Séminaire de probabilités (Strasbourg)

Laurent Miclo

Recuit simulé sans potentiel sur un ensemble fini

Séminaire de probabilités (Strasbourg), tome 26 (1992), p. 47-60

<http://www.numdam.org/item?id=SPS_1992__26__47_0>

Recuit simulé sans potentiel sur un ensemble fini

Laurent Miclo

On va prouver la convergence d'algorithmes du recuit simulé généralisés (i.e. dont les générateurs ne "dérivent" pas nécessairement d'un potentiel), grâce à des techniques utilisant l'énergie libre, et on retrouvera aussi des résultats d'ergodicité pour des processus de Markov à temps continu, non nécessairement réversibles, sur des ensembles finis.

Présentation des résultats

Les algorithmes du recuit simulé sont actuellement l'objet de nombreuses recherches, souvent motivées par des problèmes concrets apparaissant en statistique (optimisation combinatoire, restauration d'images...), ce qui amène, en général, à considérer des processus stochastiques à temps discret (pour ce sujet, on renvoie à la thèse de Catoni [1] et aux références qu'elle contient). Mais on se propose, dans cet article, de démontrer la convergence d'algorithmes du recuit simulé en temps continu sur un ensemble fini. On étudie des algorithmes généraux (n'étant pas nécessairement construits à partir d'un potentiel donné a priori), par une méthode très simple, basée sur l'étude de l'évolution de l'énergie libre. Notons que la convergence des algorithmes du recuit simulé dérivant d'un potentiel (i.e. instantanément réversibles) a déjà été démontrée, par une technique différente, par Holley et Stroock dans [2]. Néanmoins, la méthode présentée ici peut se généraliser en dimension infinie, pour certains algorithmes du recuit simulé associés à des mesures de Gibbs (cf. [3]).

Soit E un ensemble fini, muni de la topologie et de la tribu discrète. Soient $Q = (q(x,y))_{x,y\in E}$ et $\Gamma = (\gamma(x,y))_{x,y\in E}$ deux matrices. On suppose que Q est positive (i.e. $\forall\, x,y \in E,\ q(x,y) \geq 0$), et qu'elle satisfait l'hypothèse suivante d'irréductibilité :

$$(C)\qquad \begin{cases} \forall\, x,y \in E, \text{ il existe une suite } (x_i)_{0\leq i\leq N} \text{ d'éléments de } E, \text{ telle que :} \\ \quad \bullet\ x_0 = x, x_N = y \\ \quad \bullet\ \forall\, 1 \leq i \leq N,\ q(x_{i-1}, x_i) > 0 \end{cases}$$

une telle suite $(x_i)_{0\leq i\leq N}$ sera appelée un chemin reliant x à y.

Pour $\beta \geq 0$ (qui représentera l'inverse de la température), on définit

$$\forall\, x,y \in E, \qquad q_\beta(x,y) = \exp(-\beta\gamma(x,y))q(x,y)$$

puis l'opérateur L_β sur l'ensemble (noté $F(E)$) des fonctions réelles définies sur E par,

$$\forall\, \phi \in F(E),\ \forall\, x \in E, \qquad L_\beta\phi(x) = \sum_{y\in E}(\phi(y) - \phi(x))q_\beta(x,y)$$

Les L_β seront les générateurs infinitésimaux des processus de Markov qu'on considèrera. Plus précisément, soit $\Omega = D([0,+\infty[;E)$ l'ensemble des trajectoires càdlàg. de $[0,+\infty[$ dans E. Pour $t \geq 0$, $X_t \ : \Omega \to E$ désigne l'application qui, à une trajectoire, associe sa position à l'instant t, et on note

$$\begin{aligned} \mathcal{F}_t &= \sigma\{X_s \ ; \ s \leq t\} \\ \mathcal{F} &= \sigma\{X_s \ ; \ s \geq 0\} \end{aligned}$$

Soient β une fonction de $C^1(\mathbb{R}_+;\mathbb{R}_+)$ (qui représentera l'évolution de l'inverse de la température), et m une probabilité sur M (qui sera la probabilité initiale). Il est connu (cf. [2]), qu'il existe une unique probabilité P_m sur $(\Omega,\mathcal{F})$, telle que

- $X_0(P_m) = m$
- $\forall\, \phi \in F(E), \qquad ((\phi(X_t) - \phi(X_0) - \int_0^t L_{\beta_s}\phi(X_s)\, ds)_{t\geq 0}, (\mathcal{F}_t)_{t\geq 0}, P_m)$ est une martingale.

On s'intéressera aux processus canoniques $(X_t)_{t\geq 0}$, sous certaines lois P_m, et on notera m_t la loi de X_t.

Faisons quelques rappels sur la mesure invariante, dans le cas où le processus est homogène dans le temps, c'est à dire dans le cas où la température est constante ; $\forall\, t \geq 0, \beta_t = \beta \geq 0$. Il est connu que sous (C), il existe une unique probabilité invariante μ_β pour le générateur L_β, i.e. telle que

$$\forall\, \phi \in F(E), \qquad \mu_\beta(L_\beta\phi) = 0 \tag{1}$$

On a la description suivante de μ_β, donnée par Wentzell et Freidlin (cf. [4], p. 177 ; Wentzell et Freidlin s'intéressent au cas d'une mesure invariante pour une chaîne de Markov, mais on a le même résultat pour un processus de sauts, comme on s'en rend immédiatement compte en écrivant, par exemple, (1) sous forme matricielle) :

pour $x \in E$, on désigne par x-graphe un ensemble de flèches $y \to z$, où $y \in E \setminus \{x\}$, $z \in E$, $z \neq y$, tel que

- $\forall\, y \neq x$, il existe une seule flèche partant de y
- il n'y a pas de cycle fermé dans le graphe.

On note G_x l'ensemble des x-graphes. Si $g \in G_x$, on lui associe le nombre $\pi_\beta(g)$ suivant

$$\pi_\beta(g) = \prod_{(y\to z)\in g} q_\beta(y,z)$$

et on pose, pour $x \in E$,

$$M_\beta(x) = \sum_{g\in G_x} \pi_\beta(g)$$

La mesure invariante est alors donnée par

$$\forall\, x \in E, \qquad \mu_\beta(x) = \left(\sum_{y\in E} M_\beta(y)\right)^{-1} M_\beta(x)$$

Pour faire apparaître clairement la dépendance de $M_\beta(x)$ en β, posons

$$I_x = \{g \in G_x \mid \pi_0(g) \neq 0\}$$

puis, pour $g \in I_x$,

$$\begin{aligned} r(g) &= \pi_0(g) \\ \rho(g) &= -\sum_{(y\to z)\in g} \gamma(y,z) \end{aligned}$$

On a alors, pour tout $x \in E$,

$$M_\beta(x) = \sum_{g\in I_x} r(g)\exp(\beta\rho(g))$$

Il apparaît ainsi (sous l'hypothèse (C)), que μ_β est strictement positif, et si on note, pour $x \in E$,

$$W(x) = -\max_{g\in I_x} \rho(g) + \max_{y\in E}\left(\max_{g\in I_y} \rho(g)\right)$$

(remarquons que $\min_{z\in E} W(z) = 0$), alors,

$$W(x) = -\lim_{\beta\to+\infty} \beta^{-1}\ln \mu_\beta(x)$$

On associe à W et à Γ un nombre $c \geq 0$ de la manière suivante :

pour $x, y \in E$, soit $\mathcal{C}_{x,y}$ l'ensemble des chemins reliant x à y. On définit l'élévation d'un tel chemin $S = (x_i)_{0\leq i\leq N} \in \mathcal{C}_{x,y}$ par

$$e(S) = \max_{1\leq i\leq N}(\min\{W(x_{i-1}) + \gamma(x_{i-1},x_i)\ ;\ W(x_i) + \gamma(x_i,x_{i-1})\}) - W(x) - W(y)$$

et on pose

$$c = \max_{x,y\in E}\left(\inf_{S\in\mathcal{C}_{x,y}} e(S)\right)$$

Pour $\beta \geq 0$ fixé, définissons l'énergie libre (aussi appelée entropie ou information de Kullback), relativement à μ_β, d'une probabilité p sur E par

$$I_\beta(p) = \sum_{x\in E} p(x)\ln\frac{p}{\mu_\beta}(x)$$

On se propose de démontrer les résultats suivants.

<u>Théorème 1</u>

Supposons l'hypothèse (C) satisfaite.
- Si la température est constante, $I_{\beta_0}(m_t)$ tend exponentiellement vite vers 0.
- Si pour t assez grand, on a $\beta_t = K^{-1}\ln(t)$, avec $K > c$, alors

$$\lim_{t\to\infty} I_{\beta_t}(m_t) = 0$$

Applications :

Soit p une probabilité sur E ; $I_\beta(p)$ mesure d'une certaine manière la distance entre p et μ_β, notamment on a (cf. [2]) :

$$\|p - \mu_\beta\| \leq 4\sqrt{2I_\beta(p)}$$

où $\| \, . \, \|$ désigne la variation totale.

Ainsi dans le cas où la température est constante, m_t tend exponentiellement vite vers μ_{β_0}.

D'autre part, vu la forme de μ_β, il est clair qu'il existe une probabilité μ_∞ sur E, portée par l'ensemble des minima de W, telle que

$$\lim_{\beta \to +\infty} \mu_\beta = \mu_\infty$$

Ainsi, si pour t assez grand, on a $\beta_t = K^{-1}\ln(t)$, avec $K > c$, alors

$$\lim_{t \to +\infty} m_t = \mu_\infty$$

Remarque :

La seconde partie du théorème 1 est une généralisation des résultats de convergence des algorithmes du recuit simulé. Pour ceux-ci, on est dans la situation suivante ;

- Q est réversible, c'est à dire qu'il existe une probabilité α chargeant tous les points de E telle que

$$\forall\, x, y \in E, \qquad \alpha(x)q(x,y) = \alpha(y)q(y,x)$$

- γ "dérive" d'un potentiel ; il existe une fonction $U \in F(E)$ telle que

$$\forall\, x, y \in E, \qquad \gamma(x,y) = (U(y) - U(x))_+$$

(sous ces deux hypothèses, on a $W = U - \min U$ et on retrouve la définition usuelle de l'élévation).

Démonstration du théorème 1

Le problème de martingale satisfait par P_m permet de voir que pour un $x \in E$ fixé, l'application $t \mapsto m_t(x)$ est C^2 sur $[0, +\infty[$, et pour $t \geq 0$, on a :

$$\begin{aligned}
\frac{dm_t(x)}{dt} &= \sum_{y \in E} m_t(y) L_{\beta_t} 1_{\{x\}}(y) \\
&= \sum_{y \in E} m_t(y) \sum_{z \in E} (1_{\{x\}}(z) - 1_{\{x\}}(y)) q_{\beta_t}(y,z) \\
&= -m_t(x) \sum_{z \neq x} q_{\beta_t}(x,z) + \sum_{y \neq x} m_t(y) q_{\beta_t}(y,x)
\end{aligned}$$

Montrons que ceci implique, sous l'hypothèse (C), que

$$\forall\, x \in E,\ \forall\, t > 0, \quad m_t(x) > 0$$

Soit $x_0 \in E$ tel que $m(x_0) > 0$. On dira que $x \in E$ est un p-voisin de x_0, si p est le plus petit entier n tel qu'il existe un chemin $(x_i)_{0 \leq i \leq n}$ reliant x_0 à x. On va montrer l'affirmation précédente par récurrence sur p.

Pour $p = 0$, on a $x = x_0$. Or,

$$\frac{dm_t}{dt}(x_0) \geq -(\sum_{z \neq x_0} q_{\beta_t}(x_0, z))m_t(x_0)$$

d'où :

$$\frac{d}{dt}\left(m_t(x_0)\exp(\int_0^t \sum_{z \neq x_0} q_{\beta_s}(x_0, z)\ ds)\right) \geq 0$$

ce qui implique que pour tout $t \geq 0$, on a :

$$m_t(x_0)\exp(\int_0^t \sum_{z \neq x_0} q_{\beta_s}(x_0, z)\ ds) \geq m(x_0)$$

puis :

$$m_t(x_0) > 0$$

Supposons que pour tout p-voisin y de x_0, on ait :

$$\forall\, t > 0, \qquad m_t(y) > 0$$

et soit x un $(p+1)$-voisin de x_0. Il existe alors un p-voisin y_0 de x_0 tel que $q_0(y_0, x) > 0$, et donc aussi tel que pour tout $t \geq 0$, on ait $q_{\beta_t}(y, x) > 0$. Ainsi, on a :

$$\begin{aligned}
\frac{d}{dt}\left(m_t(x)\exp(\int_0^t \sum_{z \neq x} q_{\beta_s}(x, z)\ ds)\right) &= \exp(\int_0^t \sum_{z \neq x} q_{\beta_s}(x, z)\ ds) \sum_{y \neq x} q_{\beta_t}(y, x)m_t(y) \\
&\geq \exp(\int_0^t \sum_{z \neq x} q_{\beta_s}(x, z)\ ds) q_{\beta_t}(y_0, x)m_t(y_0)
\end{aligned}$$

car on a bien $y_0 \neq x$. L'expression précédente et l'hypothèse de récurrence montrent alors que pour tout $t > 0$,

$$\frac{d}{dt}\left(m_t(x)\exp(\int_0^t \sum_{z \neq x} q_{\beta_s}(x, z)\ ds)\right) > 0$$

d'où, $\forall\, t > 0$,

$$m_t(x)\exp(\int_0^t \sum_{z \neq x} q_{\beta_s}(x, z)\ ds) > 0$$

puis :

$$m_t(x) > 0$$

Ce résultat montre déjà que l'application $t \mapsto I_{\beta_t}(m_t)$ est de classe C^1 sur $]0, +\infty[$. L'expression de la dérivée est donnée par la proposition suivante.

Proposition 2

$\forall\, t > 0,$

$$\begin{aligned}\frac{dI_{\beta_t}(m_t)}{dt} &= \sum_{x\in E} \ln\left(\frac{m_t}{\mu_{\beta_t}}(x)\right) \sum_{y\in E} \alpha_{\beta_t}(y,x)\left(\frac{m_t}{\mu_{\beta_t}}(y) - \frac{m_t}{\mu_{\beta_t}}(x)\right) \\ &\quad - \sum_{x\in E} m_t(x)\frac{d}{dt}\ln[\mu_{\beta_t}](x)\end{aligned}$$

où $\alpha_{\beta_t}(y,x) = \mu_{\beta_t}(y) q_{\beta_t}(y,x)$

Démonstration :

$\forall\, t > 0,$

$$\begin{aligned}\frac{dI_{\beta_t}(m_t)}{dt} &= \sum_{x\in E} \frac{dm_t}{dt}(x) \\ &\quad + \sum_{x\in E} \ln\left(\frac{m_t}{\mu_{\beta_t}}(x)\right)\frac{dm_t(x)}{dt} \\ &\quad - \sum_{x\in E} m_t(x)\frac{d}{dt}\ln[\mu_{\beta_t}](x)\end{aligned}$$

Mais le premier terme est nul et le troisième apparaît dans la proposition 2. Il nous reste donc à nous intéresser au second terme. Pour ceci notons que l'expression donnée pour $\frac{dm_t(x)}{dt}$, avant la proposition 2, s'écrit aussi

$$\frac{dm_t(x)}{dt} = -\frac{m_t}{\mu_{\beta_t}}(x)\sum_{y\neq x}\alpha_{\beta_t}(x,y) + \sum_{y\neq x}\frac{m_t}{\mu_{\beta_t}}(y)\alpha_{\beta_t}(y,x)$$

On se rend compte sur cette expression, que la propriété de la mesure μ_{β_t} d'être invariante pour le système, si l'inverse de la température est constant et égal à β_t, s'exprime par

$$\forall\, x \in E, \qquad \sum_{y\neq x}(\alpha_{\beta_t}(x,y) - \alpha_{\beta_t}(y,x)) = 0$$

c'est à dire :

$$\sum_{y\neq x}\alpha_{\beta_t}(x,y) = \sum_{y\neq x}\alpha_{\beta_t}(y,x)$$

On a donc :

$$\frac{dm_t(x)}{dt} = \sum_{y\neq x}\alpha_{\beta_t}(y,x)\left(\frac{m_t}{\mu_{\beta_t}}(y) - \frac{m_t}{\mu_{\beta_t}}(x)\right)$$

d'où le résultat annoncé.

□

Pour un $\beta \geq 0$ fixé, le graphe (E, α_β) n'est pas nécessairement symétrique. Cependant, la proposition suivante va permettre de se ramener à ce cas.

Proposition 3

Il existe une constante $\lambda > 0$ (qui ne dépend que du cardinal de E), telle que tout $\beta \geq 0$, si on pose

$$\forall\, x, y \in E, \qquad n_\beta(x,y) = \lambda(\alpha_\beta(x,y) + \alpha_\beta(y,x))$$

alors, $\forall\, f \in F(E),\ f > 0$,

$$\sum_{x,y \in E} \alpha_\beta(x,y) \ln(f(x))(f(x) - f(y)) \geq \sum_{x,y \in E} n_\beta(x,y)(\sqrt{f}(x) - \sqrt{f}(y))^2$$

Remarques :

a) L'hypothèse (C) n'est pas nécessaire pour ce lemme, ni d'ailleurs pour l'existence d'une mesure invariante (pour une température fixée). Mais on en a besoin pour l'unicité de cette mesure (et pour la description donnée par Wentzell et Freidlin), pour l'équivalence, pour tout $t > 0$ et toute probabilité initiale m, de m_t avec la mesure de comptage sur E, et pour assurer l'existence de certaines inégalités de Sobolev logarithmiques qui apparaîtront ultérieurement. Notons qu'en fait, il y a équivalence entre l'hypothèse (C) et la condition suivante : unicité de la mesure invariante (pour une température fixée) et équivalence de cette mesure avec la mesure de comptage sur E. Or il s'agit là de la condition dont on a besoin pour définir une "bonne" énergie libre.

b) Les valeurs diagonales de Q (et de Γ) ne jouent aucun rôle, on supposera donc que $\forall\, x \in E,\ q(x,x) = 0$, ce qui implique que pour tout $\beta \geq 0$ et tout $x \in E$, on ait $\alpha_\beta(x,x) = 0$.

Démonstration :

Soit $\beta \geq 0$ fixé. On commence par décomposer le graphe (E, α_β) en boucles.

On appelle boucle un graphe B de E (c'est à dire une matrice $(B(x,y))_{x,y \in E}$ dont tous les coefficients sont positifs) pour lequel il existe une famille $(x_i)_{1 \leq i \leq n}$ d'éléments distincts de E (par convention, on posera $x_{n+1} = x_1$) telle que

$$\forall\, x, y \in E, \qquad B(x,y) = \begin{cases} 1 & \text{, s'il existe } 1 \leq i \leq n \text{ tel que } x = x_i \text{ et } y = x_{i+1} \\ 0 & \text{, sinon.} \end{cases}$$

On va montrer qu'il existe une famille $(B_j)_{1 \leq j \leq N}$ de boucles et une famille $(\delta_j)_{1 \leq j \leq N}$ de réels strictement positifs, telles que

$$\forall\, x, y \in E, \qquad \alpha_\beta(x,y) = \sum_{1 \leq j \leq N} \delta_j B_j(x,y)$$

Cette décomposition découle du fait qu'en tout point $x \in E$, "l'intensité" (pour le graphe α_β) de ce qui arrive en x, i.e. $\sum_{y\in E} \alpha_\beta(y,x)$, vaut "l'intensité" de ce qui en repart, i.e. $\sum_{y\in E} \alpha_\beta(x,y)$.

En effet, soit x_0, x_1 des éléments de E tels que $\alpha_\beta(x_0,x_1) > 0$. On construit alors, par récurrence, une suite $(x_i)_{i\geq 0}$ d'éléments de E telle que $\forall\ i \geq 0$, $\alpha_\beta(x_i,x_{i+1}) > 0$ (ce qui est possible, car si on a construit x_i, $i \geq 1$, alors $\sum_{y\in E} \alpha_\beta(y,x_i) > 0$, d'où, $\sum_{y\in E} \alpha_\beta(x_i,y) > 0$ et on peut donc choisir x_{i+1} tel que $\alpha_\beta(x_i,x_{i+1}) > 0$).

Comme E est fini, on peut extraire de $(x_i)_{i\geq 0}$ une famille d'éléments distincts $(\tilde{x}_i)_{1\leq i\leq n}$ (successifs dans $(x_i)_{i\geq 0}$), telle qu'en posant $\tilde{x}_{n+1} = \tilde{x}_1$, on ait :

$$\forall\, 1 \leq i \leq n, \qquad \alpha_\beta(\tilde{x}_i, \tilde{x}_{i+1}) > 0$$

Soit B_1 la boucle formée par les $(\tilde{x}_i)_{1\leq i\leq n}$ et posons $\delta_1 = \inf_{1\leq i\leq n} \alpha_\beta(\tilde{x}_i,\tilde{x}_{i+1})$. On considère alors le graphe $\alpha_\beta^{(1)} = \alpha_\beta - \delta_1 B_1$, qui satisfait encore

$$\forall\, x \in E, \qquad \sum_{y\in E} \alpha_\beta^{(1)}(x,y) = \sum_{y\in E} \alpha_\beta^{(1)}(y,x)$$

Cependant, on a diminué le nombre de liens :

$$\mathrm{card}\{(x,y) \in E^2\ /\ \alpha_\beta(x,y) > 0\} > \mathrm{card}\{(x,y) \in E^2\ /\ \alpha_\beta^{(1)}(x,y) > 0\}$$

Si $\alpha_\beta^{(1)} \not\equiv 0$, on lui applique le traitement précédent, pour construire une boucle B_2 et un réel strictement positif δ_2 tels qu'en posant $\alpha_\beta^{(2)} = \alpha_\beta^{(1)} - \delta_2 B_2$, on ait encore un graphe du même type, mais dont on a diminué le nombre de liens. Ainsi, en un nombre fini k d'étapes, on aboutit à $\alpha_\beta^{(k)} \equiv 0$, d'où la décomposition annoncée.

Mais, d'après le lemme 4 ci-dessous, en posant $\lambda = \inf_{2\leq p\leq |E|} \alpha_p > 0$ (avec les notations de ce lemme), on a pour toute boucle B et toute $f \in F(E)$,

$$\sum_{x,y\in E} B(x,y) \ln(\sqrt{f}(x))(f(x) - f(y)) \geq \lambda \sum_{x,y\in E} B(x,y)(\sqrt{f}(x) - \sqrt{f}(y))^2$$

d'où :

$$\begin{aligned}
\sum_{x,y\in E} \alpha_\beta(x,y) \ln(f(x))(f(x) - f(y)) &= \sum_{1\leq i\leq N} \delta_i \sum_{x,y\in E} B_i(x,y) \ln(f(x))(f(x) - f(y)) \\
&\geq 2\lambda \sum_{1\leq i\leq N} \delta_i \sum_{x,y\in E} B_i(x,y)(\sqrt{f}(x) - \sqrt{f}(y))^2 \\
&= 2\lambda \sum_{x,y\in E} \alpha_\beta(x,y)(\sqrt{f}(x) - \sqrt{f}(y))^2 \\
&= \sum_{x,y\in E} n_\beta(x,y)(\sqrt{f}(x) - \sqrt{f}(y))^2
\end{aligned}$$

en posant $n_\beta(x,y) = \lambda(\alpha_\beta(x,y) + \alpha_\beta(y,x))$.

□

__Lemme 4__

Soit $n \geq 2$ un entier. Il existe une constante $\alpha_n > 0$, telle que

$$\forall\ (f_i)_{1\leq i\leq n} \in (\mathbb{R}_+^*)^n, \qquad \sum_{1\leq i\leq n} \ln f_i(f_i^2 - f_{i+1}^2) \geq \alpha_n \sum_{1\leq i\leq n} (f_i - f_{i+1})^2$$

avec la convention $f_{n+1} = f_1$.

Démonstration :

Soit $\alpha > 0$. On considère l'application J_α suivante :

$$\begin{array}{rcl} (\mathbb{R}_+^*)^n & \rightarrow & \mathbb{R} \\ f = (f_i)_{1\leq i\leq n} & \mapsto & \sum_{1\leq i\leq n} [\ln f_i(f_i^2 - f_{i+1}^2) - \alpha(f_i - f_{i+1})^2] \end{array}$$

Notons que J_α est homogène de degré 2 :

$$\forall\ \lambda > 0,\ \forall\ f \in (\mathbb{R}_+^*)^n, \qquad J_\alpha(\lambda f) = \lambda^2 J_\alpha(f)$$

ainsi, pour montrer que J_α est positive pour un bon choix de α, il suffit de montrer que J_α est positive sur

$$A = \{f \in (\mathbb{R}_+^*)^n \ /\ \max_{1\leq i\leq n} f_i = 1\}$$

Commençons par voir qu'il existe $1 > \epsilon > 0$ (dépendant de n) tel que

$$\forall\ f \in A, \qquad \min_{1\leq i\leq n} f_i \leq \epsilon \Rightarrow J_1(f) \geq 1$$

On démontre ceci par l'absurde ; supposons que
$\forall\ p \in \mathbb{N}^*,\ \exists\ f^{(p)} \in A$, tel que

$$J_1(f^{(p)}) < 1 \text{ et } \min_{1\leq i\leq n} f_i^{(p)} \leq \frac{1}{p}$$

On peut supposer que ce dernier minimum est toujours atteint en la première coordonnée, $f_1^{(p)} \leq \dfrac{1}{p}$.

Mais $\overline{A} = \{f \in (\mathbb{R}_+)^n \ /\ \max_{1\leq i\leq n} f_i = 1\}$ est compact, il existe donc une sous-suite de $(f^{(p)})$, encore notée $(f^{(p)})$, qui converge vers un élément $f \in \overline{A}$. On a alors :

$$f_1 = 0,\ \max_{1\leq i\leq n} f_i = 1$$

Montrons que ceci implique que

$$\lim_{p\rightarrow\infty} J_1(f^{(p)}) = +\infty$$

ce qui fournira la contradiction.

Soit $I = \{i \in \{1,\dots,n\} \;/\; f_i = 0\}$. Alors, $\forall\, i \in I$,

$$(*) \qquad \liminf_{p\to\infty}[\ln f_i^{(p)}((f_i^{(p)})^2 - (f_{i+1}^{(p)})^2)] \geq 0$$

En effet, si $f_{i+1}^{(p)} \geq f_i^{(p)}$, on a :

$$\ln f_i^{(p)}((f_i^{(p)})^2 - (f_{i+1}^{(p)})^2) \geq 0$$

et si $f_{i+1}^{(p)} < f_i^{(p)}$, on a :

$$|(f_i^{(p)})^2 - (f_{i+1}^{(p)})^2| \leq (f_i^{(p)})^2$$

ainsi, dans tous les cas :

$$\ln f_i^{(p)}((f_i^{(p)})^2 - (f_{i+1}^{(p)})^2) \geq -|(f_i^{(p)})^2 \ln f_i^{(p)}|$$

or, le terme de droite tend vers 0, quand p tend vers l'infini, d'où $(*)$.

D'autre part, $\sum_{i\notin I}[\ln f_i^{(p)}((f_i^{(p)})^2 - (f_{i+1}^{(p)})^2)] - \sum_{1\leq i\leq n}(f_i^{(p)} - f_{i+1}^{(p)})^2$ tend vers une limite finie, quand p tend vers l'infini. Il reste donc à montrer qu'il existe $i \in I$ tel que

$$\lim_{p\to\infty}[\ln f_i^{(p)}((f_i^{(p)})^2 - (f_{i+1}^{(p)})^2)] = +\infty$$

Mais, il suffit, pour cela, de choisir $i \in I$ tel que $f_{i+1} > 0$ ce qui est toujours possible puisque

$$I \neq \emptyset \text{ et } I \neq \{1,\dots,n\}$$

Posons $A_\epsilon = \{f \in A \;/\; \min_{1\leq i\leq n} f_i \geq \epsilon\}$. Pour tout $0 < \alpha \leq 1$, on a $J_1 \leq J_\alpha$, ainsi, le minimum de J_α sur A est atteint en un point de A_ϵ. Prenons $\alpha = \epsilon$, on va montrer que si f est un minimum de J_ϵ sur A_ϵ, alors $f = (1,\dots,1)$. Or $J_\epsilon(1,\dots,1) = 0$, on aura donc terminé la démonstration du lemme.

Soit f un tel minimum, et supposons qu'il existe $i \in \{1,\dots,n\}$ tel que $f_i < 1$. Puisqu'on a aussi $f_i > 0$, on doit avoir $\dfrac{\partial J_\epsilon(f)}{\partial f_i} = 0$. Or,

$$\frac{\partial J_\epsilon(f)}{\partial f_i} = \frac{1}{f_i}(f_i^2 - f_{i+1}^2) + 2f_i \ln f_i - 2f_i \ln f_{i-1} - 2\epsilon(f_i - f_{i+1}) - 2\epsilon(f_i - f_{i-1})$$

Définissons les deux fonctions suivantes sur $\mathbb{R}_+^*$:

$$\begin{aligned} g(x) &= x^{-2} - 2\epsilon x^{-1} + 2\epsilon - 1 \\ h(x) &= -2\ln x + 2\epsilon x - 2\epsilon \end{aligned}$$

La relation précédente peut alors aussi s'écrire

$$(**) \qquad g\left(\frac{f_i}{f_{i+1}}\right) = h\left(\frac{f_{i-1}}{f_i}\right)$$

Cependant,

$$\begin{aligned} g'(x) &= 2x^{-2}(\epsilon - x^{-1}) \\ h'(x) &= 2(\epsilon - x^{-1}) \end{aligned}$$

ainsi g et h sont strictement décroissantes sur $]0, \epsilon^{-1}[$. Mais on a aussi $g(1) = h(1) = 0$, d'où pour $x \in]0, \epsilon^{-1}[$,

$$\begin{aligned} g(x) \geq 0 &\Leftrightarrow x \leq 1 \\ h(x) \geq 0 &\Leftrightarrow x \leq 1 \end{aligned}$$

Du fait que pour tout $1 \leq i \leq n$, on a $\dfrac{f_i}{f_{i+1}} \leq \epsilon^{-1}$ (car $f_{i+1} \geq \epsilon$), on en déduit alors, via (**), que

$$\frac{f_i}{f_{i+1}} \leq 1 \Leftrightarrow \frac{f_{i-1}}{f_i} \leq 1$$

Cette relation est donc satisfaite dès que $f_i < 1$. Mais puisque $\max_{1 \leq j \leq n} f_j = 1$, il existe $k, p \in \mathbb{N}^*$ tels que

$$\begin{aligned} &\forall\, 0 \leq l < k,\ f_{i+l} < 1 \qquad \text{et} \qquad f_{i+k} = 1 \\ &\forall\, 0 \leq q < p,\ f_{i-q} < 1 \qquad \text{et} \qquad f_{i-p} = 1 \end{aligned}$$

(quitte à poser $f_0 = f_n,\ f_{-1} = f_{n-1}, \ldots$ et $f_{n+2} = f_2,\ f_{n+3} = f_3, \ldots$)

Or, $\dfrac{f_{i+k-1}}{f_{i+k}} = f_{i+k-1} \leq 1$, d'où $\dfrac{f_{i+p-2}}{f_{i+p-1}} \leq 1$, car $f_{i+k-1} < 1$, et de proche en proche, on arrive à

$$\frac{f_{i-p}}{f_{i-p+1}} \leq 1$$

ce qui est absurde, car $1 < \dfrac{1}{f_{i-p+1}}$.

Ainsi, $f = (1, \ldots, 1)$

□

La proposition 3 va nous permettre d'appliquer certaines inégalités de Sobolev logarithmiques, pour obtenir, à partir de la proposition 2, une inégalité différentielle satisfaite par $I_{\beta_t}(m_t)$. En effet, posons pour $\beta \geq 0$ et $x, y \in E$,

$$N_\beta(x, y) = (\mu_\beta)^{-1}(x) n_\beta(x, y)$$

il est clair, par symétrie de $n_\beta(x, y)$ en x, y, que N_β est symétrique par rapport à μ_β.

D'autre part, le graphe (E, N_β) est aussi irréductible (i.e. il satisfait l'hypothèse (C)).

On sait alors que μ_β satisfait des inégalités de Sobolev logarithmiques relativement au graphe N_β, c'est à dire qu'il existe un nombre $a_\beta > 0$ tel que $\forall\, f \in F(E)$,

$$\int f^2 \ln f^2 \, d\mu_\beta \leq a_\beta \sum_{x,y} N_\beta(x, y)(f(y) - f(x))^2 \mu_\beta(x) + \int f^2 \, d\mu_\beta \ln \left[\int f^2 \, d\mu_\beta \right]$$

Plus précisément, on peut aisément adapter les résultats présentés dans [2] (dans le cas de l'algorithme du recuit classique), pour obtenir l'estimée suivante, a_β étant la plus petite constante telle que les inégalités précédentes soient vérifiées,

$$\limsup_{\beta\to+\infty} \beta^{-1} \ln a_\beta \leq c$$

où c est la constante définie avant l'énoncé du théorème 1.

En effet, vu la forme de μ_β, il existe une constante $M > 0$ telle que pour tout $x \in E$, on ait :

$$\exp(-M\beta) \leq \mu_\beta(x) \leq \exp(M\beta)$$

ainsi, d'après la démonstration du théorème 3.21 de [2], il suffit de montrer que si pour $\beta \geq 0$, c_β est la plus petite constante telle que

$$\forall\, \phi \in F(E), \qquad \int (f - \langle f\rangle_\beta)^2 \, d\mu_\beta \leq c_\beta \sum_{x,y} N_\beta(x,y)(f(y) - f(x))^2 \mu_\beta(x)$$

(où $\langle f\rangle_\beta = \int f \, d\mu_\beta$), alors,

$$\limsup_{\beta\to+\infty} \beta^{-1} \ln c_\beta \leq c$$

Pour prouver ceci, on reprend le lemme 2.7 de [2].

Pour $x, y \in E$, soit $p^{x,y} \in \mathcal{C}_{x,y}$, tel que

$$e(p^{x,y}) = \inf_{S \in \mathcal{C}_{x,y}} e(S)$$

on notera $p^{x,y} = (p_i^{x,y})_{1\leq i\leq n(x,y)}$. Pour $z, w \in E$, on pose

$$\chi_{z,w}(x,y) = \begin{cases} 1 & \text{, s'il existe } 1 \leq i < n(x,y) \text{ tel que } p_i^{x,y} = z \text{ et } p_{i+1}^{x,y} = w \\ 0 & \text{, sinon} \end{cases}$$

et

$$N = \sup_{x,y\in E} n(x,y)$$

Soit $f \in F(E)$, on a :

$$\begin{aligned}
2\int (f - \langle f\rangle_\beta)^2 \, d\mu_\beta &= \sum_{x\in E}\sum_{y\in E} (f(y) - f(x))^2 \mu_\beta(y)\mu_\beta(x) \\
&= \sum_{x\in E}\sum_{y\in E} \Big(\sum_{i=1}^{i=n(x,y)} f(p_i^{x,y}) - f(p_{i+1}^{x,y})\Big)^2 \mu_\beta(y)\mu_\beta(x) \\
&\leq \sum_{x\in E}\sum_{y\in E} n(x,y) \sum_{i=1}^{i=n(x,y)} (f(p_i^{x,y}) - f(p_{i+1}^{x,y}))^2 \mu_\beta(y)\mu_\beta(x) \\
&\leq N \sum_{x\in E}\sum_{y\in E}\sum_{z\in E}\sum_{w\in E} \chi_{z,w}(x,y)(f(z) - f(w))^2 n_\beta(w,z) \left[\frac{\mu_\beta(y)\mu_\beta(x)}{n_\beta(w,z)}\right] \\
&\leq \tilde{c}_\beta \sum_{z\in E}\sum_{w\in E} (f(z) - f(w))^2 n_\beta(w,z)
\end{aligned}$$

où :

$$\tilde{c}_\beta = N \max_{z,w\in E} \sum_{x\in E}\sum_{y\in E} \frac{\chi_{z,w}(x,y)}{n_\beta(w,z)}\mu_\beta(y)\mu_\beta(x)$$

avec la convention que $\dfrac{\chi_{z,w}(x,y)}{n_\beta(w,z)} = 0$, si $\chi_{z,w}(x,y) = 0$.

Cependant,

$$\begin{aligned}
\lim_{\beta\to+\infty} \beta^{-1}\ln\mu_\beta(x) &= -W(x)\\
\lim_{\beta\to+\infty} \beta^{-1}\ln\mu_\beta(y) &= -W(y)
\end{aligned}$$

et si $\chi_{z,w}(x,y) \neq 0$, on a :

$$\begin{aligned}
\lim_{\beta\to+\infty} \beta^{-1}\ln n_\beta(z,w) &= \lim_{\beta\to+\infty} \beta^{-1}\ln[\alpha_\beta(z,w)+\alpha_\beta(w,z)]\\
&= \lim_{\beta\to+\infty} \beta^{-1}\ln[\max\{\alpha_\beta(z,w)\ ;\ \alpha_\beta(w,z)\}]\\
&= \max\{\lim_{\beta\to+\infty} \beta^{-1}\ln\alpha_\beta(z,w)\ ;\ \lim_{\beta\to+\infty} \beta^{-1}\ln\alpha_\beta(w,z)\}\\
&= \max\{-W(z)-\gamma(z,w)\ ;\ -W(w)-\gamma(w,z)\}\\
&= -\min\{W(z)+\gamma(z,w)\ ;\ W(w)+\gamma(w,z)\}
\end{aligned}$$

Ainsi, pour $x, y \in E$,

$$\lim_{\beta\to+\infty} \beta^{-1}\ln\left[N\max_{z,w\in E}\frac{\chi_{z,w}(x,y)}{n_\beta(w,z)}\mu_\beta(y)\mu_\beta(x)\right] = e(p^{x,y})$$

d'où :

$$\lim_{\beta\to+\infty} \beta^{-1}\ln\tilde{c}_\beta = c$$

et on en déduit le résultat annoncé, du fait que $c_\beta \leq \dfrac{\tilde{c}_\beta}{2}$, pour tout $\beta \geq 0$.

On applique alors les inégalités de Sobolev logarithmiques précédentes, pour un $t \geq 0$ fixé, avec $\beta = \beta_t$ et $f = \sqrt{\dfrac{m_t}{\mu_{\beta_t}}}$, pour obtenir :

$$I_{\beta_t}(m_t) \leq a_{\beta_t}\sum_{x,y\in E} n_\beta(x,y)\left(\sqrt{\frac{m_t}{\mu_{\beta_t}}}(x) - \sqrt{\frac{m_t}{\mu_{\beta_t}}}(y)\right)^2$$

Ainsi, les propositions 2 et 3 nous montrent que $I_{\beta_t}(m_t)$ satisfait l'inégalité différentielle

$$\frac{dI_{\beta_t}(m_t)}{dt} \leq -\sum_{x\in E} m_t(x)\frac{d}{dt}\ln[\mu_{\beta_t}](x) - a_{\beta_t}^{-1}I_{\beta_t}(m_t)$$

Ceci permet de conclure que $I_{\beta_0}(m_t)$ converge exponentiellement vite vers 0, dans le cas où la température est constante, car l'inégalité précédente se réduit alors à

$$\frac{dI_{\beta_0}(m_t)}{dt} \leq -a_{\beta_0}^{-1}I_{\beta_0}(m_t)$$

Pour prouver la seconde partie du théorème 1, il nous reste à estimer, pour $x \in E$,

$$\frac{d}{dt}\ln[\mu_{\beta_t}](x)$$

mais pour $\beta \geq 0$, on a :

$$\frac{d\ln(M_\beta(x))}{d\beta} = \frac{d}{d\beta}\ln\left[\sum_{g\in I_x} r(g)\exp(\beta\rho(g))\right] = \sum_{g\in I_x}\rho(g)\frac{r(g)\exp(\beta\rho(g))}{\sum_{h\in I_x} r(h)\exp(\beta\rho(h))}$$

ce qui permet de voir qu'en posant

$$R = \max_{g\in\bigcup_{x\in E} I_x}\rho(g) - \min_{g\in\bigcup_{x\in E} I_x}\rho(g)$$

on a pour tout $x \in E$ et tout $\beta \geq 0$,

$$\left|\frac{d}{d\beta}\ln\mu_\beta(x)\right| \leq R$$

Ainsi, $\forall\, t \geq 0$,

$$\frac{dI_{\beta_t}(m_t)}{dt} \leq -a_{\beta_t}^{-1} I_{\beta_t}(m_t) + R\left|\frac{d\beta_t}{dt}\right|$$

Mais d'après un lemme classique sur les inégalités différentielles, pour obtenir :

$$\lim_{t\to+\infty} I_{\beta_t}(m_t) = 0$$

il suffit d'avoir (du fait que $I_{\beta_t}(m_t) \geq 0$) :

$$\begin{aligned}\int^{+\infty} a_{\beta_t}^{-1}\,dt &= +\infty\\ \lim_{t\to+\infty}\left|\frac{d\beta_t}{dt}\right| a_{\beta_t} &= 0\end{aligned}$$

ce qui est bien satisfait, si on prend pour t assez grand,

$$\beta_t = K^{-1}\ln(t) \qquad \text{, avec } K > c.$$

Références

[1] O. Catoni, *Thèse de doctorat de l'Université Paris XI*, Laboratoire de Statistiques Appliquées, 1990.

[2] R. Holley et D. Stroock, Annealing via Sobolev Inequalities, *C.M.P.*, vol. 115, 1988, p. 553-569.

[3] L. Miclo, Comportement asymptotique de l'énergie libre spécifique. Application à l'ergodicité et au recuit simulé en dimension infinie, *Annales de l'I.H.P.*, vol. 28 n° 2, 1992.

[4] A.D. Wentzell et M.I. Freidlin, *Random Perturbations of Dynamical Systems*, Springer-Verlag, 1984.