sinon} \end{cases}$$

fournit un exemple très simple pour lequel $k(p) = n - 1$.

Revenons au cas général, à partir des inégalités de Cauchy-Schwarz,

$$\forall\, k \geq 1,\ \forall\, f \in \mathcal{F}(S),\ \forall\, x \in S, \qquad (P^{k+1}f(x))^2 \leq P((P^kf)^2)(x)$$

et de l'invariance de μ par rapport à P, on remarque que

$$\forall\, k \geq 1, \forall\, f \in \mathcal{F}(S), \qquad \mathcal{E}_{I-P^{k+1*}P^{k+1}}(f,f) \geq \mathcal{E}_{I-P^{k*}P^k}(f,f)$$

ce qui montre que

$$\mathbb{N}^* \ni k \mapsto \alpha(I - P^{k*}P^k)$$

est une application croissante (en convenant que $\alpha(I - P^{k*}P^k) = 0$ si $p^{k*}p^k$ n'est pas irréductible). Celle-ci admet d'ailleurs $\alpha(I - E_\mu^* E_\mu) = \alpha(I - E_\mu)$ comme limite en l'infini, où E_μ est la matrice correspondant à l'espérance par rapport à μ et qui est donnée par

$$\forall\, x, y \in S, \qquad E_\mu(x,y) = \mu(y)$$

Indiquons que la valeur de $\alpha(I - E_\mu)$ a été astucieusement calculée par Diaconis et Saloff-Coste (cf. le théorème 5.1 de l'appendice de [4]) et qu'elle vaut $(1-2\underline{\mu})/\ln(1/\underline{\mu}-1)$, où $\underline{\mu} = \min_{x\in S} \mu(x)$.

Il apparaît également que pour tout $k \geq k(p)$, $p^{k*}p^k$ sera irréductible, puisque ceci revient à dire que $\alpha(I - P^{k*}P^k) > 0$, inégalité elle-même équivalente à

$$\forall\, f \in \mathcal{F}(S), \qquad \mathcal{E}_{I-P^{k*}P^k}(f,f) = 0 \Rightarrow f \text{ est constant}$$

<u>Remarques</u> :

a) Dans le même ordre d'idées, notons pour tout $n \in \mathbb{N}$,

$$V_n = \{f \in \mathcal{F}(S) \,/\, \|P^n f\|_2 = \|f\|_2\}$$

Puisque pour tout $x \in S$, on a $(P^n f(x))^2 \leq P^n(f^2)(x)$ avec égalité si et seulement si f est constante sur l'ensemble $\{y \in S \,/\, p^n(x,y) > 0\}$, il apparaît que V_n est un sous-espace vectoriel de $\mathcal{F}(S)$. De plus, en utilisant les inégalités $\|P^{n+1}f\|_2 \leq \|P^n f\|_2 \leq \cdots \leq \|Pf\|_2 \leq \|f\|_2$ valables pour tout $f \in \mathcal{F}(S)$, on montre que pour tout $n \in \mathbb{N}$,

$$f \in V_{n+1} \iff Pf \in V_n \text{ et } f \in V_1$$

De ceci il découle facilement que si pour un $n \in \mathbb{N}$ on a $V_{n+1} = V_n$, alors pour tout $r \in \mathbb{N}$, $V_{n+r} = V_n$. Or en supposant p irréductible apériodique, on a $V_{k(p)} = \mathrm{Vect}(\mathbf{1})$, ainsi (V_n) est une suite de sous-espaces vectoriels qui commence par être strictement décroissante ($V_0 = \mathcal{F}(S)$) puis qui finit par valoir $\mathrm{Vect}(\mathbf{1})$. On en déduit que l'on a toujours $k(p) \leq \mathrm{card}(S) - 1$ (l'égalité étant possible comme on l'a déjà vu dans l'exemple ci-dessus sur $\mathbb{Z}/(n\mathbb{Z})$). Ceci nous donne un résultat sur les chemins qui n'est pas facile à obtenir directement : en effet, $k(p)$ est le plus petit $k \in \mathbb{N}$ tel que pour tous $x, y \in S$, on puisse trouver deux suites finies $(q^{(i)})_{0\leq i\leq n}$ et $(\breve{q}^{(i)})_{0\leq i\leq n}$ de chemins (pour p) de longueur k satisfaisant $q_0^{(0)} = x$, $\breve{q}_0^{(n)} = y$, $q_k^{(i)} = \breve{q}_k^{(i)}$ pour tout $0 \leq i \leq n$ et $q_0^{(i+1)} = \breve{q}_0^{(i)}$ pour tout $0 \leq i < n$. Les résultats précédents montrent donc que l'on peut trouver de telles suites avec $k = \mathrm{card}(S) - 1$.

b) D'autre part, remarquons que P^*P est diagonalisable et n'admet que des valeurs propres positives, la plus grande étant 1. Puisque cette diagonalisation est orthogonale dans $L^2(\mu)$, $\|Pf\|_2 = \|f\|_2$ équivaut à l'appartenance de f à l'espace propre associé à la valeur propre 1 et l'irréductibilité de p^*p est donc équivalente au fait que celle-ci est de multiplicité 1. Or il est bien connu que P^*P et PP^* ont même spectre, ainsi l'irréductibilité de p^*p est équivalente à celle de pp^*, on a notamment $k(p^*) = k(p)$ et $I - P^*P$ satisfait une inégalité de Sobolev-logarithmique si et seulement si il en est de même pour $I - PP^*$. Cependant il n'est pas clair en général, mis à part dans les cas où p est normal (notamment si p est réversible, car on a alors $P^* = P$), que les constantes $\alpha(I - P^*P)$ et $\alpha(I - PP^*)$ soient égales. Or dans la section suivante c'est la première qui interviendra, alors que dans les sections 4 et 5 ce sera la seconde.

c) Si p est irréductible apériodique et réversible, notons que $k(p) = 1$, car s'il existait $f \in \mathcal{F}(S)$ non constante telle que $\mathcal{E}_{I-P^*P}(f,f) = 0$, ceci impliquerait que l'espace propre associé à la valeur propre 1 pour $P^2 = P^*P$ serait au moins de dimension 2, i.e. soit l'espace propre associé à 1 pour P est au moins de dimension 2, ce qui est exclus par l'irréductibilité, soit -1 est valeur propre de P, ce qui est en contradiction avec l'apériodicité.

On peut en déduire un résultat un peu plus général : supposons que p soit tel que

$$\forall\, x,y \in S, \qquad p(x,y) > 0 \Longleftrightarrow p(y,x) > 0 \tag{5}$$

alors on a aussi $k(p) = 1$.

En effet, $k(p)$ ne dépend que de la matrice d'incidence $i(p)$ associée à p, qui est donnée par

$$\forall\, x,y \in S, \qquad i(p)(x,y) = \mathbf{1}_{p(x,y)>0}$$

ainsi $k(p) = k(\tilde{p})$, où $\tilde{p}$ est la matrice markovienne définie par

$$\forall\, x,y \in S, \qquad \tilde{p}(x,y) = \begin{cases} 1/\mathrm{card}\{z \in S \,/\, p(x,z) > 0\} & \text{, si } p(x,y) > 0 \\ 0 & \text{, sinon} \end{cases}$$

or cette dernière est réversible si et seulement si p satisfait (5).

3 Hypercontractivité à temps discret

Traditionnellement dans le cas des processus de Markov finis, une inégalité d'hypercontractivité du type (1) se prouve pour $P = \exp(t_0 L)$ où $t_0 > 0$ et L est un générateur irréductible sur S. Plus précisément, on montre que toute une famille d'inégalités est satisfaite :

$$\forall\, t \geq 0, \forall\, f \in \mathcal{F}(S), \qquad \|\exp(tL)f\|_{q(t)} \leq \|f\|_{q_0}$$

où $q : \mathbb{R}_+ \to [2, +\infty[$ est une certaine application telle que $q(0) = q_0 \geq 2$. On les prouve en les dérivant par rapport à $t \geq 0$ et en imposant aux dérivées d'être négatives grâce à des inégalités de Sobolev-logarithmiques (cf. par exemple Gross [8] ou Diaconis et Saloff-Coste [4]). Ainsi, si on veut montrer (1), on peut commencer par chercher s'il n'existerait pas un générateur irréductible L sur S tel que $P = \exp(L)$. Cependant ceci est rarement satisfait, ne serait-ce que parce que cela implique déjà que pour tous $x, y \in S$, $p(x,y) > 0$. Nous allons plutôt essayer d'estimer directement la différence $\|Pf\|_q - \|f\|_{q_0}$ et voir comment les inégalités de Sobolev-logarithmiques de la section précédente peuvent permettre de la rendre négative.

On supposera donc que $I - P^*P$ satisfait une inégalité de Sobolev-logarithmique avec constante $\alpha(I - P^*P) > 0$, c'est-à-dire que P^*P est irréductible, ce qui entraîne notamment l'irréductibilité et l'apériodicité de P. Notons également

$$\nu(P) = \max\{1/p(x,y) \,/\, x,y \in S,\, p(x,y) > 0\} - 1$$

(on a donc $\nu(P) \geq 1$ par apériodicité de P), et si $q \geq 2$ et $\nu > 0$ sont donnés, on pose

$$g(q,\nu) = \frac{(1+\nu)^q - 1 - q\nu}{((1+\nu)^{q/2}-1)^2}$$

Si $q = 2$, on a $g(2,\cdot) \equiv 1$, mais si $q > 2$ est fixé, on vérifie aisément que l'application $\mathbb{R}_+^* \ni \nu \mapsto g(q,\nu)$ est strictement décroissante et satisfait

$$g(q,0_+) = 2\frac{q-1}{q}, \qquad g(q,+\infty) = 1$$

D'autre part, on peut aussi montrer qu'à $\nu > 0$ fixé, $[2,+\infty[\ni q \mapsto g(q,\nu)$ commence par être croissante puis décroît, les limites étant $g(2,\nu) = 1 = g(+\infty,\nu)$.

Avec ces notations, le principal résultat de cette section s'énonce alors,

Proposition 2

Pour tous $q_0 \geq 2$ et $f \in \mathcal{F}(S)$, on a

$$\|Pf\|_q \leq \|f\|_{q_0}$$

dès que q satisfait $q \leq [1 + g(q,\nu(P))\alpha(I - P^*P)]q_0$, et donc notamment si $q = [1 + \alpha(I - P^*P)]q_0$

La preuve de cette proposition est basée sur deux lemmes très simples :

Lemme 3

Fixons $f \in \mathcal{F}(S)$, l'application

$$[0,1] \ni t \mapsto \|f\|_{1/t}$$

est convexe. Notamment pour tous $q \geq q_0 \geq 1$,

$$\|f\|_q - \|f\|_{q_0} \leq \frac{q-q_0}{qq_0}\|f\|_q^{1-q}\mathcal{L}(f^{q/2})$$

Démonstration :

Soit $0 < \theta < 1$ et $0 < s < t < 1$, on veut estimer $\|f\|_{1/(\theta s+(1-\theta)t)}$. Pour ceci soit $0 < r < 1$ tel que

$$\frac{1}{\theta s + (1-\theta)t} = \frac{r}{s} + \frac{1-r}{t}$$

on calcule que $r = \theta s/(\theta s + (1-\theta)t)$.

Par l'inégalité de Hölder on a

$$\begin{aligned}
\|f\|_{1/(\theta s+(1-\theta)t)} &\leq \|f\|_{1/s}^{r(\theta s+(1-\theta)t)/s}\|f\|_{1/t}^{(1-r)(\theta s+(1-\theta)t)/t} \\
&= \|f\|_{1/s}^{\theta}\|f\|_{1/t}^{1-\theta} \\
&\leq \theta\|f\|_{1/s} + (1-\theta)\|f\|_{1/t}
\end{aligned}$$

par convexité de l'application exponentielle. Pour la seconde affirmation, il reste à vérifier que

$$\begin{aligned}\frac{d}{dt}\|f\|_{1/t} &= -\|f\|_{1/t}^{1-1/t}\int \ln\left(\frac{f^{1/t}}{\|f\|_{1/t}^{1/t}}\right) f^{1/t}\,d\mu \\ &= -\|f\|_{1/t}^{1-1/t}\,\mathcal{L}(f^{\frac{1}{2t}})\end{aligned}$$

□

Le second lemme concerne une petite inégalité qui précise la stricte convexité de l'application $\mathbb{R}_+ \ni t \mapsto t^q$, pour $q \geq 2$. C'est elle qui a déterminé la fonction $g(\cdot,\cdot)$ précédente.

Lemme 4

Soit $\nu \geq 0$ et $q \geq 2$ fixés. Pour tous $t \geq 0$ et $-t \leq s \leq \nu t$, on a

$$(t+s)^q \geq t^q + qt^{q-1}s + g(q,\nu)((t+s)^{q/2} - t^{q/2})^2$$

Démonstration :

En effet, $t > 0$ étant fixé, notons pour $u = s/t$, $h(u)$ le membre de gauche moins le membre de droite le tout divisé par t^q. En dérivant deux fois, il apparaît que h a le comportement suivant :

$$\begin{array}{c|ccccccc} u & -1 & & 0 & & u(q,\nu) & & \nu \\ \hline h & h(-1) & \searrow & 0 & \nearrow & h(u(q,\nu)) & \searrow & 0 \end{array}$$

pour une certaine valeur $0 < u(q,\nu) < \nu$, ce qui permet de conclure au résultat annoncé.

□

Démonstration de la proposition 2 :

Soit $q \geq q_0 \geq 2$ a priori quelconques, et comme d'habitude, on commence par se ramener à ne considérer que des fonctions f positives. On écrit ensuite

$$\|Pf\|_q - \|f\|_{q_0} = \|Pf\|_q - \|f\|_q + \|f\|_q - \|f\|_{q_0}$$

et on majore l'expression $\|f\|_q - \|f\|_{q_0}$ comme dans lemme 3.

Pour l'autre terme, on utilise l'inégalité de concavité $a^{1/q} - b^{1/q} \leq \frac{1}{q}b^{1/q-1}(a-b)$, valable pour tous réels $a, b \geq 0$, pour obtenir

$$\|Pf\|_q - \|f\|_q \leq \frac{1}{q}\|f\|_q^{1-q}(\|Pf\|_q^q - \|f\|_q^q)$$

ce qui nous ramène à estimer $\int (Pf)^q\,d\mu - \int f^q\,d\mu$. Pour ceci on se sert du lemme 4 appliqué avec $\nu = \nu(P)$, $t = Pf(x)$ et $t + s = f(y)$, où $x, y \in S$ sont tels que $p(x,y) > 0$.

Mais $x \in S$ étant fixé, on intègre ces inégalités en y par rapport à $p(x,y)$ pour obtenir

$$P(f^q)(x) \geq (P(f)(x))^q + g(q,\nu(P)) \sum_{y\in S} p(x,y)(f^{q/2}(y) - (Pf)^{q/2}(x))^2$$

Cependant, toujours à $x \in S$ fixé, on a

$$\begin{aligned}
\sum_{y\in G} p(x,y)\left(f^{q/2}(y) - (Pf)^{q/2}(x)\right)^2 &\geq \min_{c\in\mathbb{R}} \sum_{y\in G} p(x,y)\left(f^{q/2}(y) - c\right)^2 \\
&= \sum_{y\in G} p(x,y)\left(f^{q/2}(y) - P(f^{q/2})(x)\right)^2 \\
&= P(f^q)(x) - \left(P(f^{q/2})(x)\right)^2
\end{aligned}$$

d'où

$$P(f^q)(x) \geq (P(f)(x))^q + g(q,\nu(P))\left(P(f^q)(x) - \left(P(f^{q/2})(x)\right)^2\right) \tag{6}$$

Il reste alors à intégrer ces inégalités en x par rapport à μ et à utiliser l'invariance de μ par P pour faire apparaître que

$$\|Pf\|_q^q - \|f\|_q^q \leq -g(q,\nu(P))\mathcal{E}(f^{q/2}, f^{q/2})$$

Finalement on a donc montré que

$$\|Pf\|_q - \|f\|_{q_0} \leq \frac{1}{q}\|f\|_q^{1-q}\left[\frac{q-q_0}{q_0}\mathcal{L}(f^{q/2}) - g(q,\nu(P))\mathcal{E}(f^{q/2}, f^{q/2})\right] \tag{7}$$

qui est un analogue pour des différences finies de l'estimation classique de la dérivée dans la preuve de l'hypercontractivité.

Si maintenant $q \leq [1 + g(q,\nu(P))\alpha(I - P^*P)]q_0$, on peut conclure par l'inégalité de Sobolev-logarithmique

$$\alpha(I - P^*P)\mathcal{L}(f^{q/2}) \leq \mathcal{E}(f^{q/2}, f^{q/2})$$

$\square$

<u>Remarques :</u>

Dans les applications, on a souvent $\nu(P)$ très grand, et on peut alors se contenter de prendre $q = (1 + \alpha(I - P^*P))q_0$. Notons d'ailleurs que pour la preuve du résultat d'hypercontractivité avec ce q, le lemme 4 est inutile, car on a besoin de l'inégalité (6) qu'avec $g(q,\nu(P))$ remplacé par 1, et dans ce cas elle se réduit à

$$\left(P(f^{q/2})(x)\right)^2 \geq (P(f)(x))^q$$

qui est trivialement satisfait pour $q \geq 2$.

Néanmoins nous avons tenu à présenter le lemme 4, car c'est une inégalité du même genre qui interviendra dans la section suivante, et curieusement la différence entre $g(q,0_+)$ et $g(q,+\infty)$ est la "même" que celle qui intervient dans le lemme

2.6 de Diaconis et Saloff-Coste [4] entre les situations réversibles et non réversibles. Malheureusement, nous ne sommes pas arrivé à exploiter ceci pour montrer par exemple que si P est réversible, alors on peut obtenir un meilleur résultat que la proposition 2.

Dans le même ordre de considérations, notons que (7), où on remplace $g(q, \nu(P))$ par 1, permet de retrouver le lemme 2.6 de Diaconis et Saloff-Coste [4] dans les cas non réversibles : soit L un générateur markovien sur S, matriciellement $L = (L(x,y))_{x,y\in S}$ avec pour tous $x \neq y \in S$, $L(x,y) \geq 0$, et pour tout $x \in S$, $\sum_{y\in S} L(x,y) = 0$. Supposons L irréductible et soit $P = \exp(tL)$ pour un $t > 0$. Si $q : \mathbb{R}_+ \to \mathbb{R}_+$ est une application dérivable telle que $q(0) = q_0 \geq 2$, écrivons que

$$\begin{aligned}
&\frac{\|\exp(tL)f\|_{q(t)} - \|f\|_{q_0}}{t} \\
&\leq \frac{1}{q(t)} \|f\|_{q(t)}^{1-q(t)} \left[\frac{q(t) - q_0}{t} \frac{\mathcal{L}(f^{q(t)/2})}{q_0} - \frac{\mathcal{E}_{I-\exp(tL^*)\exp(tL)}(f^{q(t)/2}, f^{q(t)/2})}{t} \right]
\end{aligned}$$

Or on sait (voir la preuve du théorème 3.5 de Diaconis et Saloff-Coste [4]) que

$$\frac{d}{dt} \|\exp(tL)f\|_{q(t)} \bigg|_{t=0} = \|f\|_{q_0}^{1-q_0} \left[\frac{dq(t)}{dt}\bigg|_{t=0} \frac{\mathcal{L}(f^{q_0/2})}{q_0^2} - \mathcal{E}_{-L}(f, f^{q_0-1}) \right]$$

et d'autre part il est clair que

$$\lim_{t\to 0_+} \frac{\mathcal{E}_{I-\exp(tL^*)\exp(tL)}(f^{q(t)/2}, f^{q(t)/2})}{t} = \mathcal{E}_{-L^*-L}(f^{q_0/2}, f^{q_0/2}) = 2\mathcal{E}_{-L}(f^{q_0/2}, f^{q_0/2})$$

Ainsi en passant à la limite en t petit, on retrouve bien que pour tout $q_0 \geq 2$ et tout $f \in \mathcal{F}(S)$,

$$\mathcal{E}_{-L}(f^{q_0/2}, f^{q_0/2}) \leq \frac{q_0}{2} \mathcal{E}_{-L}(f, f^{q_0-1})$$

4 Evolution de l'entropie à temps discret

Un des moyens de montrer la convergence en loi vers l'équilibre des processus de Markov finis irréductibles et homogènes (à temps continu), est d'étudier l'évolution de l'entropie, par rapport à la mesure invariante μ du système, de la loi à un instant donné. En effet, il n'est pas très difficile de majorer la dérivée par rapport au temps de cette quantité par l'opposé de deux fois la forme de Dirichlet associée au symétrisé additif de l'opposé du générateur (si le processus est réversible, on peut même faire intervenir quatre fois cette forme de Dirichlet, voir Diaconis et Saloff-Coste [4]). Par le biais d'une inégalité de Sobolev-logarithmique, ceci permet d'obtenir une inégalité différentielle simple satisfaite par l'entropie, qui s'intègre pour prouver que cette quantité converge exponentiellement vite vers 0, la vitesse étant donnée par la constante de Sobolev-logarithmique considérée.

Nous allons voir comment ceci peut également s'adapter pour traiter de l'ergodicité des chaînes de Markov finies homogènes irréductibles et apériodiques.

Comme dans la section précédente, on a besoin d'une inégalité qui quantifie d'une certaine manière la stricte convexité de l'application $\mathbb{R}_+ \ni u \mapsto u\ln(u)$.

Lemme 5

Pour tous $t \geq 0$ et $s \geq -t$, on a

$$(t+s)\ln(t+s) \geq t\ln(t) + (1+\ln(t))s + (\sqrt{t+s} - \sqrt{t})^2$$

Démonstration :

A $t > 0$ fixé, soit pour $u = s/t$, $h(u)$ le membre de gauche moins celui de droite, le tout divisé par t. En dérivant deux fois on montre que h a le tableau de variations suivant :

$$\begin{array}{c|ccccccc}
u & -1 & & u_0 & & 0 & & +\infty \\
\hline
 & & & h(u_0) & & & & +\infty \\
h & & \nearrow & & \searrow & & \nearrow & \\
 & 0 & & & & 0 & &
\end{array}$$

pour une certaine valeur $-1 < u_0 < 0$, caractérisée par $\ln(1+u_0) - 1 + (1+u_0)^{-1/2} = 0$.

$\square$

En supposant que $I - PP^*$ satisfait une inégalité de Sobolev-logarithmique avec constante $\alpha(I - PP^*) > 0$, on en déduit le résultat suivant :

Proposition 6

Pour tout $m \in \mathcal{P}(S)$, on a

$$\operatorname{Ent}(mP) \leq (1 - \alpha(I - PP^*))\operatorname{Ent}(m)$$

Démonstration :

Désignons par f la densité de m par rapport à μ. On vérifie immédiatement que la densité de mP par rapport à μ est alors donnée par P^*f.

En appliquant le lemme 5 avec $t = P^*f(x)$ et $t + s = f(y)$, pour $x, y \in S$, puis en multipliant cette inégalité par $p^*(x, y)$ et en sommant sur $y \in S$, on obtient que pour tout $x \in S$,

$$P^*(f\ln(f))(x) \geq P^*f(x)\ln(P^*f(x)) + \sum_{y\in S} p^*(x,y)\left(\sqrt{f(y)} - \sqrt{P^*f(x)}\right)^2$$

Cependant, toujours à $x \in S$ fixé, on a comme dans la section précédente

$$\sum_{y\in S} p^*(x,y)\left(\sqrt{f(y)} - \sqrt{P^*f(x)}\right)^2 \;\; \geq \;\; P^*f(x) - \left(P^*(\sqrt{f})(x)\right)^2$$

ainsi en intégrant par rapport à μ en x et en utilisant l'invariance de μ par rapport à P^*, on fait apparaître que

$$(8) \qquad \operatorname{Ent}(mP) = \sum_{x\in S} P^*f(x)\ln(P^*f(x))\,\mu(x) \;\; \leq \;\; \operatorname{Ent}(m) - \mathcal{E}_{I-PP^*}(\sqrt{f}, \sqrt{f})$$

Il reste à utiliser l'inégalité de Sobolev-logarithmique

$$\mathrm{Ent}(m) = \mathcal{L}(\sqrt{f}) \leq \frac{1}{\alpha(I - PP^*)}\mathcal{E}_{I-PP^*}(\sqrt{f}, \sqrt{f})$$

pour conclure au résultat escompté.

□

Remarques :

a) Pour $0 < \nu \leq 1$, soit $g(\nu) = ((1-\nu)\ln(1-\nu)+\nu)/(\sqrt{1-\nu}-1)^2$, on vérifie sans difficulté que pour tous $t \geq 0$ et $s \geq -\nu t$, on a

(9) $$(t+s)\ln(t+s) \geq t\ln(t) + (1+\ln(t))s + g(\nu)(\sqrt{t+s}-\sqrt{t})^2$$

Remarquons que g est strictement décroissante et que $g(0_+) = 2$ et $g(1) = 1$. Une nouvelle coïncidence veut que la différence entre ces valeurs est la même que celle qui intervient dans le lemme 2.7 de Diaconis et Saloff-Coste [4] entre les situations réversibles et non réversibles. Notons aussi que l'on peut se servir d'une inégalité du type (9) avec un $\nu > 0$ dépendant de m et p, si l'on sait a priori que m charge tous les points. Puis quand on finit par savoir plus précisément que f est assez proche de $\mathbf{I}$, on peut utiliser plutôt une variante de l'inégalité de trou spectral à la place d'une véritable inégalité de Sobolev-logarithmique (cf. les inégalités énergie-entropie de Bakry [1], que l'on écrit ici sous la forme

$$\forall\, f \in \mathcal{F}(S),\ \|f\|_2 = 1, \qquad \mathcal{L}(f) \leq \alpha_{\mathcal{L}(f)}(I - PP^*)\mathcal{E}(f,f)$$

avec pour tout $h > 0$, $\alpha_h(I - PP^*) = h^{-1}\inf_{f\in\mathcal{F}(S)\setminus\mathrm{Vect}(\mathbf{I}),\,\|f\|_2=1,\,\mathcal{L}(f)=h}\mathcal{E}(f,f) \geq \alpha(I - PP^*)$, et on se sert du fait que $\lim_{h\to 0_+}\alpha(h)$ est égal à la moitié du trou spectral).

b) Par des arguments identiques à ceux de la fin de la section précédente, en prenant $P = \exp(tL)$ avec t petit et L générateur irréductible, on montre que (8) permet de retrouver le lemme 2.7 de Diaconis et Saloff-Coste [4] : pour tout $f \in \mathcal{F}(S)$, $f \geq 0$,

$$\mathcal{E}_{-L}(\sqrt{f}, \sqrt{f}) \leq \frac{1}{2}\mathcal{E}_{-L}(\ln(f), f)$$

(car il suffit de l'avoir pour toutes ces fonctions qui satisfont $\int f\, d\mu = 1$).

c) Le choix précédent de $t = P^*f(x)$ est commode et est suggéré par la preuve de l'inégalité de Jensen pour la fonction convexe $\mathbb{R}_+ \ni t \mapsto t\ln(t)$ et la probabilité donnée par $p^*(x, \cdot\,)$ à $x \in S$ fixé (qui permet de voir que l'entropie est toujours décroissante le long des itérés du semi-groupe agissant sur $\mathcal{P}(S)$), mais n'est pas optimal en général : en effectuant les mêmes calculs, on obtient pour tout $t \geq 0$,

$$\begin{aligned} P^*(f\ln(f))(x) &- \left(P^*(f)(x) - (P^*(\sqrt{f})(x))^2\right) \\ &\geq t\ln(t) + (1+\ln(t))(P^*(f)(x) - t) + \left(P^*(\sqrt{f})(x) - \sqrt{t}\right)^2 \end{aligned}$$

et à $x \in S$ fixé, on est donc amené à optimiser le membre de droite en t. On vérifie facilement que le maximum est atteint pour

$$t = \left(\frac{P^*(f)(x)}{P^*(\sqrt{f})(x)}\right)^2$$

si $P^*(\sqrt{f})(x) \neq 0$ (sinon de toute façon, l'inégalité précédente s'écrit $0 \geq 0$ pour tout $t > 0$). En remplaçant t par cette valeur, on obtient

$$\begin{aligned} & P^*(f\ln(f))(x) - \left(P^*(f)(x) - (P^*(\sqrt{f})(x))^2\right) \\ \geq \quad & -P^*(f)(x) + (P^*(\sqrt{f})(x))^2 + 2\ln\left(\frac{P^*(f)(x)}{P^*(\sqrt{f})(x)}\right) P^*(f)(x) \end{aligned}$$

c'est-à-dire

$$P^*(f\ln(f))(x) - \ln(P^*(f)(x))P^*(f)(x) \geq \ln\left(\frac{P^*(f)(x)}{(P^*(\sqrt{f})(x))^2}\right) P^*(f)(x)$$

En intégrant par rapport à μ en x, il apparaît donc que

$$\mathrm{Ent}(m) - \mathrm{Ent}(mP) \geq \int [\ln(P^*(f)) - \ln((P^*(\sqrt{f}))^2)] P^*(f)\, d\mu$$

ainsi en utilisant l'inégalité de concavité

$$\forall\, a, b > 0, \qquad \ln(a) - \ln(b) \geq \frac{1}{a}(a - b)$$

on retrouve la majoration (8).

De manière similaire le choix de $t = Pf(x)$ dans la section précédente n'est pas a priori le meilleur, cependant le problème d'optimisation en t n'est plus alors aussi trivial (et dépend de l'exposant q considéré).

Comme application immédiate de la proposition 6, on obtient une vitesse de convergence vers l'équilibre des chaînes de Markov irréductibles et apériodiques :

Corollaire 7

Soit p une matrice markovienne irréductible et apériodique. Alors pour toute probabilité initiale $m \in \mathcal{P}(S)$, tout $k \geq k(p)$ et tout $n \in \mathbb{N}$, on a

$$\mathrm{Ent}(mP^n) \leq (1 - \alpha(I - P^k P^{k*}))^{\lfloor n/k \rfloor} \mathrm{Ent}(m)$$

où $\lfloor \cdot \rfloor$ représente la partie entière.

En fait on aurait pu se servir de ce résultat pour montrer que $(i) \Rightarrow (ii)$ dans la proposition 1, du moins si l'on y suppose p sans point transient, car il permet de voir que si pour un $k \geq 1$, $\alpha(I - P^k P^{k*}) > 0$, alors on a unicité de la mesure invariante.

Notons également que contrairement aux processus de Markov à temps continu, ici l'entropie ne décroît pas nécessairement strictement en dehors de l'équilibre μ. Plus précisément, par le cas d'égalité dans l'inégalité de Jensen pour l'application strictement convexe $\mathbb{R}_+ \ni x \mapsto x\ln(x)$, il apparaît que $\mathrm{Ent}(mP) = \mathrm{Ent}(m)$ si et seulement si pour tout $x \in S$ fixé, la densité $f = dm/d\mu$ est constante sur l'ensemble $\{y \in S \,/\, p^*(x,y) > 0\}$ (ce qui équivaut aussi à $\mathcal{E}_{I-PP^*}(\sqrt{f}, \sqrt{f}) = 0$), ainsi dans l'exemple sur $\mathbb{Z}/n\mathbb{Z}$ de la section 2, si on part d'une masse de Dirac en 1, il faut attendre le temps $n - 1$ pour que l'entropie commence à décroître strictement à la prochaine transition (car $\mathrm{Ent}(mP) = \mathrm{Ent}(m)$ équivaut ici à $m(0) = 2m(n-1)$).

Si p est une matrice apériodique irréductible telle que $k(p) = 1$, on pourrait conjecturer que

$$\mathbb{N} \ni n \mapsto \frac{\mathrm{Ent}(mP^{n+1}) - \mathrm{Ent}(mP^n)}{\mathrm{Ent}(mP^n)}$$

est une application décroissante, ce qui permettrait de sortir du cadre d'ensembles d'états S finis et obtenir en général une décroissance exponentielle de l'entropie, mais avec une vitesse qui dépend de la loi initiale m par l'intermédiaire de la constante $(\mathrm{Ent}(mP) - \mathrm{Ent}(m))/\mathrm{Ent}(m) < 0$. Cependant ceci est faux : considérons à nouveau l'exemple ci-dessus sur $\mathbb{Z}/n\mathbb{Z}$ avec $n \geq 3$ et prenons $m = \frac{1}{5}\delta_{n-2} + \frac{4}{5}\delta_0$, on vérifie immédiatement que $\mathrm{Ent}(mP^2) = \mathrm{Ent}(mP) < \mathrm{Ent}(m)$. Quitte à considérer pour un $\epsilon > 0$ assez petit la matrice donnée par

$$\forall\, x, y \in S, \qquad p_\epsilon(x, y) = \frac{p(x, y) + \epsilon}{1 + n\epsilon}$$

on obtient alors un contre-exemple à la conjecture précédente.

5 Application aux algorithmes de recuit généralisés

L'ergodicité forte en loi des algorithmes de recuit généralisés à temps discret, si la température décroît suffisamment lentement, est bien connue (cf. les articles de Hwang et Sheu [11] et [12]). Dans [14], nous avons présenté une preuve très courte de ce résultat pour les équivalents de ces processus à temps continu (voir aussi la simplification donnée ultérieurement dans la section 2 de [16]), tout en retrouvant la constante critique qui y apparaît, en étudiant l'évolution de l'entropie à l'aide du comportement à basse température, donné par Holley et Stroock dans [10], de certaines constantes de Sobolev-logarithmiques. Nous allons voir ici comment les techniques décrites dans la section précédente étendent cette démonstration aux cas d'algorithmes à temps discret. Le résultat n'étant pas original, la justification de cette section se trouve plutôt dans l'espoir que la relative simplicité de la preuve lui permettra de s'adapter à des situations plus complexes que l'on rencontre dans la pratique.

On considère toujours un ensemble fini S muni d'un noyau de probabilités de transitions p irréductible (pas nécessairement apériodique), mais on suppose que l'on dispose en plus d'une fonction de coût V, définie sur $S \times S$ privé de sa diagonale et à valeurs dans $\overline{\mathbb{R}_+}$, compatible avec (S, p) (qui dans ce contexte est appelé noyau de communication a priori) :

$$\forall\, x \neq y \in S, \qquad p(x, y) = 0 \Longleftrightarrow V(x, y) = +\infty$$

Pour $\beta \geq 0$ donné (représentant l'inverse de la température), soit p_β le noyau irréductible défini par

$$\forall\, x, y \in S, \qquad p_\beta(x, y) = \begin{cases} \exp(-\beta V(x, y))p(x, y) & \text{, si } x \neq y \\ 1 - \sum_{z \neq x} p_\beta(x, z) & \text{, sinon.} \end{cases}$$

Pour ne pas considérer des situations homogènes, on suppose que V n'est pas seulement à valeurs dans $\{0, +\infty\}$. Ceci assure que pour tout $\beta > 0$, il existe un $x_0 \in S$ tel que $p_\beta(x_0, x_0) > 0$ et donc que p_β est de plus apériodique.

Si $m_0 \in \mathcal{P}(S)$ et une évolution de l'inverse de la température $\beta : \mathbb{N} \to \mathbb{R}_+$ sont donnés, on notera $X = (X_n)_{n\in\mathbb{N}}$ une chaîne de Markov inhomogène de loi initiale m_0 et dont le noyau de probabilités de transitions à tout instant $n \in \mathbb{N}$ est p_{β_n}. Quand l'évolution β est telle que $\lim_{n\to\infty} \beta_n = +\infty$, le processus précédent est appelé un algorithme de recuit généralisé.

Pour donner des conditions d'ergodicité forte en loi de ces chaînes, il apparaît une constante $c \geq 0$ que nous allons maintenant décrire, mais il faut pour cela faire quelques rappels sur les probabilités invariantes μ_β associées aux p_β pour $\beta \geq 0$: il existe (voir le lemme 3.1 p. 177 de Freidlin et Wentzell [6]) deux applications $\rho : S \to \mathbb{R}_+^*$ et $U : S \to \mathbb{R}_+$ (appelée parfois énergie virtuelle ou quasi-potentiel associé à (S, p, V)), telles que pour tout $x \in S$ fixé, on ait pour β grand,

$$\mu_\beta(x) \sim \rho(x) \exp(-\beta U(x))$$

Ceci montre notamment que quand β devient grand, les probabilités μ_β convergent vers une certaine mesure μ_∞ dont le support est l'ensemble des minima globaux de U.

Pour $x, y \in S$, soit $\mathcal{C}_{x,y}$ l'ensemble des chemins (dont les transitions sont permises par p) allant de x à y. Si $q = (q_i)_{0\leq i\leq n}$ est un tel chemin, son élévation est le nombre

$$e(q) = \begin{cases} \max_{0\leq i<n}(U(x_i) + V(x_i, x_{i+1})) & \text{, si } n \geq 1 \\ U(x) & \text{, si } n = 0 \text{ (et donc } x = y) \end{cases}$$

(où on a convenu que pour tout $z \in S$, $V(z, z) = 0$ si $p(z, z) > 0$), et on pose

$$H(x, y) = \min_{q\in\mathcal{C}_{x,y}} e(q)$$

cette quantité étant appelée la hauteur de communication de x à y (pour le triplet (S, p, V)). Enfin on note

$$c = \max_{x,y\in S} H(x, y) - U(x) - U(y) \geq 0$$

Le résultat classique que nous voulons retrouver s'énonce alors,

Proposition 8

Supposons que l'évolution β satisfasse

$$\begin{aligned} \lim_{n\to+\infty} \beta_n &= +\infty \\ \limsup_{n\to+\infty} \beta_n / \ln(n) &< 1/c \\ \limsup_{n\to+\infty} n\, |\beta_{n+1} - \beta_n| &< +\infty \end{aligned}$$

alors la loi de X_n converge pour n grand vers μ_∞.

Bien que la proposition reste vérifiée dans le cas où $c = 0$, cette situation n'est pas très intéressante en théorie du recuit simulé et on se restreindra désormais à $c > 0$.

Exemple :

Une évolution typique de l'inverse de la température qui satisfait ces conditions est celle donnée par

$$\forall\, n \in \mathbb{N}, \qquad \beta_n = b^{-1}\ln(1+n)$$

avec $b > c$. Profitons-en pour préciser que ces conditions ne sont pas optimales, on pourrait d'ailleurs les étendre un peu en reprenant plus soigneusement la preuve qui suit, cependant il semblerait qu'aucune de ces généralisations ne puissent permettre de traiter le cas où pour tout $n \in \mathbb{N}$, $\beta_n = c^{-1}\ln(1+n)$, or il est connu que cette évolution assure également la convergence en loi vers μ_∞ (cf. par exemple [17], mais il est à noter que la démonstration est alors beaucoup plus complexe et passe par une évaluation précise des temps et positions de sortie de certains ensembles, car nous ne sommes pas arrivé à la déduire de l'étude de l'évolution de fonctionnelles (comme l'entropie) définies sur $\mathcal{P}(S)$, et bien que la preuve soit donnée pour des processus à temps continu, elle peut s'adapter au cas du temps discret). Néanmoins, notons que si l'on prend pour tout $n \in \mathbb{N}$, $\beta_n = b^{-1}\ln(1+n)$, avec $0 < b < c$, alors la conclusion de la proposition 8 est fausse, car on sait (cf. Hwang et Sheu [12] ou [15] dont les résultats sont donnés pour des algorithmes de recuit simulé classiques à temps continu, mais qui peuvent également s'étendre à la situation d'algorithmes généralisés à temps discret) que dans ce cas la loi limite dépend effectivement de la loi initiale, et en ce sens, les conditions de cette proposition ne sont pas si mauvaises.

La preuve de cette proposition va se faire en plusieurs étapes. Tout d'abord, en reprenant les notations de la section 2, remarquons que $k(p_\beta)$ ne dépend pas de $\beta > 0$, on notera désormais cette quantité $\underline{k}$. Pour $l \geq \underline{k}$, soit $\lambda(I - P_\beta^l P_\beta^{l*})$ la plus petite valeur propre non nulle de l'opérateur $I - P_\beta^l P_\beta^{l*}$, qui est auto-adjoint dans $L^2(\mu_\beta)$, c'est-à-dire la plus grande constante $\lambda > 0$ telle que

$$(10) \qquad \forall\, f \in \mathcal{F}(S), \qquad \lambda \int (f - \mu_\beta(f))^2 \, d\mu_\beta \leq \int f(I - P_\beta^l P_\beta^{l*}) f \, d\mu_\beta$$

Commençons par rappeler le comportement pour β grand de ce trou spectral, et pour ceci introduisons les quantités suivantes :

$$\forall\, x \neq y \in S, \qquad W_l(x,y) = -\lim_{\beta \to +\infty} \beta^{-1} \ln((p_\beta^l p_\beta^{l*})(x,y))$$

puis

$$c(U, W_l) = \max_{x,y \in S} \min_{q=(q_i)_{0 \leq i \leq n} \in \mathcal{S}_{x,y}} \max_{0 \leq i < n} (U(q_i) + W_l(q_i, q_{i+1})) \wedge (U(q_{i+1}) + W_l(q_{i+1}, q_i))$$

où $\mathcal{S}_{x,y}$ représente l'ensemble des suites finies de S dont le premier élément est x et le dernier y.

Lemme 9

Il existe une certaine constante $K > 0$ assez grande, telle que pour tout $\beta \geq 0$,

$$K^{-1}\exp(-c(U, W_l)\beta) \leq \lambda(I - P_\beta^l P_\beta^{l*}) \leq K \exp(-c(U, W_l)\beta)$$

La majoration est donnée à titre indicatif, car elle n'est pas nécessaire à la preuve de la proposition 8.

Démonstration :

Il s'agit là d'une estimation classique, dès que l'on a remarqué que U est la fonctionnelle des grandes déviations satisfaites pour β grand par la mesure invariante de $p^l_\beta p^{l*}_\beta$, car celle-ci n'est autre que μ_β.

Pour la minoration, on renvoit à la preuve présentée dans [14], qui adapte celle de Holley et Stroock [10] donnée dans le cas d'algorithmes de recuit classiques réversibles et basée sur des inégalités de type Poincaré (plus précisément, on utilise ici que pour tous $x \neq y \in S$ tels que $W_l(x,y) < +\infty$, il existe $\widehat{\rho}_l(x,y) > 0$ tel que pour β grand, $p^l_\beta p^{l*}_\beta(x,y) \sim \widehat{\rho}_l(x,y)\exp(-\beta W_l(x,y))$).

Pour la majoration, on peut également reprendre la démonstration de Holley et Stroock [10] qui est basée sur le choix d'une bonne fonction f dans (10) : soit C un cycle relatif à la fonction de coût W_l (voir Hwang et Sheu [12] pour une définition récursive des cycles, ou la section 3 de [17] qui les fait apparaître comme les traces sur S des cycles usuels relatifs à la structure de Hajek virtuelle associée à $(S, \widehat{\rho}_l, W_l)$, qui est une notion définie un peu plus loin), dont le complémentaire contient au moins un des minima globaux de U et qui est de hauteur de sortie maximale parmi ceux-ci. Cette hauteur de sortie de C vaut alors $c(U, W_l)$ et il suffit de prendre pour f l'indicatrice de C dans (10) pour obtenir la majoration escomptée.

□

Pour tous $x \neq y \in S$, notons

$$H_{U,W_l}(x,y) = \min_{q=(q_i)_{0\leq i\leq n}\in \mathcal{S}_{x,y}} \max_{0\leq i\leq n-1} U(q_i) + W_l(q_i, q_{i+1})$$

et remarquons que

$$(11)\quad \min_{q=(q_i)_{0\leq i\leq n}\in \mathcal{S}_{x,y}} \max_{0\leq i<n} (U(q_i) + W_l(q_i, q_{i+1})) \wedge (U(q_{i+1}) + W_l(q_{i+1}, q_i)) = H_{U,W_l}(x,y)$$

car puisque μ_β est réversible pour le noyau $p^l_\beta p^{l*}_\beta$, on a pour tous $z, z' \in S$, $U(z) + W_l(z,z') = U(z') + W_l(z', z)$ (néanmoins il est connu que l'égalité (11) ci-dessus est toujours satisfaite, même si μ_β n'avait été qu'invariante).

L'étape suivante est cruciale et consiste à trouver un $k \geq 1$ tel que

$$c(W_k, U) = c$$

Pour décrire cet entier, on va faire quelques rappels sur la structure du paysage d'énergie associé à la fonction de coût V.

On dit que $x \in S$ est un minimum local, si et seulement si pour tout $y \in S$ tel que $U(y) < U(x)$, on a $H(x,y) > U(x)$, ainsi notamment les minima globaux de U, qui sont ceux qui annulent cette application, en sont. Il ne s'agit pas là d'une notion de minima locaux pour U (bien que ceux-ci en soient tous), mais par rapport à une fonction $\widetilde{U}$ qui prolonge naturellement U sur un graphe $(\widetilde{S}, \widetilde{p})$ qui lui même étend canoniquement le graphe (S, p), comme on l'a déjà présenté dans [17] : Pour tous $x \neq y \in S$ tels que $p(x,y) > 0$, soit $x \cdot y$ un nouveau point n'appartenant pas à S. On prend alors

$$\widetilde{S} = S \sqcup \{x \cdot y \,/\, x \neq y \in S \text{ tels que } p(x,y) > 0\}$$

et

$$\forall\, z,z' \in \widetilde{S}, \qquad \widetilde{p}(z,z') = \begin{cases} p(z,z') & \text{, si } z = z' \in S \\ p(z,z') & \text{, si } z \in S \text{ et } z' = z \cdot y, \text{ pour un } y \in S \\ 1 & \text{, si } z = x \cdot z' \text{ et } z' \in S, \text{ pour un } x \in S \\ 0 & \text{, sinon} \end{cases}$$

puis

$$\forall\, z \in \widetilde{S}, \qquad \widetilde{U}(z) = \begin{cases} U(z) & \text{, si } z \in S \\ U(x) + V(x,y) & \text{, si } z = x \cdot y, \text{ avec } x,y \in S \end{cases}$$

Cependant, si V dérivait déjà d'un potentiel $\widetilde{U}$ satisfaisant la condition de réversibilité faible de Hajek, c'est-à-dire si pour tous $x \neq y \in S$ tels que $p(x,y) > 0$, on a $V(x,y) = (\widetilde{U}(y) - \widetilde{U}(x))_+$, et si en posant pour tous $x,y \in S$,

$$H_{\widetilde{U}}(x,y) = \min_{q=(q_i)_{0\leq i\leq n} \in \mathcal{C}_{x,y}} \max_{0\leq i\leq n} \widetilde{U}(q_i)$$

on a que $H_{\widetilde{U}}$ est symétrique en ses deux arguments, alors cette construction est en fait inutile, car son but est de ramener d'une certaine manière toutes les situations à celle-ci.

Revenons au cas général, il apparaît que H correspond à la hauteur de communication classique relative au potentiel $\widetilde{U}$ (i.e. la restriction à $S \times S$ de $H_{\widetilde{U}}$, qui elle est définie sur $\widetilde{S} \times \widetilde{S}$), et on a vu dans [17] que cette application $\widetilde{U}$ satisfaisait une condition de réversibilité faible de Hajek (sur $(\widetilde{S},\widetilde{p})$), ce qui permit de retrouver que H est symétrique en ses deux arguments (ce résultat est dû initialement à Trouvé, cf. le lemme 1.33 p. 34 de [20], qui l'a prouvé par une méthode différente). En conséquence, $(\widetilde{S},\widetilde{p},\widetilde{U})$ sera appelé la structure de Hajek virtuelle associée à (S,p,V).

Notons L l'ensemble des minima locaux (qui est donc la trace sur S de l'ensemble des véritables minima locaux de $\widetilde{U}$ sur $\widetilde{S}$), sur lequel on introduit une relation d'équivalence $\bowtie$ par

$$\forall\, x,y \in L, \qquad x \bowtie y \iff H(x,y) = U(x) = U(y)$$

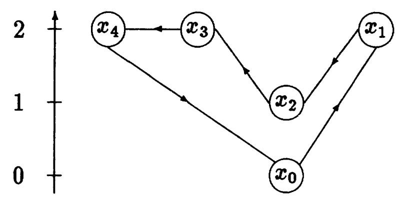
Appelons $C_1, \cdots, C_s$ les classes d'équivalence correspondantes. Du fait de l'hypothèse faite sur V, pour tout $1 \leq i \leq s$, il existe au moins un $x \in C_i$ tel que pour tout $\beta > 0$, $p_\beta(x,x) > 0$. On notera $\widehat{C}_i$ leur ensemble.

Soit $q = (q_i)_{0\leq i\leq n}$ un chemin (on désignera désormais par chemin, une suite finie d'éléments de S dont toutes les transitions sont permises par les p_β, pour $\beta > 0$, et non pas seulement par p, c'est-à-dire que l'on admet éventuellement certaines transitions supplémentaires qui restent en un même point, et pour tout $x \in S$, on conviendra d'ailleurs que $V(x,x) = 0$ ou $V(x,x) = +\infty$, suivant que $p_1(x,x) > 0$ ou $p_1(x,x) = 0$), on dira que q est descendant (respectivement montant) si

$$\forall\, 0 < i < n, \qquad V(q_i, q_{i+1}) = 0$$

(resp. si pour tout $0 \leq i < n-1$, $U(q_i) + V(q_i, q_{i+1}) = U(q_{i+1})$). Notons, en utilisant que pour tous $x \neq y \in S$ tels que $p(x,y) > 0$, on a $\widetilde{U}(x \cdot y) \geq U(x) \vee U(y)$, que les chemins qui leur sont canoniquement associés dans $(\widetilde{S},\widetilde{p}_\beta)$ sont effectivement descendants (resp. montants) pour $\widetilde{U}$, sauf éventuellement la première (resp. dernière) transition. Si V dérivait déjà d'un potentiel de Hajek, descendant (resp. montant) signifiera véritablement décroissant (resp. croissant) pour le potentiel.

Il est clair (en travaillant par exemple sur $(\widetilde{S}, \widetilde{p}_\beta))$, que pour tout $x \in S$, il existe au moins un chemin descendant (resp. montant) issu de x et aboutissant à un élément de $\widehat{C}_1 \sqcup \cdots \sqcup \widehat{C}_s$ (resp. partant de $\widehat{C}_1 \sqcup \cdots \sqcup \widehat{C}_s$ et aboutissant en x). Notons $D(x)$ (resp. $M(x)$) l'ensemble des éléments de $\widehat{C}_1 \sqcup \cdots \sqcup \widehat{C}_s$ pour lesquels il existe un tel chemin descendant (resp. montant). Sur l'exemple simple suivant on voit qu'il se peut que $D(x)$ et $M(x)$ soient différents :



On a $S = \{x_0, x_1, x_2, x_3, x_4\}$ et les flèches désignent les transitions strictement positives pour p (toutes ici de valeur 1), les autres transitions étant nulles. De plus, la fonction de coût V dérive du potentiel de Hajek dont les valeurs sont indiquées en ordonnée. Avec ces définitions il apparaît que $D(x_4) = \{x_0\}$ et que $M(x_4) = \{x_2\}$. On pourrait d'ailleurs remplacer x_3 et x_4 par un seul point, mais on les a distingués pour que cet exemple puisse aussi servir dans la remarque (b) de la fin de cette section.

Revenons au cas général, pour tous $x \in S$ et $y \in D(x)$ (resp. $y \in M(x)$) fixés, considérons $\widehat{\mathcal{C}}^{(\mathrm{d})}_{x,y}$ (resp. $\widehat{\mathcal{C}}^{(\mathrm{m})}_{y,x}$) l'ensemble des chemins descendants (resp. montants) allant de x à y (resp. de y à x) qui sont d'élévation minimale parmi de tels chemins. Rappelons par ailleurs que la longueur d'un chemin $q = (q_i)_{0 \leq i \leq n}$ est définie par $l(q) = n$ et notons

$$\begin{aligned} d(x,y) &= \min_{q \in \widehat{\mathcal{C}}^{(\mathrm{d})}_{x,y}} l(q) \\ m(y,x) &= \min_{q \in \widehat{\mathcal{C}}^{(\mathrm{m})}_{y,x}} l(q) \end{aligned}$$

On pose ensuite provisoirement

$$k = \max_{x \in S,\, y \in D(x)} d(x,y)$$

Proposition 10

On a $c(W_k, U) = c$, ce qui implique que $k \geq \underline{k}$, et ainsi il existe une constante $K \geq 1$ telle que

$$\forall\, \beta \geq 0, \qquad K^{-1} \exp(-c\beta) \leq \lambda(I - P_\beta^k P_\beta^{k*}) \leq K \exp(-c\beta)$$

Démonstration :

Pour tous $x \in S$ et $y \in D(x)$, il existe au moins un chemin de $\widehat{\mathcal{C}}^{(\mathrm{d})}_{x,y}$ de longueur k , quitte à obliger le chemin à rester plusieurs pas en y. Choisissons un tel chemin que l'on notera désormais $q^{(\mathrm{d})}(x,y) = (q_i^{(\mathrm{d})}(x,y))_{0 \leq i \leq k}$. Il sera commode aussi de faire intervenir son retourné $\check{q}^{(\mathrm{d})}(y,x) = (q^{(\mathrm{d})}_{k-i}(x,y))_{0\le i\le k}$, qui n'est plus nécessairement un chemin au sens précédent.

Rappelons également que si $q = (q_i)_{0\le i\le k}$ est une suite de k éléments de S, sa valuation par rapport à p_β (resp. p^*_β) est la quantité

$$v_\beta(q) = \prod_{i=0}^{k-1} p_\beta(q_i, q_{i+1}) \le p^k_\beta(q_0, q_k)$$

(resp. $v^*_\beta(q) = \prod_{i=0}^{k-1} p^*_\beta(q_i, q_{i+1}) \le p^{*k}_\beta(q_0, q_k)$).

Ainsi pour les valuations $v_\beta(q^{(\mathrm{d})}(x,y))$ et $v^*_\beta(\check{q}^{(\mathrm{d})}(y,x))$, on a le comportement suivant pour β grand :

$$\begin{aligned}
v_\beta(q^{(\mathrm{d})}(x,y)) &\sim \exp(-\beta V(x, q^{(\mathrm{d})}_0(x,y)))r^{(\mathrm{d})}(x,y)\\
&= \exp(-\beta[e(q^{(\mathrm{d})}(x,y)) - U(x)])r^{(\mathrm{d})}(x,y)\\
v^*_\beta(\check{q}^{(\mathrm{d})}(y,x)) &= \mu_\beta(x)v_\beta(q^{(\mathrm{d})}(x,y))/\mu_\beta(y)\\
&\sim \exp(-\beta[e(q^{(\mathrm{d})}(x,y)) - U(y)])\check{r}^{(\mathrm{d})}(y,x)
\end{aligned}$$

où $r^{(\mathrm{d})}(x,y)$ et $\check{r}^{(\mathrm{d})}(y,x)$ sont des quantités strictement positives.

Ceci nous amène à considérer une nouvelle fonction de coût $\check{W}$ (non nécessairement compatible avec (S,p)) dont l'intérêt apparaîtra par la suite et qui est définie par

$$\forall\ x \neq y \in S, \qquad \check{W}(x,y) = \begin{cases} e(q^{(\mathrm{d})}(x,y)) - U(x) & \text{, si } y \in D(x)\\ e(q^{(\mathrm{d})}(y,x)) - U(x) & \text{, si } x \in D(y)\\ +\infty & \text{, sinon}\end{cases}$$

que l'on prolonge sur la diagonale de $S\times S$ par $W(x,x) = 0$ ou $W(x,x) = +\infty$ suivant que $x \in \widehat{C}_1 \sqcup \cdots \sqcup \widehat{C}_s$ ou $x \notin \widehat{C}_1 \sqcup \cdots \sqcup \widehat{C}_s$.

Par ailleurs, pour tous $x \neq y$, soit

$$\begin{aligned}
\widehat{W}(x,y) &= -\lim_{\beta\to+\infty} \beta^{-1}\ln(p^k_\beta(x,y))\\
&= \min_{x_1,\cdots,x_{k-1}\in S} V(x,x_1) + V(x_1,x_2) + \cdots + V(x_{k-1},y)
\end{aligned}$$

Du fait que pour tous $z, z' \in S$, $U(z) + V(z,z') \ge U(z')$, on s'aperçoit sur cette dernière expression que

$$\forall\ x \neq y \in S, \qquad \widehat{W}(x,y) \ge H(x,y) - U(x)$$

Remarquons également que pour tous $x \neq y$,

$$W(x,y) = \min_{z\in S} \widehat{W}(x,z) + U(y) + \widehat{W}(y,z) - U(z)$$

où désormais W désigne W_k.

En utilisant que pour tout $z \in S$ fixé, on a

$$\begin{aligned}
\widehat{W}(x,z) + U(y) + \widehat{W}(y,z) - U(z) &\ge H(x,z) - U(x) + H(y,z) - U(z)\\
&= H(x,z) + H(z,y) - U(z) - U(x)\\
&\ge H(x,y) - U(x)
\end{aligned}$$

il apparaît que l'on a aussi

$$(12) \qquad \forall\, x,y \in S, \qquad W(x,y) \geq H(x,y) - U(x)$$

D'après les inégalités $H(x,y) \leq H(x,z) \vee H(z,y)$, valables pour tous $x,y,z \in S$, et (12), il est clair que $H_{U,W} \geq H$ (en convenant comme d'habitude que pour tout $x \in S$, $H_{U,W}(x,x) = U(x)$). Mais vérifions qu'en fait on a une égalité

$$H_{U,W} = H$$

En utilisant que $\lim_{\beta\to+\infty} \beta^{-1} \ln[v_\beta(q^{(\mathrm{d})}(y,y))] = 0 = \lim_{\beta\to+\infty} \beta^{-1} \ln[v^*_\beta(\check{q}^{(\mathrm{d})}(y,y))]$ pour tout $y \in \widehat{C}_1 \sqcup \cdots \sqcup \widehat{C}_s$, il apparaît facilement que $W \leq \check{W}$, d'où $H_{U,W} \leq H_{U,\check{W}}$ (cette dernière application étant définie de manière similaire, avec $\check{W}$ remplaçant W), et on peut donc se contenter de montrer que $H_{U,\check{W}} \leq H$.

Par construction des quantités précédentes et en se ramenant sur $(\widetilde{S},\widetilde{p})$, il suffit de voir que si une fonction de coût V dérive d'un potentiel U satisfaisant la condition de Hajek sur un graphe irréductible (S,p), alors pour tous $x \neq y \in S$, il existe une suite finie $q = (q_i)_{1\leq i\leq n} \in \mathcal{S}_{x,y}$ qui soit une succession de chemins véritablement descendants à des minima locaux (pour U) et de retournés de tels chemins à partir de ces minima locaux et telle que

$$\max_{0\leq i\leq n} U(q_i) = H(x,y)$$

Pour ceci, notons pour tout $1 \leq i \leq s$, A_i le domaine d'attraction de $\widehat{C}_i$, c'est-à-dire l'ensemble des $x \in S$ tels que $D(x) \cap \widehat{C}_i \neq \emptyset$. Soit $\widetilde{q} = (\widetilde{q}_j)_{0\leq j\leq m} \in \mathcal{C}_{x,y}$ un véritable chemin tel que $e(\widetilde{q}) = H(x,y)$. On construit q à partir de $\widetilde{q}$ de la manière suivante: soit $1 \leq i_0 \leq s$ tel que $x \in A_{i_0}$ et soit $m_0 = \max\{0 \leq j \leq m \,/\, \forall\, 0 \leq l \leq j,\, \widetilde{q}_l \in A_{i_0}\}$. On peut alors remplacer le tronçon $(\widetilde{q}_i)_{0\leq i\leq m_0}$ par une descente à partir de x en $\widehat{C}_{i_0}$ suivi du retourné d'une descente de $\widetilde{q}_{m_0}$ en $\widehat{C}_{i_0}$. Ensuite si $m_0 = m$ c'est gagné, et sinon il existe $1 \leq i_1 \leq s$ tel que $\widetilde{q}_{m_0} \in A_{i_1}$ et tel qu'en posant $m_1 = \max\{m_0 \leq j \leq m \,/\, \forall\, m_0 \leq l \leq j,\, \widetilde{q}_l \in A_{i_0}\}$, on ait $m_1 > m_0$ (ceci provenant du fait que l'on a nécessairement $U(\widetilde{q}_{m_0}) > U(\widetilde{q}_{m_0+1})$). On remplace alors le trajet $(\widetilde{q}_j)_{m_0\leq j\leq m_1}$ par une descente en $\widehat{C}_{i_1}$ suivi du retourné d'une autre telle descente. En procédant ainsi, on finit par arriver en y, d'où l'existence de q qui est la concaténation de ces suites finies de descentes et de retournées de descentes.

L'identité $H_{U,W} = H$ implique alors par le biais de (11) que $c(U,W) = c$.

□

Si l'on avait posé

$$k = \max_{x\in S,\, y\in M(x)} m(y,x)$$

en travaillant plutôt à partir des chemins montants et de leur retournés (ce qui amène à introduire pour $x \in S$ et $y \in M(x)$, des $q^{(\mathrm{m})}(y,x)$ et des $\check{q}^{(\mathrm{m})}(x,y)$) et des domaines de répulsion des $\widehat{C}_i$, $R_i = \{x \in S, \,/\, M(x) \cap \widehat{C}_i \neq \emptyset\}$ (en fait plutôt ceux associés à la structure de Hajek virtuelle), on montre d'une manière similaire que pour une certaine constante $K > 1$,

$$\forall\, \beta \geq 0, \qquad K^{-1} \exp(-c\beta) \leq \lambda(I - P^{k*}_\beta P^k_\beta) \leq K \exp(-c\beta)$$

Il reste à utiliser que $\lambda(I - P_\beta^k P_\beta^{k*}) = \lambda(I - P_\beta^{k*} P_\beta^k)$ pour conclure que

$$\text{(13)} \qquad \forall\, \beta \geq 0, \qquad K^{-1}\exp(-c\beta) \leq \lambda(I - P_\beta^k P_\beta^{k*}) \leq K\exp(-c\beta)$$

est toujours vérifié avec k défini par

$$\text{(14)} \qquad k = \left(\max_{x\in S,\, y\in D(x)} d(x,y)\right) \wedge \left(\max_{x\in S,\, y\in M(x)} m(y,x)\right)$$

ce qui admet aussi pour conséquence que $k \geq \underline{k}$.

Soit $\alpha(I - P_\beta^k P_\beta^{k*})$ la constante de Sobolev-logarithmique associée à l'opérateur $I - P_\beta^k P_\beta^{k*}$, c'est-à-dire la plus grande constante $\alpha > 0$ telle que

$$\forall\, f \in \mathcal{F}(S), \qquad \alpha \int f^2 \ln(f^2)\, d\mu_\beta \leq \int f(I - P_\beta^k P_\beta^{k*}) f\, d\mu_\beta$$

Un calcul de Holley et Stroock (cf. le théorème 3.21 p. 568 de [10]) permet de déduire de la minoration de $\lambda(I - P_\beta^k P_\beta^{k*})$ dans (13) et du comportement pour β grand de μ_β, qu'il existe une constante $K > 0$ assez grande, telle que

$$\forall\, \beta \geq 1, \qquad \alpha(I - P_\beta^k P_\beta^{k*}) \geq K^{-1}\beta^{-1}\exp(-c\beta)$$

(plus simplement encore et d'une manière peut-être plus explicite, ceci découle aussi du corollaire 5.4 de l'appendice de Diaconis et Saloff-Coste [4]).

D'autre part, la constante de Sobolev-logarithmique étant toujours majorée par la moitié du trou spectral de l'opérateur symétrique considéré (voir par exemple Diaconis et Saloff-Coste [4]), on a en fin de compte un encadrement de la forme

$$\text{(15)} \qquad \forall\, \beta \geq 1, \qquad K^{-1}\beta^{-1}\exp(-c\beta) \leq \alpha(I - P_\beta^k P_\beta^{k*}) \leq K\exp(-c\beta)$$

(notons que l'appendice de Diaconis et Saloff-Coste [4], qui donne une expression exacte pour la constante de Sobolev-logarithmique dans le cas où S n'est constitué que de deux points, permet de voir que l'on ne peut pas espérer avoir en général une estimation du type (13) pour $\alpha(I - P_\beta^k P_\beta^{k*})$, mais que la minoration (15) est du bon ordre).

On a donc prouvé l'estimation suivante :

Proposition 11

Si k est l'entier donné par (14), on a

$$\lim_{\beta\to+\infty} \beta^{-1}\ln(\alpha(I - P_\beta^k P_\beta^{k*})) = -c$$

Remarquons que de la même manière, on montre que $\lim_{\beta\to+\infty}\beta^{-1}\ln(\alpha(I - P_\beta^{k*} P_\beta^k)) = -c$, ce qui aurait pu être intéressant si on avait cherché à utiliser plutôt l'hypercontractivité, dans l'esprit de Concordet [2], ce qui permet plus précisément de montrer que la densité converge vers $\mathbb{1}$ et de donner des estimées de la vitesse de convergence.

Des arguments standards de comparaisons permettent d'étendre ce résultat, mais introduisons d'abord l'ensemble suivant :

$$R = \{x \in S \;/\; \lim_{\beta\to+\infty} p_\beta(x,x) > 0\}$$

(on a aussi, puisque pour tout $x \in S$ l'application $\mathbb{R}_+ \ni \beta \mapsto p_\beta(x,x)$ est soit constante soit strictement croissante, $R = \{x \in S \,/\, \exists\, \beta \geq 0$ avec $p_\beta(x,x) > 0\} = \{x \in S \,/\, \forall\, \beta > 0, p_\beta(x,x) > 0\}$).

Supposons maintenant que l'on dispose de k (qui désormais désignera la constante définie ci-dessus) matrices markoviennes $p(1,t), \cdots, p(k,t)$ dépendant d'un paramètre $t \geq 0$ telles que pour une certaine fonction $\beta : \mathbb{R}_+ \to \mathbb{R}_+$ vérifiant $\lim_{t\to+\infty} \beta_t = +\infty$, on ait les limites suivantes pour tout $1 \leq i \leq k$,

$$(16)\begin{cases} \forall\, x \neq y \in S, & \lim_{t\to+\infty} \beta_t^{-1} \ln(p(i,t)(x,y)) = -V(x,y) \\ \forall\, x \in S, & \lim_{t\to+\infty} \beta_t^{-1} \ln(p(i,t)(x,x)) = \begin{cases} 0 & , \text{si } x \in R \\ -\infty & , \text{si } x \notin R \end{cases} \end{cases}$$

Posons

$$\alpha[I - P(1,t)\cdots P(k,t)(P(1,t)\cdots P(k,t))^*] = \inf_{f\in\mathcal{F}\setminus\mathrm{Vect}(\mathbb{1})} \frac{\mathcal{E}_{I-P(1,t)\cdots P(k,t)(P(1,t)\cdots P(k,t))^*}(f,f)}{\mathcal{L}_{\widehat{\mu}_t}(f)}$$

avec $\widehat{\mu}_t$ la probabilité invariante de $p(1,t)\cdots p(k,t)$ (qui intervient aussi dans la forme de Dirichlet au numérateur comme mesure d'intégration et dans la définition de l'adjoint qui est compris ici dans $L^2(\widehat{\mu}_t)$), qui est bien définie du moins pour t assez grand.

Le résultat suivant se base sur la proposition 11 pour la généraliser.

Proposition 12

Les constantes de Sobolev-logarithmique précédentes satisfont

$$\lim_{t\to+\infty} \beta_t^{-1} \ln\left(\alpha[I - P(1,t)\cdots P(k,t)(P(1,t)\cdots P(k,t))^*]\right) = -c$$

Démonstration :

Il suffit de vérifier que

$$\lim_{t\to+\infty} \beta_t^{-1} \ln\left(\frac{\alpha[I - P(1,t)\cdots P(k,t)(P(1,t)\cdots P(k,t))^*]}{\alpha[I - P_{\beta_t}^k P_{\beta_t}^{k*}]}\right) = 0$$

Cependant la condition (16) permet de voir que

$$\forall\, x,y \in S, \qquad \lim_{t\to+\infty} \beta_t^{-1} \ln((p(1,t)\cdots p(k,t))(x,y)) = \lim_{t\to+\infty} \beta_t^{-1} \ln((p_{\beta_t}^k)(x,y))$$

Par la formule de Freidlin et Wentzell pour la mesure invariante, il apparaît alors que

$$(17) \qquad \forall\, x \in S, \qquad \lim_{t\to+\infty} \beta_t^{-1} \ln(\widehat{\mu}_t) = \lim_{t\to+\infty} \beta_t^{-1} \ln(\mu_{\beta_t}) = -U(x)$$

En conséquence, on a donc également pour tous $x, y \in S$,

$$\lim_{t\to+\infty} \beta_t^{-1} \ln((p(1,t)\cdots p(k,t))^*(x,y)) = \lim_{t\to+\infty} \beta_t^{-1} \ln(p_{\beta_t}^{k*}(x,y))$$

et

$$(18) \qquad \lim_{t\to+\infty} \beta_t^{-1} \ln([p(1,t)\cdots p(k,t)(p(1,t)\cdots p(k,t))^*](x,y)) = \lim_{t\to+\infty} \beta_t^{-1} \ln([p_{\beta_t}^k p_{\beta_t}^{k*}](x,y))$$

Grâce à (17) et (18), via une application de l'argument de comparaison du lemme 3.3 de Diaconis et Saloff-Coste [4], on obtient le résultat annoncé.

□

Nous pouvons maintenant appliquer les calculs de la section précédente. On considère donc une chaîne de Markov inhomogène X vérifiant les hypothèses de la proposition 8 et soit $0 \leq r < k$ fixé.

Pour simplifier les notations, convenons de poser pour tout $n \in \mathbb{N}$, $\widehat{m}_n$ la loi de X_{kn+r}, $\widehat{p}_n$ la matrice produit $p_{\beta_{kn+r}} \cdots p_{\beta_{kn+k+r-1}}$ et $\widehat{\mu}_n$ la probabilité invariante qui lui est associée. Notons que pour n grand, on a aussi

$$\forall\, x \in S, \qquad \widehat{\mu}_n(x) \sim \rho(x)\exp(-\beta_{kn+r}U(x))$$

En effet, il est facile de voir, d'après la troisième condition de la proposition 8, que les entrées des matrices $\widehat{p}_n$ et $p^k_{\beta_{kn+r}}$ ne diffèrent que de facteurs de la forme $(1+\mathcal{O}(1/n))$ (du moins dès que $\beta_{kn+r} > 0$, car sinon, si de plus $\beta_{kn+k+r-1} > 0$, il se peut que pour un $x \in S$, on ait $(p^k_0)(x,x) = 0$ et $\widehat{p}_n(x,x) > 0$), ainsi par la formule explicite de la mesure invariante donnée par Freidlin et Wentzell, il apparaît, en considérant $\mu_{\beta_{kn+r}}$ comme la mesure invariante de $p^k_{\beta_{kn+r}}$, que pour n grand, on a

$$(19) \qquad \forall\, x \in S, \qquad \widehat{\mu}_n(x) = \mu_{\beta_{kn+r}}(x)(1+\mathcal{O}(1/n))$$

d'où l'équivalence annoncée, et d'après la première condition de la proposition 8, le fait que

$$(20) \qquad \forall\, x \in S, \qquad \lim_{n\to\infty} \widehat{\mu}_n(x) = \mu_\infty = \begin{cases} \rho(x) & \text{, si } U(x) = 0 \\ 0 & \text{, sinon.} \end{cases}$$

Mais on peut déduire de (19) un autre résultat qui nous sera fort utile : en effet, de la même manière, on prouve que pour n grand,

$$\forall\, x \in S, \qquad \mu_{\beta_{k(n+1)+r}}(x) = \mu_{\beta_{kn+r}}(x)(1+\mathcal{O}(1/n))$$

et puisque pour tout $x \in S$, $\widehat{\mu}_{n+1}(x) = \mu_{\beta_{k(n+1)+r}}(x)(1+\mathcal{O}(1/n))$, on obtient que

$$\forall\, x \in S, \qquad \frac{\widehat{\mu}_{n+1}(x)}{\widehat{\mu}_n(x)} = 1 + \mathcal{O}(1/n)$$

puis que

$$\left\| \ln\left(\frac{\widehat{\mu}_{n+1}}{\widehat{\mu}_n}\right) \right\|_\infty = \mathcal{O}(1/n)$$

où $\|\cdot\|_\infty$ représente la norme uniforme sur $\mathcal{F}(S)$.

D'autre part, pour $n \in \mathbb{N}$, appelons $\widehat{\alpha}_n$ la constante de Sobolev-logarithmique associée à l'opérateur $I - \widehat{P}_n\widehat{P}_n^*$:

$$\widehat{\alpha}_n = \inf_{f\in\mathcal{F}\setminus\mathrm{Vect}(\mathbb{1})} \frac{\mathcal{E}_{I-\widehat{P}_n\widehat{P}_n^*}(f,f)}{\mathcal{L}_{\widehat{\mu}_n}(f)}$$

Une application de (12) montre que

$$(21) \qquad \lim_{n\to\infty} \beta^{-1}_{kn+r}\ln(\widehat{\alpha}_n) = -c$$

ce qui combiné avec la seconde hypothèse de la proposition 8, donne

$$\liminf_{n\to\infty}\frac{\ln(\widehat{\alpha}_n)}{\ln(n)}=\liminf_{n\to\infty}\frac{\ln(kn+r)}{\ln(n)}\frac{\beta_{kn+r}}{\ln(kn+r)}\frac{\ln(\widehat{\alpha}_n)}{\beta_{kn+r}}>-1$$

En notant $\gamma=-(\liminf_{n\to\infty}\ln(\widehat{\alpha}_n)/\ln(n)-1)/2<1$, il existe donc deux constantes $n_0\geq 0$ et $K>0$ telles que pour tout $n\geq n_0$, $\widehat{\alpha}_n\geq K/(n+1)^\gamma$. On peut d'ailleurs prendre $n_0=\max\{n\in\mathbb{N}\,/\,\beta_{kn+r}=0\}+1$ (ou $n_0=0$ si cet ensemble est vide) et trouver un $K>0$ correspondant, et comme on peut supposer β à valeurs dans $\mathbb{R}_+^*$ pour ce qui nous intéresse, on prendra ci-dessous $n_0=0$.

Nous disposons désormais de toutes les estimations nécessaires à l'étude de l'évolution de l'entropie $\mathrm{Ent}(\widehat{m}_n|\widehat{\mu}_n)$. On aura remarqué que pour prouver la proposition 8, il suffit, par le biais de la majoration (2) et de (20), de voir que $\lim_{n\to\infty}\mathrm{Ent}(\widehat{m}_n|\widehat{\mu}_n)=0$ (pour tout $1\leq r<k$ fixé). Pour ceci, on écrit

$$\begin{aligned}\mathrm{Ent}(\widehat{m}_{n+1}|\widehat{\mu}_{n+1})-\mathrm{Ent}(\widehat{m}_n|\widehat{\mu}_n) &= \mathrm{Ent}(\widehat{m}_{n+1}|\widehat{\mu}_{n+1})-\mathrm{Ent}(\widehat{m}_{n+1}|\widehat{\mu}_n)\\ &\quad+\mathrm{Ent}(\widehat{m}_{n+1}|\widehat{\mu}_n)-\mathrm{Ent}(\widehat{m}_n|\widehat{\mu}_n)\end{aligned}$$

Pour évaluer la première différence du membre de droite, notons que

$$\begin{aligned}a_n &\stackrel{\text{déf}}{=} |\mathrm{Ent}(\widehat{m}_{n+1}|\widehat{\mu}_{n+1})-\mathrm{Ent}(\widehat{m}_{n+1}|\widehat{\mu}_n)|\\ &= \left|\int\ln(\widehat{\mu}_{n+1}/\widehat{\mu}_n)\,d\widehat{m}_{n+1}\right|\\ &\leq \left\|\ln\left(\frac{\widehat{\mu}_{n+1}}{\widehat{\mu}_n}\right)\right\|_\infty\\ &\leq \frac{K'}{n+2}\end{aligned}$$

pour tout $n\in\mathbb{N}$, pour un bon choix de $K'>0$.

Quant à la seconde différence, d'après la section 4, on a

$$\mathrm{Ent}(\widehat{m}_{n+1}|\widehat{\mu}_n)-\mathrm{Ent}(\widehat{m}_n|\widehat{\mu}_n)\leq-\widehat{\alpha}_n\mathrm{Ent}(\widehat{m}_n|\widehat{\mu}_n)$$

d'où en fin de compte, en posant pour tout $n\in\mathbb{N}$, $b_n=1-\widehat{\alpha}_n\geq 0$,

$$\begin{aligned}&\mathrm{Ent}(\widehat{m}_{n+1}|\widehat{\mu}_{n+1})\\ &\leq b_n\mathrm{Ent}(\widehat{m}_n|\widehat{\mu}_n)+a_n\\ &\leq a_n+b_na_{n-1}+b_nb_{n-1}a_{n-2}+\cdots+b_nb_{n-1}\cdots b_1a_0+b_nb_{n-1}\cdots b_0\mathrm{Ent}(\widehat{m}_0|\widehat{\mu}_0)\end{aligned}$$

Il reste donc à voir que cette dernière expression tend vers 0 pour n grand. On utilise pour ceci les inégalités

$$b_n=1-\widehat{\alpha}_n\leq\exp(-\widehat{\alpha}_n)\leq\exp(-K/(n+1)^\gamma)$$

qui permettent de voir pour tout $0\leq j\leq n$,

$$b_n\cdots b_j\leq\exp(-K((n+1)^{-\gamma}+\cdots+(j+1)^{-\gamma}))\leq\exp(-K(1-\gamma)^{-1}((n+2)^{1-\gamma}-(j+1)^{1-\gamma})$$

On est donc ramené à voir que

$$\lim_{n\to\infty}\exp(-K(1-\gamma)^{-1}(n+2)^{1-\gamma})\left[\sum_{j=0}^{n}\frac{1}{j+2}\exp(K(1-\gamma)^{-1}(j+2)^{1-\gamma})\right]=0$$

et pour cela on majore la somme entre crochet par l'intégrale suivante (car l'application $\mathbb{R}_+^* \ni t \mapsto \exp(K(1-\gamma)^{-1}t^{1-\gamma})/t$ est décroissante puis croissante), que l'on intègre par parties

$$\begin{aligned}
&\int_1^{n+3} \frac{1}{t} \exp(K(1-\gamma)^{-1}t^{1-\gamma})\, dt \\
&= \int_1^{n+3} \frac{1}{t^{1-\gamma}} \frac{1}{t^\gamma} \exp(K(1-\gamma)^{-1}t^{1-\gamma})\, dt \\
&= K^{-1}([t^{\gamma-1}\exp(K(1-\gamma)^{-1}t^{1-\gamma})]_1^{n+3} + (1-\gamma)\int_1^{n+3} t^{\gamma-2}\exp(K(1-\gamma)^{-1}t^{1-\gamma})\, dt)
\end{aligned}$$

Choisissons $A \geq 1$ tel que $K(1-\gamma)^{-1}A^{1-\gamma} \geq 2$ et écrivons que

$$\begin{aligned}
&K^{-1}(1-\gamma)\int_1^{n+3} \frac{1}{t^{2-\gamma}} \exp(K(1-\gamma)^{-1}t^{1-\gamma})\, dt \\
&\leq K^{-1}(1-\gamma)\int_1^{A} \frac{1}{t^{2-\gamma}} \exp(K(1-\gamma)^{-1}t^{1-\gamma})\, dt + \frac{1}{2}\int_1^{n+3} \frac{1}{t} \exp(K(1-\gamma)^{-1}t^{1-\gamma})\, dt
\end{aligned}$$

ce qui fait apparaître que pour une certaine constante $K'' > 0$,

$$\int_1^{n+3} \frac{1}{t} \exp(K(1-\gamma)^{-1}t^{1-\gamma})\, dt \leq K'' + 2K^{-1}\frac{1}{(n+3)^{1-\gamma}} \exp(K(1-\gamma)^{-1}(n+3)^{1-\gamma})$$

puis que pour n grand,

$$\exp(-K(1-\gamma)^{-1}(n+2)^{1-\gamma})\left[\sum_{j=0}^{n} \frac{1}{j+2} \exp(K(1-\gamma)^{-1}(j+2)^{1-\gamma})\right] = \mathcal{O}\left(\frac{1}{n^{1-\gamma}}\right)$$

(et en étant un peu plus soigneux, il est possible de voir que le membre de gauche multiplié par $n^{1-\gamma}$ tend pour n grand vers K^{-1}), d'où le résultat annoncé, et une majoration de la vitesse de convergence.

Remarques :

a) Rappelons que les algorithmes de recuit simulé classiques sont ceux pour lesquels il existe une mesure μ telle que le noyau a priori p soit réversible par rapport à μ et pour lesquels V dérive d'un potentiel U.

Dans cette situation un encadrement de la forme (15) est en fait valable pour $\alpha(I - P_\beta P_\beta^*)$. En effet, comme d'habitude par les techniques de Holley et Stroock, on se ramène à une estimation du trou spectral λ_β de $I - P_\beta P_\beta^*$.

Or ici, il apparaît que la mesure invariante μ_β est la mesure de Gibbs donnée par

$$\forall\, x \in S, \qquad \mu_\beta(x) = \exp(-\beta U(x))\mu(x)$$

et que p_β est réversible par rapport à cette mesure, montrant ainsi que $P_\beta^* = P_\beta$.

On en déduit que $1 - \lambda_\beta$ est la plus grande valeur propre de P_β^2 différente de 1. Or d'après les calculs qui précèdent, il existe $k \in \mathbb{N}^*$ tel que la plus petite valeur propre non nulle $\widetilde{\lambda}_\beta$ de $I - P_\beta^k P_\beta^{k*} = I - P_\beta^{2k}$ satisfasse pour une certaine constante $K \geq 1$, un encadrement de la forme

$$\forall\, \beta \geq 1, \qquad K^{-1}\exp(-c\beta) \leq \widetilde{\lambda}_\beta \leq K\exp(-c\beta)$$

Mais en utilisant que toutes les valeurs propres de P_β^2 sont positives, il apparaît que pour β grand,

$$\widetilde{\lambda}_\beta = 1 - (1-\lambda_\beta)^k \sim k\lambda_\beta$$

et il en découle qu'il existe une autre constante $K > 1$ telle que

$$(22) \qquad \forall\, \beta \geq 1, \qquad K^{-1}\exp(-c\beta) \leq \lambda_\beta \leq K\exp(-c\beta)$$

Frigerio et Grillo (voir les équations (2.12) et (2.13) de [7]) affirment qu'il est aussi possible d'obtenir directement (22) à partir des calculs classiques de Holley et Stroock. Cependant remarquons que si l'on note pour $\beta \geq 0$, $-1 \leq \lambda_1(\beta) \leq \cdots \leq \lambda_{\mathrm{card}(S)-1}(\beta) < \lambda_{\mathrm{card}(S)}(\beta) = 1$ les valeurs propres de p_β ($\mathrm{Vect}(\mathbf{I})$ étant l'espace propre associé à $1 = \lambda_{\mathrm{card}(S)}(\beta)$), les estimations de [10] portent sur $1 - \lambda_{\mathrm{card}(S)-1}(\beta)$, or on a $\lambda_\beta = 1 - (\lambda^2_{\mathrm{card}(S)-1}(\beta) \vee \lambda_1^2(\beta))$, et il faudrait encore connaître le comportement de $1 + \lambda_1(\beta)$.

Mais montrons que celui-ci n'est en fait pas gênant, car

$$\lim_{\beta\to+\infty} 1 + \lambda_1(\beta) > 0$$

En effet, soit p_∞ la matrice markovienne limite pour β grand des p_β, par un résultat usuel de perturbation (cf. Kato [13], théorème 5.1 p. 107), $\lambda_1(\beta)$ converge vers la plus petite valeur propre de p_∞ pour β grand (même si p_∞ n'est plus nécessairement diagonalisable, contrairement ici aux p_β, on aura noté que les valeurs propres de p_∞ restent réelles). Il suffit donc de voir que -1 n'est pas valeur propre de p_∞. Pour ceci considérons par l'absurde un vecteur propre $f \in \mathcal{F}(S) \setminus \{0\}$ associé. On a donc pour tout $n \in \mathbb{N}$,

$$(23) \qquad P_\infty^n f = (-1)^n f$$

Or remarquons que les classes de récurrences de p_∞ sont exactement les classes d'équivalence pour la relation $\bowtie$ dans L, i.e. les $C_1, \cdots, C_s$. Posons pour tout $1 \leq i \leq s$, $p_\infty^{(i)}$ la restriction de p_∞ à C_i, il s'agit de matrices markoviennes irréductibles et apériodiques (car pour tout $x \in \widehat{C}_i$, $p_\infty(x,x) > 0$). On notera $\mu_\infty^{(i)}$ les probabilités invariantes associées (sur C_i). Alors, pour tout $x \in S$ fixé, on a

$$\lim_{n\to+\infty} P_\infty^n f(x) = \sum_{i=1}^{s} \gamma_i(x)\mu_\infty^{(i)}(f)$$

où pour tout $1 \leq i \leq s$, les limites suivantes existent bien

$$\gamma_i(x) = \lim_{n\to+\infty}\uparrow P_\infty^n(\mathbf{I}_{C_i})(x)$$

Ainsi en vertu de (23), $(-1)^n f$ doit converger pour n grand, c'est-à-dire que $f \equiv 0$, ce qui est la contradiction recherchée.

b) Souvent on peut aussi prendre un k plus petit que celui donné par (14) tout en étant assuré d'avoir encore (15). Par exemple, pour $x \in S$, $y \in D(x)$ et $q \in \widehat{\mathcal{C}}_{x,y}^{(\mathrm{d})}$, soit $j(q)$ la plus grande longueur d'un sous-chemin de q qui, en dehors de ses points de départ et d'arrivée, ne rencontre pas $R = \{z \in S \,/\, p_1(z,z) > 0\}$, posons ensuite

$$d(x,y) = \min_{q \in \mathcal{C}_{x,y}^{(\mathrm{d})}} j(q)$$

et définissons de manière similaire $m(y,x)$ pour $x \in S$ et $y \in M(x)$. On peut alors prendre dans les preuves précédentes le k construit comme dans (14) mais à partir de ces nouvelles quantités. Néanmoins même celui-ci n'est pas minimal en général, comme on peut le voir sur les algorithmes de recuit classiques.

On pourrait conjecturer que l'on a toujours pour une certaine constante $K \geq 1$,

$$\forall \beta \geq 0, \qquad K^{-1}\exp(-c\beta) \leq \lambda(I - P_\beta^{\underline{k}} P_\beta^{\underline{k}*}) \leq K\exp(-c\beta)$$

mais c'est faux :

En effet, considérons un triplet (S,p,V) tel que $\underline{k} > 1$, par exemple celui donné par une figure précédente. On rajoute à S un point $\overline{x}$ qui n'y appartenait pas, et on pose $\overline{S} = S \sqcup \{\overline{x}\}$. Soit $x_0 \in S$ tel que $U(x_0) = 0$ et tel qu'il existe $x_1 \in S$ avec $0 < V(x_0,x_1) < +\infty$. On définit un noyau de probabilités de transitions $\overline{p}$ par

$$\forall\, x,y \in \overline{S}, \qquad \overline{p}(x,y) = \begin{cases} p(x,y) & \text{, si } x,y \in S, \text{ avec } x \neq x_0 \\ p(x,y)/2 & \text{, si } x = x_0 \text{ et } y \in S \\ 1/2 & \text{, si } x = x_0 \text{ et } y = \overline{x} \\ 1/\mathrm{card}(S) & \text{, si } x = \overline{x} \text{ et si } y \in S \\ 0 & \text{, sinon.} \end{cases}$$

Pour $K \geq 0$, on munit $(\overline{S},\overline{p})$ de la fonction de coût $\overline{V}_K$ définie par

$$\forall\, x \neq y \in \overline{S}, \qquad \overline{V}_K(x,y) = \begin{cases} V(x,y) & \text{, si } x,y \in S \\ K & \text{, si } x = x_0 \text{ et } y = \overline{x} \\ 0 & \text{, si } x = \overline{x} \text{ et si } y \in S \\ +\infty & \text{, sinon.} \end{cases}$$

Par la formule de Freidlin et Wentzell, il apparaît que si K est assez grand, alors la restriction à S du quasi-potentiel $\overline{U}$ de $(\overline{S},\overline{p},\overline{V}_K)$ n'est autre que U et que $\overline{U}(\overline{x}) = K$. Par ailleurs, sur l'expression de c, il est clair que cette constante sera aussi celle associée à $(\overline{S},\overline{p},\overline{V}_K)$, quitte à prendre K encore plus grand. Or on se convainc facilement, puisque $\underline{k} > 1$, que pour tout $C > c$ assez grand, on peut choisir K toujours plus grand (et plus précisément, pour C grand, K et C seront équivalents) de manière à ce que $\lambda(I - \overline{P}_\beta \overline{P}_\beta^*)$ soit de l'ordre de $\exp(-C\beta)$ (au sens d'une minoration et d'une majoration du type (22)) et non pas de l'ordre de $\exp(-c\beta)$, ce qui fournit le contre-exemple cherché.

Pour d'autres remarques sur les preuves de cette section et tout particulièrement sur les manières d'affaiblir la seconde condition de (16) tout en gardant le bénéfice de la proposition 12, on renvoit à la prépublication 07-96 du Laboratoire de Statistique et Probabilités de l'Université Toulouse III qui correspond à cet article.

Références

[1] D. Bakry. L'hypercontractivité et son utilisation en théorie des semigroupes. In P. Bernard, editor, *Lectures on Probability Theory. Ecole d'Eté de Probabilités de Saint-Flour XXII-1992*, Lecture Notes in Mathematics 1581. Springer-Verlag, 1994.

[2] D. Concordet. Estimation de la densité du recuit simulé. *Annales de l'Institut Henri Poincaré*, 30(2):265–302, 1994.

[3] P. Diaconis and L. Saloff-Coste. Nash inequalities for finite Markov chains. A paraître dans *Journal of Theoretical Probability*, 1992.

[4] P. Diaconis and L. Saloff-Coste. Logarithmic Sobolev inequalities for finite Markov chains. Préprint, Octobre 1995.

[5] J.A. Fill. Eigenvalue bounds on convergence to stationarity for nonreversible Markov chains, with an application to the exclusion process. *The Annals of Applied Probability*, 1(1):62–87, 1991.

[6] M.I. Freidlin and A.D. Wentzell. *Random Perturbations of Dynamical Systems.* A Series of Comprehensive Studies in Mathematics 260. Springer-Verlag, 1984.

[7] A. Frigerio and G. Grillo. Simulated annealing with time-dependent energy function. *Mathematische Zeitschrift*, 213:97–116, 1993.

[8] L. Gross. Logarithmic Sobolev inequalities. *American Journal of Mathematics*, 97(4):1061–1083, 1976.

[9] R. Holley and D. Stroock. Logarithmic Sobolev inequalities and stochastic Ising models. *Journal of Statistical Physics*, 46:1159–1194, 1987.

[10] R. Holley and D. Stroock. Simulated annealing via Sobolev inequalities. *Communications in Mathematical Physics*, 115:553–569, 1988.

[11] C.R. Hwang and S.J. Sheu. Large-time behavior of perturbed diffusion Markov processes with applications to the second eigenvalue problem for Fokker-Planck operators and simulated annealing. *Acta Applicandae Mathematicae*, 19:253–295, 1990.

[12] C.R. Hwang and S.J. Sheu. Singular perturbed Markov chains and exact behaviors of simulated annealing processes. *Journal of Theoretical Probability*, 5(2):223–249, 1992.

[13] T. Kato. *Perturbation Theory for Linear Operators.* Classics in Mathematics. Springer, 1980.

[14] L. Miclo. Recuit simulé sans potentiel sur un ensemble fini. In J. Azéma, P.A. Meyer, and M. Yor, editors, *Séminaire de Probabilités XXVI*, Lecture Notes in Mathematics 1526, pages 47–60. Springer-Verlag, 1992.

[15] L. Miclo. Une étude des algorithmes de recuit simulé sous-admissibles. *Annales de la Faculté des sciences de Toulouse*, 4:819–877, 1995.

[16] L. Miclo. Sur les problèmes de sortie discrets inhomogènes. Préprint à paraître dans *The Annals of Applied Probability*, 1995.

[17] L. Miclo. Sur les temps d'occupations des processus de Markov finis inhomogènes à basse température. Préprint, 1995.

[18] O.S. Rothaus. Diffusion on compact Riemannian manifolds and logarithmic Sobolev inequalities. *Journal of Functional Analysis*, 42:102–109, 1981.

[19] D.W. Stroock. Logarithmic Sobolev inequalities for Gibbs states. In G. Dell'Antonio and U. Mosco, editors, *Dirichlet Forms*, Lecture Notes in Mathematics 1563, pages 194–228. Springer-Verlag, 1993.

[20] A. Trouvé. *Parallélisation massive du recuit simulé*. PhD thesis, Université Paris 11, Janvier 1993. Thèse de doctorat.

Laboratoire de Statistique et Probabilités
Université Paul Sabatier et CNRS
118, route de Narbonne
31062 Toulouse cedex, France