ELSEVIER Stochastic Processes and their Applications 65 (1996) 281–298

stochastic processes and their applications

Recuit simulé partiel

Laurent Miclo[1]

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

Received April 1996

Abstract

Let $(L_\theta)_{\theta \in N}$ be a family of elliptic diffusion operators on a compact and connected smooth manifold M, whose terms of first order are indexed by a parameter θ living in N, the n-dimensional torus. For each fixed θ, we associate to L_θ its invariant probability μ_θ. Let f be a smooth function on $M \times N$ and define for $\theta \in N$, $F(\theta) = \int f(x,\theta)\mu_\theta(\mathrm{d}x)$. We study partial simulated annealing algorithms (using only quite directly L_θ and f) to find the global minima of F. This paper presents a new proof of the convergence of these algorithms, using $n+2$ partial entropies associated naturally to the problem. This approach is simpler than the one exposed previously in (Miclo, 1994), which furthermore was restricted to the case $n = 1$, but we need to speed up much more the diffusion interacting with the simulated annealing algorithm (and in practice, this is embarrassing).

Keywords: Diffusions interacting with simulated annealing processes; Evolution of partial entropies

Résumé

On s'intéresse aux interactions entre des diffusions et des algorithmes de recuit simulé qui permettent de trouver les minima globaux (sur N, le tore de dimension n) de potentiels de la forme $F(\theta) = \int f(x,\theta)\mu_\theta(\mathrm{d}x)$ où μ_θ est la probabilité invariante associée à une diffusion non dégénérée sur une variété riemannienne compacte et connexe, dont la dérive est paramétrée par $\theta \in N$. On présente une nouvelle démonstration de la convergence de ces algorithmes, plus simple que celle restreinte au cas $n = 1$ que nous avions déjà exposée dans (Miclo, 1994), car basée sur l'étude des évolutions conjointes de $n + 2$ entropies partielles associées naturellement au problème. Cependant, cette méthode exige que l'on accélère fortement la diffusion adjointe, ce qui en pratique est gênant.

1. Description de l'algorithme

On va reprendre ici le problème décrit dans (Miclo, 1994) et en présenter une approche plus simple. Cette méthode permet de traiter le cas où l'espace des

[1] Ce travail a été effectué à l'Institut de Recherche Mathématique Avancée, Université Louis Pasteur et CNRS, Strasbourg, France.

PII S0304-4149(96)00107-X

paramètres est un tore de dimension quelconque (et non plus seulement le cercle), mais elle exige que les paramètres déterministes γ_t et h_t, dont dépend l'évolution de l'algorithme à l'instant t, convergent exponentiellement plus vite vers 0 en temps grand que la température partielle β_t^{-1} (alors que dans l'article précédent, il suffisait qu'ils convergent polynômialement plus vite).

Rappelons succinctement les motivations, fournies par Etienne Pardoux, qui sont à l'origine du problème que l'on se pose. On considère un système dont l'évolution stochastique dans $\mathbb{R}^m$ est décrite par

$$\mathrm{d}\tilde{X}_t^\theta = d\tilde{B}_t + \tilde{b}(\tilde{X}_t^\theta, \theta)\,\mathrm{d}t$$

où $\tilde{B}$ est un mouvement brownien sur $\mathbb{R}^m$ indépendant de la condition initiale $\tilde{X}_0$ et où $\tilde{b}(\cdot,\theta) : \mathbb{R}^m \to \mathbb{R}^m$ est un champ de vecteurs régulier paramétré par $\theta \in \mathbb{R}^n$, élément par lequel on se permet de contrôler la diffusion $(\tilde{X}^\theta)$.

Pour une certaine fonction régulière $\tilde{f}$ donnée sur $\mathbb{R}^m \times \mathbb{R}^n$, on est amené à s'intéresser à la quantité

$$I_T(\theta) = \frac{1}{T}\int_0^T \tilde{f}(\tilde{X}_s^\theta, \theta)\,\mathrm{d}s$$

et plus précisément, à sa limite quand T devient très grand. Le théorème ergodique affirme (sous de bonnes hypothèses sur $\tilde{b}$) que $I_T(\theta)$ converge p.s. vers

$$I_\infty(\theta) = \int \tilde{f}(x,\theta)\tilde{\mu}_\theta(\mathrm{d}x)$$

où $\tilde{\mu}_\theta$ est l'unique probabilité invariante associée à la diffusion $(\tilde{X}^\theta)$. On cherche à minimiser globalement en $\theta \in \mathbb{R}^n$ cette fonction (contrôle asymptotique) et le but de l'article est de présenter un algorithme qui effectue cette opération en ne faisant intervenir que d'une manière relativement simple les deux données a priori, $\tilde{b}$ et $\tilde{f}$ (on ne se permet notamment pas d'utiliser les probabilités $\tilde{\mu}_\theta$, pour lesquelles il n'existe pas en général de formules explicites). Pour simplifier, on se placera dans le cas où tous les espaces considérés sont compacts.

Soient M une variété riemannienne compacte et connexe, et $N = T^n$, où $T = \{z \in \mathbb{C} / |z| = 1\}$ est le cercle unitaire (que l'on identifiera souvent avec $\mathbb{R}/2\pi\mathbb{Z}$), muni de sa structure riemannienne usuelle. Comme d'habitude, on notera $\langle\,,\rangle$, $|\,|$, ∇, div, Δ et λ, le produit scalaire, la norme, le gradient, la divergence, le laplacien et la probabilité associés à la structure riemannienne de M.

On se donne une famille de champs de vecteurs sur M, $b(\cdot,\theta)$, paramétrée par $\theta \in N$, et on suppose que l'application

$$M \times N \to TM$$

$$(x,\theta) \mapsto b(x,\theta)$$

est de classe C^∞.

Pour tout $\theta \in N$ fixé, on associe au champ $b(\cdot,\theta)$ l'unique probabilité μ_θ sur M invariante pour l'opérateur L_θ défini sur $C^2(M)$ par

$$L_\theta \cdot = \tfrac{1}{2}\Delta \cdot + \langle b(x,\theta), \nabla \cdot \rangle$$

c'est-à-dire l'unique probabilité μ_θ qui satisfait pour toute fonction $\phi \in C^2(M)$, $\mu_\theta(L_\theta \phi) = 0$.

Soit $f \in C^\infty(M \times N)$, on pose, pour $\theta \in N$,

$$F(\theta) = \int_M f(x,\theta)\mu_\theta(\mathrm{d}x).$$

Comme expliqué précédemment, le problème est de trouver les minima globaux de cette fonction F par une procédure simple à appliquer.

On est amené, pour cela, à s'intéresser à des algorithmes de recuit partiel sur $M^{n+1} \times N$. Introduisons tout d'abord quelques notations: pour $h > 0$, $x, y \in M$, $\theta = (\theta_1, \ldots, \theta_n) \in N$ et $1 \leqslant i \leqslant n$, posons

$$l_h^{(i)}(x,y,\theta) = \frac{\partial}{\partial \theta_i} f(x,\theta) + \frac{1}{h}\,[f(y,\theta) - f(x,\theta)]$$

et

$$e_i = (\delta_{ij})_{1 \leqslant j \leqslant n} \in (\mathbb{R}/2\pi\mathbb{Z})^n$$

où δ_{ij} désigne le symbole de Kronecker. On convient aussi de noter $e_0 = 0 \in (\mathbb{R}/2\pi\mathbb{Z})^n$.

Remarque. Si l'on ne veut pas calculer la dérivée partielle de f par rapport à θ_i, on peut aussi considérer l'application

$$\tilde{l}_h^{(i)}(x,y,\theta) = \frac{1}{h}\,[f(y,\theta + he_i) - f(x,\theta)]$$

car il est facile de vérifier que ce qui suit reste valable, avec les $(\tilde{l}_h^{(i)})_{1 \leqslant i \leqslant n}$ à la place des $(l_h^{(i)})_{1 \leqslant i \leqslant n}$.

Pour pouvoir interpréter les opérateurs que l'on considère comme les (restrictions aux fonctions de classe C^2 de) générateurs de processus de diffusions, il est commode d'écrire le laplacien sur M sous la forme

$$\frac{1}{2}\Delta = \sum_{\alpha=1}^{r} A_\alpha^2 + A_0$$

où les A_α, pour $0 \leqslant \alpha \leqslant r$, sont des champs de vecteurs réguliers sur M (considérés ici comme des opérateurs de dérivation sur M). On peut toujours mettre le laplacien sous cette forme, quitte à prendre r suffisamment grand, d'ailleurs les opérateurs L_θ sont parfois donnés directement sous une telle forme,

$$L_\theta = \sum_{\alpha=1}^{r} \tilde{A}_\alpha^2 + \tilde{A}_{0,\theta}$$

où l'opérateur $\sum_{\alpha=1}^{r} \tilde{A}_\alpha^2$ est elliptique (il est bien connu qu'il existe alors une structure riemannienne sur M dans laquelle les opérateurs L_θ prennent la forme donnée initialement, $L_\theta = \frac{1}{2}\Delta + \langle b(x,\theta), \nabla \rangle$, voir Ikeda et Watanabe (1981)), et on peut dans ce cas prendre $A_\alpha = \tilde{A}_\alpha$, pour $1 \leqslant \alpha \leqslant r$.

Un exemple du type d'algorithme à considérer peut se décrire sur $M^{n+1} \times T^n$ par (où $\circ$ représente l'intégration stochastique au sens de Stratonovitch)

$$\begin{aligned}
dX_t^{(0)} &= \gamma_t^{-1/2} \sum_{\alpha=1}^{r} A_\alpha(X_t^{(0)}) \circ dB_t^\alpha + \gamma_t^{-1}[A_0(X_t^{(0)}) + b(X_t^{(0)}, \Theta_t)]\, dt,\\
dX_t^{(1)} &= \gamma_t^{-1/2} \sum_{\alpha=1}^{r} A_\alpha(X_t^{(1)}) \circ dB_t^\alpha + \gamma_t^{-1}[A_0(X_t^{(1)}) + b(X_t^{(1)}, \Theta_t + h_t e_1)]\, dt,\\
dX_t^{(2)} &= \gamma_t^{-1/2} \sum_{\alpha=1}^{r} A_\alpha(X_t^{(2)}) \circ dB_t^\alpha + \gamma_t^{-1}[A_0(X_t^{(2)}) + b(X_t^{(2)}, \Theta_t + h_t e_2)]\, dt,\\
&\ \ \vdots \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (1)\\
dX_t^{(n)} &= \gamma_t^{-1/2} \sum_{\alpha=1}^{r} A_\alpha(X_t^{(n)}) \circ dB_t^\alpha + \gamma_t^{-1}[A_0(X_t^{(n)}) + b(X_t^{(n)}, \Theta_t + h_t e_n)]\, dt,\\
d\Theta_t^{(1)} &= dW_t^{(1)} - \beta_t l_{h_t}^{(1)}(X_t^{(0)}, X_t^{(1)}, \Theta_t)\, dt,\\
&\ \ \vdots\\
d\Theta_t^{(n)} &= dW_t^{(n)} - \beta_t l_{h_t}^{(n)}(X_t^{(0)}, X_t^{(n)}, \Theta_t)\, dt
\end{aligned}$$

où $(B^1, \ldots, B^r, W^{(1)}, \ldots, W^{(n)})$ est un mouvement brownien sur $\mathbb{R}^r \times T^n$ (indépendant de la v.a. initiale $(X_0^{(0)}, \ldots, X_0^{(n)}, \Theta_0^{(1)}, \ldots, \Theta_0^{(n)})$) et où β, γ et $h \in C^1([0, +\infty[, \mathbb{R}_+^*)$ sont des évolutions à déterminer pour que Θ_t tende, en temps grand, vers les minima globaux de F (a priori, β, γ et h seront tels que $\lim_{t\to+\infty} \beta_t^{-1} = 0$, $\lim_{t\to+\infty} \gamma_t = 0$ et $\lim_{t\to+\infty} h_t = 0$).

Remarquons que le générateur $\mathbb{L}_t$ à l'instant $t \geqslant 0$ de ce processus ne dépend pas uniquement de b, f, β_t, γ_t et h_t, mais aussi du choix des champs de vecteurs $(A_\alpha)_{0 \leqslant \alpha \leqslant r}$ nécessaires à la décomposition du laplacien (on peut vérifier sur certains exemples qu'il est possible effectivement d'obtenir des $\mathbb{L}_t$ différents). Cependant, le fait important est que pour tout $t \geqslant 0$, pour tout $f \in C^2(M \times N)$ et pour tout $0 \leqslant i \leqslant n$, l'action du générateur $\mathbb{L}_t$ sur la fonction f_i définie sur $M^{n+1} \times N$ par

$$\forall (x_0, x_1, \ldots, x_n, \theta) \in M^{n+1} \times N, \quad f_i(x_0, x_1, \ldots, x_n, \theta) = f(x_i, \theta)$$

soit donnée par

$$\begin{aligned}
\mathbb{L}_t[f_i](x_0, x_1, \ldots, x_n, \theta) &= \frac{1}{2} \Delta f(x_i, \theta) + \langle b(x_i, \theta), \nabla f(x_i, \theta) \rangle \\
&\quad + \frac{1}{2} \sum_{j=1}^{n} \frac{\partial^2}{\partial \theta_j^2} f(x_i, \theta) - \beta_t \sum_{j=1}^{n} l_{h_t}^{(j)}(x_0, x_j, \theta) \frac{\partial}{\partial \theta_j} f(x_i, \theta).
\end{aligned}$$

On ne se servira dans la suite que de ce résultat, et on appelera désormais algorithme de recuit partiel, un processus de diffusion $(X_t^{(0)}, X_t^{(1)}, \ldots, X_t^{(n)}, \Theta_t^{(1)}, \ldots, \Theta_t^{(n)})_{t \geqslant 0}$ sur $M^{n+1} \times N$, dont la famille des générateurs $(\mathbb{L}_t)_{t \geqslant 0}$ satisfait à cette propriété. On aurait pu ainsi considérer l'algorithme de recuit partiel décrit heuristiquement sur $M^{n+1} \times T^n$ par

$$
\begin{aligned}
\mathrm{d}X_t^{(0)} &= \gamma_t^{-1/2}\,\mathrm{d}B_t^{(0)} + \gamma_t^{-1} b(X_t^{(0)}, \Theta_t)\,\mathrm{d}t,\\
\mathrm{d}X_t^{(1)} &= \gamma_t^{-1/2}\,\mathrm{d}B_t^{(1)} + \gamma_t^{-1} b(X_t^{(1)}, \Theta_t + h_t e_1)]\,\mathrm{d}t,\\
\mathrm{d}X_t^{(2)} &= \gamma_t^{-1/2}\,\mathrm{d}B_t^{(2)} + \gamma_t^{-1} b(X_t^{(2)}, \Theta_t + h_t e_2)]\,\mathrm{d}t,\\
&\vdots\\
\mathrm{d}X_t^{(n)} &= \gamma_t^{-1/2}\,\mathrm{d}B_t^{(n)} + \gamma_t^{-1} b(X_t^{(n)}, \Theta_t + h_t e_n)]\,\mathrm{d}t,\\
\mathrm{d}\Theta_t^{(1)} &= \mathrm{d}W_t^{(1)} - \beta_t l_{h_t}^{(1)}(X_t^{(0)}, X_t^{(1)}, \Theta_t)\,\mathrm{d}t,\\
&\vdots\\
\mathrm{d}\Theta_t^{(n)} &= \mathrm{d}W_t^{(n)} - \beta_t l_{h_t}^{(n)}(X_t^{(0)}, X_t^{(n)}, \Theta_t)\,\mathrm{d}t
\end{aligned}
\tag{2}
$$

où $(B^{(0)}, B^{(1)}, \dots, B^{(n)}, W^{(1)}, \dots, W^{(n)}$ est un mouvement brownien sur $M^{n+1} \times T^n$ (indépendant de la v.a. initiale $(X_0^{(0)}, \dots, X_0^{(n)}, \Theta_0^{(1)}, \dots, \Theta_0^{(n)})$) c'est-à-dire un algorithme dont la restriction à $C^2(M^{n+1} \times N)$ du générateur $\tilde{\mathbb{L}}_t$ à l'instant $t \geqslant 0$ est donné par

$$
\tilde{\mathbb{L}}_t \cdot = \frac{1}{2}\sum_{i=0}^{n} [\Delta_i \cdot + 2\langle b(x_i, \theta), \nabla_i \cdot \rangle] + \frac{1}{2}\sum_{j=1}^{n}\left[\frac{\partial^2}{\partial\theta_j^2}\cdot - 2\beta_t l_{h_t}^{(j)}(x_0, x_j, \theta)\frac{\partial}{\partial\theta_j}\cdot\right]
$$

où les i en indices indiquent que l'on considère la structure riemannienne du $(i+1)^{\text{ième}}$ facteur de M^{n+1}.

Cependant, l'algorithme (1) est plus facile à mettre en oeuvre pratiquement, car il ne nécessite de simuler que r "browniens directeurs" indépendants sur $\mathbb{R}$ (au lieu de $(n+1)r$ pour (2)).

On suppose désormais qu'un algorithme de recuit partiel quelconque nous est donné. On se propose dans cet article de prouver le résultat énoncé ci-dessous, qui montre que l'on peut l'utiliser pour localiser les minima globaux de F, en observant où tend à se diriger Θ_t. Pour une justification heuristique de cette convergence (en loi), on renvoie à la fin de cette section.

Si on part d'une probabilité initiale m sur $M^{n+1} \times N$, on note m_t son image à l'instant $t \geqslant 0$, par l'algorithme considéré (c'est-à-dire la loi de $(X_t^{(0)}, X_t^{(1)}, \dots, X_t^{(n)}, \Theta_t^{(1)}, \dots, \Theta_t^{(n)})$), et n_t la projection de m_t sur N (*i.e.* la loi de $(\Theta_t^{(1)}, \dots, \Theta_t^{(n)})$).

Posons également $F_0 = \min_{\theta \in N} F(\theta)$.

Théorème 1

Il existe une constante $c \geqslant 0$, telle que si on prend pour t assez grand

$$\beta_t = k^{-1}\ln(t), \qquad \gamma_t = t^{-3}, \qquad h_t = t^{-1/2}$$

avec $k > c$, alors, pour toute constante $\delta > 0$,

$$\lim_{t \to +\infty} n_t(\{F - F_0 \geqslant \delta\}) = 0$$

On peut donner une description simple de la constante c qui apparaît dans le théorème: A $x, y \in N$, on associe $\mathscr{C}_{x,y}$ l'ensemble des applications continues $\phi:[0,1] \to N$ telles que $\phi(0) = x$ et $\phi(1) = y$, et on définit l'elévation (relativement à F) d'un élément $\phi \in \mathscr{C}_{x,y}$ par

$$e(\phi) = \sup_{0 \leqslant t \leqslant 1} F(\phi(t)) - F(x) - F(y) + \inf_{z \in N} F(z).$$

On peut alors prendre pour c la constante

$$c(F) = 2 \sup_{x, y \in N} \left(\inf_{\phi \in \mathscr{C}_{x,y}} e(\phi) \right).$$

Remarquons que l'on retrouve la même constante que si l'on avait considéré un algorithme de recuit simulé classique associé à F (d'ailleurs, dans le cas trivial où f ne dépend pas de x, le processus $(\Theta_t)_{t \geqslant 0}$ est un algorithme de recuit usuel). Ainsi la condition donnée ci-dessus pour β est en un certain sens optimale, car on sait que si on prend $\beta_t = k^{-1} \ln(t)$ avec $k < c(F)$ dans un algorithme de recuit classique, alors en général (notamment dans le cas où il y a un unique minimum global), il n'y a pas convergence vers les minima globaux (mais aussi vers certains minima locaux). Cependant, le fait de retrouver cette bonne constante semble lié à l'exposant 3 qui apparaît pour le comportement en temps grand de γ^{-1} (mais il est probable que ceci n'est dû qu'à la méthode utilisée), et comme on le verra, quitte à augmenter k, on peut diminuer cet exposant.

D'autre part, notons que si l'on ne sait pas en général calculer F, on peut du moins majorer facilement $c(F)$ par

$$c(F) \leqslant 2\left(\sup_{\theta \in N} F(\theta) - \inf_{\theta \in N} F(\theta) \right) \leqslant 2\left(\sup_{(x,\theta) \in M \times N} f(x,\theta) - \inf_{(x,\theta) \in M \times N} f(x,\theta) \right).$$

Dans (Miclo, 1994), où on ne considérait que le cas $n = 1$, on a étudié précisément le comportement en temps grand de la probabilité invariante instantanée μ_t associée à l'algorithme (2), en prouvant notamment que sous de bonnes conditions sur β, γ et h, sa projection sur N tendait à se concentrer au voisinage des minima globaux de F, puis en se servant des résultats obtenus, on a également montré que l'entropie de m_t par rapport à μ_t tendait vers 0 en temps grand (sous une hypothèse supplémentaire sur la température partielle β_t^{-1}), ce qui permettait de conclure.

Dans cet article, on ne va plus considérer l'entropie de m_t par rapport à μ_t, car on a du mal à bien estimer cette probabilité, d'autant plus qu'elle peut ne pas être régulière (comme par exemple dans le cas de l'algorithme (1)), mais on va s'intéresser à des "entropies partielles" de m_t par rapport à certaines mesures que l'on connaît bien et desquelles on espère que m_t se rapproche (partiellement) en temps grand.

Soit $m_t(\mathrm{d}x_0, \mathrm{d}x_1, \ldots, \mathrm{d}x_n \mid \theta)$ la projection sur M^{n+1} de la probabilité conditionnelle sous m_t sachant que la coordonnée sur N vaut θ. On définit $m_t^{(i)}(\mathrm{d}x_i \mid \theta)$, pour $0 \leqslant i \leqslant n$, comme étant la projection de cette probabilité sur le $(i+1)^{\text{ième}}$ facteur de M^{n+1} (que l'on convient désormais de noter $M^{(i)}$), c'est-à-dire $m_t^{(i)}(\mathrm{d}x_i \mid \theta) = \mathbb{E}_m[X_t^{(i)} \in \mathrm{d}x_i \mid \Theta_t = \theta]$.

De par la présence du terme γ_t qui tend vers 0, on espère qu'en temps grand cette probabilité conditionnelle $m_t(\mathrm{d}x_i\,|\,\theta)$ va se rapprocher de $\mu_{\theta+h_te_i}(\mathrm{d}x_i)$. On définit alors pour $t>0$ et $0\leqslant i\leqslant n$, les entropies partielles $I_t^{(i)}$ par

$$\begin{aligned} I_t^{(i)} &= \int_{M^{n+1}\times N} \ln\left(\frac{m_t^{(i)}(x_i\,|\,\theta)}{\mu_{\theta+h_te_i}(x_i)}\right) m_t(\mathrm{d}x_0,\mathrm{d}x_1,\ldots,\mathrm{d}x_n,\mathrm{d}\theta) \\ &= \int_{M^{(i)}\times N} \ln\left(\frac{m_t^{(i)}(x_i\,|\,\theta)}{\mu_{\theta+h_te_i}(x_i)}\right) m_t^{(i)}(\mathrm{d}x_i,\mathrm{d}\theta). \end{aligned}$$

Comme d'habitude, $m_t^{(i)}(x_i\,|\,\theta)$ (respectivement $\mu_{\theta+h_te_i}(x_i)$) représente la densité de $m_t^{(i)}(\mathrm{d}x_i\,|\,\theta)$ (respectivement $\mu_{\theta+h_te_i}(\mathrm{d}x_i)$) par rapport à la probabilité riemannienne $\lambda(\mathrm{d}x_i)$ sur $M^{(i)}$. On aura remarqué que les résultats de Taniguchi (1983) permettent de voir que pour tout $t>0$ et tout $0\leqslant i\leqslant n$, la probabilité $m_t^{(i)}(\mathrm{d}x_i,\mathrm{d}\theta)$ (qui est la projection de m_t sur $M^{(i)}\times N$) est absolument continue par rapport à la probabilité riemannienne $\lambda(\mathrm{d}x_i)(2\pi)^{-1}\,\mathrm{d}\theta_1\cdots(2\pi)^{-1}\,\mathrm{d}\theta_n$ sur $M^{(i)}\times T^n$, et que sa densité est strictement positive et de classe C^∞ sur $M^{(i)}\times N$. Taniguchi ne considère que le cas où les champs de vecteurs ne dépendent pas du temps, mais on peut se ramener à cette situation, comme me l'a fait remarquer Rémi Léandre, en considérant le temps comme une coordonnée supplémentaire (voir alors la remarque de Taniguchi (1983 p. 283) de son article), ou on peut reprendre directement les calculs, qui s'adaptent facilement (voir aussi l'article de Stroock (1981)). Ainsi $I_t^{(i)}$ est une quantité finie tout $t>0$. On conviendra dans la suite de noter $I_t=\sum_{0\leqslant i\leqslant n}I_t^{(i)}$.

D'autre part, sachant que $m_t^{(0)}(\mathrm{d}x_0\,|\,\theta)$ est proche de $\mu_\theta(\mathrm{d}x_0)$ et que pour $1\leqslant i\leqslant n$, $m_t^{(i)}(\mathrm{d}x_i\,|\,\theta)$ est proche de $\mu_{\theta+h_te_i}(\mathrm{d}x_i)$, l'évolution sur la $i^{\text{ième}}$ composante de N sera à peu près donnée par

$$\begin{aligned} \mathrm{d}\Theta_t^{(i)} &= \mathrm{d}W_t^{(i)} - \beta_t\left(\int m_t(\mathrm{d}x_0,\mathrm{d}x_1,\ldots,\mathrm{d}x_n,\,|\,\Theta_t)\,l_{h_t}^{(i)}(x_0,x_i,\Theta_t)\right)\mathrm{d}t, \\ &= \mathrm{d}W_t^{(i)} - \beta_t\left(\int \frac{\partial}{\partial\theta_i}f(x_0,\Theta_t) - \frac{1}{h_t}f(x_0,\Theta_t)m_t^{(0)}(\mathrm{d}x_0\,|\,\Theta_t)\right. \\ &\quad \left. + \frac{1}{h_t}\int f(x_i,\Theta_t)m_t^{(i)}(\mathrm{d}x_i\,|\,\Theta_t)\right)\mathrm{d}t, \\ &\simeq \mathrm{d}W_t^{(i)} - \beta_t\left(\int \frac{\partial}{\partial\theta_i}f(x_0,\Theta_t) - \frac{1}{h_t}f(x_0,\Theta_t)\mu_{\Theta_t}(\mathrm{d}x_0)\right. \\ &\quad \left. + \frac{1}{h_t}\int f(x_i,\Theta_t)\mu_{\Theta_t+h_te_i}(\mathrm{d}x_i)\right)\mathrm{d}t, \\ &= \mathrm{d}W_t^{(i)} - \beta_t\left(\int \frac{\partial}{\partial\theta_i}f(x_0,\Theta_t)\mu_{\Theta_t}(\mathrm{d}x_0) + \int f(x_0,\Theta_t)\frac{1}{h_t}\,[\mu_{\Theta_t+h_te_i}(x_0)\right. \\ &\quad \left. - \mu_{\Theta_t}(x_0)]\,\lambda(\mathrm{d}x_0)\right)\mathrm{d}t. \end{aligned}$$

Or le terme entre parenthèses du dernier membre de droite converge, quand h_t tend vers 0, vers $\partial F(\Theta_t)/\partial\theta_i$.

On est ainsi ramené à un algorithme de recuit simulé usuel associé à F sur N et on s'attend donc à ce que la loi de Θ_t soit proche de

$$\nu_{\beta_t}(\mathrm{d}\theta) = Z_{\beta_t}^{-1}\exp(-2\beta_t F(\theta))(2\pi)^{-1}\,\mathrm{d}\theta_1 \cdots (2\pi)^{-1}\,\mathrm{d}\theta_n$$

où Z_{β_t} est la constante de normalisation

$$Z_{\beta_t} = \int \exp(-2\beta_t F(\theta))(2\pi)^{-1}\mathrm{d}\theta_1 \cdots (2\pi)^{-1}\,\mathrm{d}\theta_n.$$

On est donc amené à considérer l'entropie partielle

$$J_t = \int \ln\left(\frac{n_t(\theta)}{\nu_{\beta_t}(\theta)}\right) n_t(\mathrm{d}\theta)$$

qui est aussi une quantité finie pour tout $t > 0$.

On va étudier les évolutions conjointes des entropies partielles précédentes dans le but de prouver que J_t tend vers 0, sous les hypothèses du théorème 1.

Ce résultat suffit en effet à montrer que pour tout $\delta > 0$,

$$\lim_{t \to +\infty} n_t(\{F - F_0 \geqslant \delta\}) = 0$$

comme on peut le voir en utilisant la relation entre variation totale et entropie

$$\| n_t - \nu_{\beta_t} \| \leqslant 4\sqrt{2J_t}$$

(où $\|\cdot\|$ représente la variation totale de mesures signées) et le fait que ν_{β_t} tend à se concentrer au voisinage des minima globaux de F.

2. Démonstration du théorème 1

Commençons par faire quelques rappels sur la probabilité invariante μ_θ: Pour tout $\theta \in N$ fixé, elle est absolument continue par rapport à la probabilité riemannienne λ et elle admet une densité strictement positive de classe C^∞. De plus, pour tout $x \in M$ fixé, $N \ni \theta \mapsto \mu_\theta(x)$ est au moins deux fois différentiable et $\mu_\theta(x)$ et ses deux premières dérivées partielles $\partial\mu_\theta(x)/\partial\theta_i$ et $\partial^2\mu_\theta(x)/\partial\theta_i\partial\theta_j$ sont continues sur $M \times N$ (voir par exemple les propositions 5 et 7 de (Miclo, 1994)). Ainsi, il existe une constante $A \geqslant 1$ telle que pour tout $1 \leqslant i, j \leqslant n$ et tout $(x, \theta) \in M \times N$,

$$\mu_\theta(x) \vee \mu_\theta^{-1}(x) \leqslant A, \quad \left|\frac{\partial}{\partial\theta_i}\ln(\mu_\theta(x))\right| \leqslant A, \quad \left|\frac{\partial^2}{\partial\theta_i\partial\theta_j}\,\mu_\theta(x)\right| \leqslant A,$$

Nous pouvons maintenant étudier l'évolution de J_t, pour $t > 0$.

Proposition 2. *Il existe deux constantes $c_1, c_2 > 0$ et un entier $p > 0$, tels que pour tout $t > 0$,*

$$\frac{\mathrm{d}J_t}{\mathrm{d}t} \leqslant c_1\left(\left|\frac{\mathrm{d}\beta_t}{\mathrm{d}t}\right| + (\beta_t h_t)^2 + \beta_t^2(h_t^{-2} \vee 1)I_t\right) - c_2(\beta_t \vee 1)^{-p}\exp(-c(F)\beta_t)J_t$$

où $c(F)$ est la constante définie dans la première section.

Démonstration. D'après les propriétés de régularité des n_t pour $t > 0$, il est clair que l'on peut dériver sous l'intégrale dans l'expression de J_t, et on obtient:

$$\frac{\mathrm{d}J_t}{\mathrm{d}t} = \int \frac{\partial}{\partial t} \ln(n_t(\theta))n_t(\mathrm{d}\theta) - \int \frac{\partial}{\partial t} \ln(\nu_{\beta_t}(\theta))n_t(\mathrm{d}\theta) + \int \ln\left(\frac{n_t(\theta)}{\nu_{\beta_t}(\theta)}\right)\frac{\partial}{\partial t} n_t(\theta)\Lambda(\mathrm{d}\theta)$$

(ici et dans toute la suite, $\Lambda(\mathrm{d}\theta)$ désignera la probabilité $(2\pi)^{-1}\mathrm{d}\theta_1 \cdots (2\pi)^{-1}\mathrm{d}\theta_n$ sur N)

$$= \int \frac{\partial}{\partial t} n_t(\theta)\Lambda(\mathrm{d}\theta) + 2\frac{\mathrm{d}\beta_t}{\mathrm{d}t}\int F(\theta)(n_t(\theta) - \nu_{\beta_t}(\theta))\Lambda(\mathrm{d}\theta)$$

$$+ \int \ln\left(\frac{n_t(\theta)}{\nu_{\beta_t}(\theta)}\right)\frac{\partial}{\partial t} n_t(\theta)\Lambda(\mathrm{d}\theta).$$

Le premier terme est nul, car il vaut $\mathrm{d}\int n_t(\mathrm{d}\theta)/\mathrm{d}t$, et le second est borné par $4\|F\|_\infty|\mathrm{d}\beta_t/\mathrm{d}t|$. Quant au dernier terme à estimer, il s'écrit aussi

$$\frac{\mathrm{d}}{\mathrm{d}s}\int \ln\left(\frac{n_t(\theta)}{\nu_{\beta_t}(\theta)}\right)m_s(\mathrm{d}x, \mathrm{d}\theta)\bigg|_{s=t}$$

où on convient désormais, pour simplifier l'écriture, de noter $x = (x_0, x_1, \dots, x_n)$ un élément générique de M^{n+1} et $\mathrm{d}x = (\mathrm{d}x_0, \mathrm{d}x_1, \dots, \mathrm{d}x_n)$.

Notons $\mathbb{L}_t = \gamma_t^{-1}L_{1,t} + L_{2,t}$ la restriction à $C^2(M^{n+1} \times N)$ du générateur à l'instant $t \geqslant 0$ de l'algorithme de recuit partiel considéré, où $L_{1,t}$ est un opérateur de diffusion (éventuellement dégénéré) sur M^{n+1} (paramétré par $\theta \in N$) et où $L_{2,t}$ est un opérateur de diffusion sur N (paramétré par $x \in M^{n+1}$). L'hypothèse faite sur la famille $(\mathbb{L}_t)_{t \geqslant 0}$ implique qu'une telle décomposition est possible et que

$$L_{2,t}\cdot = \sum_{1 \leqslant i \leqslant n} \frac{1}{2}\frac{\partial^2}{\partial \theta_i^2}\cdot - \beta_t I_{h_t}^{(i)}(x_0, x_i, \theta)\frac{\partial}{\partial \theta_i}\cdot$$

Par définition du générateur, on a

$$\frac{\mathrm{d}}{\mathrm{d}s}\int \ln\left(\frac{n_t(\theta)}{\nu_{\beta_t}(\theta)}\right)m_s(\mathrm{d}x, \mathrm{d}\theta)\bigg|_{s=t} = \int \mathbb{L}_t\left[\ln\left(\frac{n_t(\theta)}{\nu_{\beta_t}(\theta)}\right)\right]m_t(\mathrm{d}x, \mathrm{d}\theta)$$

$$= \int L_{2,t}\left[\ln\left(\frac{n_t(\theta)}{\nu_{\beta_t}(\theta)}\right)\right]m_t(\mathrm{d}x, \mathrm{d}\theta)$$

$$= \int \tilde{L}_{2,t}\left[\ln\left(\frac{n_t(\theta)}{\nu_{\beta_t}(\theta)}\right)\right]n_t(\theta)\Lambda(\mathrm{d}\theta)$$

où $\tilde{L}_{2,t}$ est l'opérateur sur $C^2(N)$ défini par

$$\tilde{L}_{2,t}\cdot = \sum_{1 \leqslant i \leqslant n} \frac{1}{2}\frac{\partial^2}{\partial \theta_i^2}\cdot - \beta_t \tilde{R}_t^i(\theta)\frac{\partial}{\partial \theta_i}\cdot$$

le champ de vecteurs $(\tilde{R}_t^i)_{1 \leqslant i \leqslant n}$ sur N étant donné par

$$\tilde{R}_t^i(\theta) = \int l_{h_t}^{(i)}(x_0, x_i, \theta) m_t(\mathrm{d}x \mid \theta)$$

$$= \int \left[\frac{\partial}{\partial \theta_i} f(x_0, \theta) - \frac{1}{h_t} f(x_0, \theta) \right] m_t^{(0)}(\mathrm{d}x_0 \mid \theta) + \frac{1}{h_t} \int f(x_i, \theta) m_t^{(i)}(\mathrm{d}x_i \mid \theta).$$

Posons, pour $1 \leqslant i \leqslant n$, $\hat{R}_t^i(\theta) = \partial F(\theta)/\partial\theta_i$ et $\hat{L}_{2,t}$ l'opérateur associé au champ de vecteurs $-\beta_t(\hat{R}_t^i)_{1 \leqslant i \leqslant n}$:

$$\hat{L}_{2,t} \cdot = \sum_{1 \leqslant i \leqslant n} \frac{1}{2} \frac{\partial^2}{\partial \theta_i^2} \cdot - \beta_t \hat{R}_t^i(\theta) \frac{\partial}{\partial \theta_i} \cdot .$$

On a

$$\int \tilde{L}_{2,t}\left[\ln\left(\frac{n_t(\theta)}{\nu_{\beta_t}(\theta)} \right) \right] n_t(\theta) \Lambda(\mathrm{d}\theta)$$

$$= \int \hat{L}_{2,t}\left[\ln\left(\frac{n_t(\theta)}{\nu_{\beta_t}(\theta)} \right) \right] n_t(\theta) \Lambda(\mathrm{d}\theta) + \int (\tilde{L}_{2,t} - \hat{L}_{2,t}) \left[\ln\left(\frac{n_t(\theta)}{\nu_{\beta_t}(\theta)} \right) \right] n_t(\theta) \Lambda(\mathrm{d}\theta). \quad (3)$$

Remarquons que $\hat{L}_{2,t}$ est un opérateur de diffusion, ainsi

$$\hat{L}_{2,t}\left[\ln\left(\frac{n_t(\theta)}{\nu_{\beta_t}(\theta)} \right) \right] = \frac{\nu_{\beta_t}(\theta)}{n_t(\theta)} \hat{L}_{2,t}\left[\frac{n_t(\theta)}{\nu_{\beta_t}(\theta)} \right] - \frac{1}{2} \left(\frac{\nu_{\beta_t}(\theta)}{n_t(\theta)} \right)^2 \sum_{1 \leqslant i \leqslant n} \left(\frac{\partial}{\partial \theta_i} \frac{n_t(\theta)}{\nu_{\beta_t}(\theta)} \right)^2 .$$

Mais ν_{β_t} est la probabilité invariante associée à $\hat{L}_{2,t}$, d'où

$$\int \hat{L}_{2,t}\left[\ln\left(\frac{n_t(\theta)}{\nu_{\beta_t}(\theta)} \right) \right] n_t(\theta) \Lambda(\mathrm{d}\theta)$$

$$= \int \hat{L}_{2,t}\left[\frac{n_t(\theta)}{\nu_{\beta_t}(\theta)} \right] \nu_{\beta_t}(\theta) \Lambda(\mathrm{d}\theta) - \frac{1}{2} \int \left(\frac{\nu_{\beta_t}(\theta)}{n_t(\theta)} \right)^2 \sum_{1 \leqslant i \leqslant n} \left(\frac{\partial}{\partial \theta_i} \frac{n_t(\theta)}{\nu_{\beta_t}(\theta)} \right)^2 n_t(\theta) \Lambda(\mathrm{d}\theta)$$

$$= -2 \int \sum_{1 \leqslant i \leqslant n} \left(\frac{\partial}{\partial \theta_i} \sqrt{\frac{n_t(\theta)}{\nu_{\beta_t}(\theta)}} \right)^2 \nu_{\beta_t}(\mathrm{d}\theta)$$

(notons que l'on ne s'est pas servi ici la réversibilité de ν_{β_t} par rapport à $\hat{L}_{2,t}$, on réutilisera cet argument par la suite).

D'autre part, on peut borner le second terme du membre de droite de (3) par

$$\left| \int (\tilde{L}_{2,t} - \hat{L}_{2,t}) \left[\ln\left(\frac{n_t(\theta)}{\nu_{\beta_t}(\theta)} \right) \right] n_t(\theta) \Lambda(\mathrm{d}\theta) \right|$$

$$= \beta_t \left| \sum_{1 \leqslant i \leqslant n} \int (\tilde{R}_t^i - \hat{R}_t^i)(\theta) \left[\frac{\partial}{\partial \theta_i} \ln\left(\frac{n_t(\theta)}{\nu_{\beta_t}(\theta)} \right) \right] n_t(\theta) \Lambda(\mathrm{d}\theta) \right|$$

$$= 2\beta_t \left| \sum_{1 \leqslant i \leqslant n} \int (\tilde{R}_t^i - \hat{R}_t^i)(\theta) \left[\frac{\partial}{\partial \theta_i} \sqrt{\left(\frac{n_t(\theta)}{\nu_{\beta_t}(\theta)} \right)} \right] \sqrt{\frac{\nu_{\beta_t}(\theta)}{n_t(\theta)}} n_t(\theta) \Lambda(\mathrm{d}\theta) \right|$$

$$\leqslant 2\beta_t \sum_{1 \leqslant i \leqslant n} \sqrt{\int (\tilde{R}_t^i - \hat{R}_t^i)^2(\theta) n_t(\theta) \Lambda(\mathrm{d}\theta)} \sqrt{\int \left(\frac{\partial}{\partial \theta_i} \sqrt{\frac{n_t(\theta)}{\nu_{\beta_t}(\theta)}} \right)^2 \nu_{\beta_t}(\theta) \Lambda(\mathrm{d}\theta)}$$

$$\leqslant \sum_{1 \leqslant i \leqslant n} \left(\beta_t^2 \int (\tilde{R}_t^i - \hat{R}_t^i)^2(\theta) n_t(\mathrm{d}\theta) + \int \left(\frac{\partial}{\partial \theta_i} \sqrt{\frac{n_t(\theta)}{\nu_{\beta_t}(\theta)}} \right)^2 \nu_{\beta_t}(\mathrm{d}\theta) \right).$$

On obtient donc

$$\int \tilde{L}_{2,t}\left[\ln\left(\frac{n_t(\theta)}{v_{\beta_t}(\theta)}\right)\right] n_t(\theta)\Lambda(\mathrm{d}\theta)$$

$$\leqslant \beta_t^2 \sum_{1\leqslant i\leqslant n} \int (\tilde{R}_t^i - \hat{R}_t^i)^2(\theta) n_t(\mathrm{d}\theta) - \int \sum_{1\leqslant i\leqslant n} \left(\frac{\partial}{\partial\theta_i}\sqrt{\frac{n_t(\theta)}{v_{\beta_t}(\theta)}}\right)^2 v_{\beta_t}(\mathrm{d}\theta). \qquad (4)$$

Evaluons le premier terme. Pour un $1 \leqslant i \leqslant n$ et un $\theta \in N$ fixés, on a

$$|\tilde{R}_t^i - \hat{R}_t^i|(\theta) \leqslant \left|\int \left[\frac{\partial}{\partial\theta_i} f(x_0,\theta) - \frac{1}{h_t} f(x_0,\theta)\right][m_t^{(0)}(\mathrm{d}x_0|\theta) - \mu_\theta(\mathrm{d}x_0)]\right|$$

$$+ \left|\frac{1}{h_t}\int f(x_i,\theta)[m_t^{(i)}(\mathrm{d}x_i|\theta) - \mu_{\theta+h_t e_i}(\mathrm{d}x_i)]\right|$$

$$+ \left|\int l_{h_t}^{(i)}(x_0,x_i,\theta)\mu_\theta(\mathrm{d}x_0)\mu_{\theta+h_t e_i}(\mathrm{d}x_i) - \frac{\partial}{\partial\theta_i} F(\theta)\right|.$$

Or, en utilisant l'inégalité entre variation totale et entropie, on a

$$\left|\int \left[\frac{\partial}{\partial\theta_i} f(x_0,\theta) - \frac{1}{h_t} f(x_0,\theta)\right][m_t^{(0)}(\mathrm{d}x_0|\theta) - \mu_\theta(\mathrm{d}x_0)]\right|$$

$$\leqslant K(h_t^{-1} \vee 1)\sqrt{\int \ln\left(\frac{m_t^{(0)}(x_0|\theta)}{\mu_\theta(x_0)}\right) m_t^{(0)}(\mathrm{d}x_0|\theta)},$$

$$\left|\frac{1}{h_t}\int f(x_i,\theta)[m_t^{(i)}(\mathrm{d}x_i|\theta) - \mu_{\theta+h_t e_i}(\mathrm{d}x_i)]\right|$$

$$\leqslant K(h_t^{-1} \vee 1)\sqrt{\int \ln\left(\frac{m_t^{(i)}(x_i|\theta)}{\mu_{\theta+h_t e_i}(x_i)}\right) m_t^{(i)}(\mathrm{d}x_i|\theta)}$$

avec

$$K = 4\sqrt{2}\left(\max_{1\leqslant i\leqslant n}\left\|\frac{\partial}{\partial\theta_i} f\right\|_\infty + \|f\|_\infty\right).$$

D'autre part, il est clair que

$$\left|\int l_{h_t}^{(i)}(x_0,x_i,\theta)\mu_\theta(\mathrm{d}x_0)\mu_{\theta+h_t e_i}(\mathrm{d}x_i) - \frac{\partial}{\partial\theta_i} F(\theta)\right|$$

$$= \left|\int f(x_0,\theta)\left(\frac{1}{h_t}[\mu_{\theta+h_t e_i}(x_0) - \mu_\theta(x_0)] - \frac{\partial}{\partial\theta_i}\mu_\theta(x_0)\right)\lambda(\mathrm{d}x_0)\right|$$

$$\leqslant A h_t \|f\|_\infty$$

où A est constante définie avant la proposition 2.

Ainsi, en posant $K' = 3\max[K^2(n+1); (A\|f\|_\infty)^2 n]$, on voit que pour tout $t > 0$,

$$\sum_{1\leqslant i\leqslant n}\int (\tilde{R}_t^i - \hat{R}_t^i)^2(\theta) n_t(\mathrm{d}\theta) \leqslant 3\sum_{1\leqslant i\leqslant n}[A^2\|f\|_\infty^2 h_t^2 + K^2(h_t^{-2} \vee 1)(I_t^{(0)} + I_t^{(i)})]$$

$$\leqslant K'(h_t^2 + (h_t^{-2} \vee 1) I_t).$$

Reste à nous intéresser au second terme du membre de droite de (4); En utilisant les inégalités de Sobolev logarithmiques satisfaites par ν_{β_t} (voir le théorème (1.14) de Holley et Stroock (1988) et le théorème (3.23) de Holley et al. (1989)), on voit qu'il existe une constante $c_2 > 0$ et un entier $p > 0$ (qui ne dépendent que de la norme uniforme de F et de ses premières dérivées, ainsi que de la dimension n) tels que pour tout $t > 0$,

$$\int \sum_{1 \leqslant i \leqslant n} \left(\frac{\partial}{\partial \theta_i} \sqrt{\frac{n_t(\theta)}{\nu_{\beta_t}(\theta)}} \right)^2 \nu_{\beta_t}(\mathrm{d}\theta) \geqslant c_2(\beta_t \vee 1)^{-p} \exp(-c(F)\beta_t) J_t$$

(pour plus de détails, voir par exemple (Miclo, 1992)).

Ceci termine la démonstration de la Proposition 2, si on y prend $c_1 = \max(K'; 4 \|f\|_\infty)$. □

La proposition précédente montre, comme l'on s'y attendait, que l'évolution à l'instant $t > 0$ de J_t est liée à I_t. De même l'évolution de I_t est liée à celle de J_t, mais d'une autre manière:

Proposition 3. *Il existe deux constantes $k_1, k_2 > 0$, telles que pour tout $t > 0$,*

$$\frac{\mathrm{d}I_t}{\mathrm{d}t} \leqslant k_1 \left(\left| \frac{\mathrm{d}h_t}{\mathrm{d}t} \right| + \left| \frac{\mathrm{d}\beta_t}{\mathrm{d}t} \right| + (\beta_t^2 \vee 1)(h_t^{-2} \vee 1) \right) - (n+1) \frac{\mathrm{d}J_t}{\mathrm{d}t} - k_2 \gamma_t^{-1} I_t .$$

Démonstration. Soit $0 \leqslant i \leqslant n$ fixé. Notons que pour tout $(x_i, \theta) \in M \times N$, $m_t^{(i)}(x_i | \theta) = m_t^{(i)}(x_i, \theta) n_t^{-1}(\theta)$, ainsi, comme dans la proposition précédente, il n'y a aucun problème pour dériver sous l'intégrale dans l'expression de $I_t^{(i)}$, et on trouve

$$\frac{\mathrm{d}I_t^{(i)}}{\mathrm{d}t} = \int \frac{\partial}{\partial t} \ln(m_t^{(i)}(x_i | \theta)) m_t^{(i)}(\mathrm{d}x_i, \mathrm{d}\theta) - \int \frac{\partial}{\partial t} \ln(\mu_{\theta + h_t e_i}(x_i)) m_t^{(i)}(\mathrm{d}x_i, \mathrm{d}\theta)$$
$$+ \int \ln \left(\frac{m_t^{(i)}(x_i | \theta)}{\mu_{\theta + h_t e_i}(x_i)} \right) \frac{\partial}{\partial t} m_t^{(i)}(x_i, \theta) \lambda(\mathrm{d}x_i) \Lambda(\mathrm{d}\theta).$$

Le premier terme est nul car il vaut

$$\int \frac{\partial}{\partial t} \ln(m_t^{(i)}(x_i, \theta)) m_t^{(i)}(\mathrm{d}x_i, \mathrm{d}\theta) - \int \frac{\partial}{\partial t} \ln(n_t(\theta)) m_t^{(i)}(\mathrm{d}x_i, \mathrm{d}\theta)$$
$$= \int \frac{\partial}{\partial t} m_t^{(i)}(x_i, \theta) \lambda(\mathrm{d}x_i) \Lambda(\mathrm{d}\theta) - \int \frac{\partial}{\partial t} (n_t(\theta) \Lambda(\mathrm{d}\theta) = 0 - 0.$$

Le second se majore facilement à l'aide de la constante A définie avant la proposition 2:

$$\left| \int \frac{\partial}{\partial t} \ln(\mu_{\theta + h_t e_i}(x_i)) m_t^{(i)}(\mathrm{d}x_i, \mathrm{d}\theta) \right| \leqslant A \left| \frac{\mathrm{d}h_t}{\mathrm{d}t} \right| .$$

Pour estimer le dernier terme, on considère l'opérateur de diffusion $L_{1,t}^{(i)}$ (paramétré par $\theta \in N$) défini sur $C^2(M^{(i)})$ par

$$L_{1,t}^{(i)} \cdot = \tfrac{1}{2} \Delta_i \cdot + \langle b(x_i, \theta + h_t e_i), \nabla_i \cdot \rangle_i$$

(les indices $0 \leqslant i \leqslant n$ indiquent que l'on considère la structure riemannienne de $M^{(i)}$).

Par hypothèse, on a que pour tout $f \in C^2(M^{(i)})$ et tout $x_i \in M^{(i)}$, $L_{1,t}^{(i)} f(x_i) = L_{1,t} \hat{f}_i(x)$, où $L_{1,t}$ est l'opérateur introduit dans la démonstration de la proposition 2, où $\hat{f}_i$ est défini sur M^{n+1} par $\hat{f}_i(x) = f(x_i)$ et où $x \in M^{n+1}$ est tel que sa $(i+1)^{\text{ième}}$ coordonnée soit x_i. Ainsi,

$$\begin{aligned}
&\int \ln\left(\frac{m_t^{(i)}(x_i\,|\,\theta)}{\mu_{\theta+h_te_i}(x_i)}\right) \frac{\partial}{\partial t}\, m_t^{(i)}(x_i,\theta)\lambda(\mathrm{d}x_i)\Lambda(\mathrm{d}\theta) \\
&= \int \mathbb{L}_t\left[\ln\left(\frac{m_t^{(i)}(x_i\,|\,\theta)}{\mu_{\theta+h_te_i}(x_i)}\right)\right] m_t(\mathrm{d}x,\mathrm{d}\theta) \\
&= \gamma_t^{-1} \int L_{1,t}^{(i)}\left[\ln\left(\frac{m_t^{(i)}(x_i\,|\,\theta)}{\mu_{\theta+h_te_i}(x_i)}\right)\right] m_t^{(i)}(x_i,\theta)\lambda(\mathrm{d}x_i)\Lambda(\mathrm{d}\theta) \\
&\quad + \int L_{2,t}\left[\ln\left(\frac{m_t^{(i)}(x_i\,|\,\theta)}{\mu_{\theta+h_te_i}(x_i)}\right)\right] m_t(\mathrm{d}x,\mathrm{d}\theta).
\end{aligned} \tag{5}$$

Cependant, du fait que pour tout $\theta \in N$ fixé, $L_{1,t}^{(i)}$ est un opérateur de diffusion sur $M^{(i)}$ dont $\mu_{\theta+h_te_i}$ est la mesure invariante, on a

$$\int L_{1,t}^{(i)}\left[\ln\left(\frac{m_t^{(i)}(x_i\,|\,\theta)}{\mu_{\theta+h_te_i}(x_i)}\right)\right] m_t^{(i)}(x_i\,|\,\theta)\lambda(\mathrm{d}x_i) = -2\int \left| \nabla_i \sqrt{\frac{m_t^{(i)}(x_i\,|\,\theta)}{\mu_{\theta+h_te_i}(x_i)}}\right|^2 \mu_{\theta+h_te_i}(\mathrm{d}x_i).$$

On utilise alors les inégalités de Sobolev logarithmiques satisfaites par les $\mu_{\theta+h_te_i}$ pour voir qu'il existe une constante $k_2 > 0$ indépendante de $t > 0$, de $\theta \in N$ et de $0 \leqslant i \leqslant n$, telle que

$$\int \left| \nabla_i \sqrt{\frac{m_t^{(i)}(x_i\,|\,\theta)}{\mu_{\theta+h_te_i}(x_i)}}\right|^2 \mu_{\theta+h_te_i}(\mathrm{d}x_i) \geqslant \frac{k_2}{2}\int \ln\left(\frac{m_t^{(i)}(x_i\,|\,\theta)}{\mu_{\theta+h_te_i}(x_i)}\right) m_t^{(i)}(\mathrm{d}x_i\,|\,\theta)$$

(En effet, remarquons que $\ln(\mu_{\theta+h_te_i}(x_i))$ est borné uniformément en $t \geqslant 0$, $\theta \in N$, $0 \leqslant i \leqslant n$ et $x_i \in M^{(i)}$, ce qui suffit pour prouver l'existence d'une constante k_2 satisfaisant l'inégalité ci-dessus, cf. le lemma 3.13 de Holley et Stroock (1988).)

Ainsi,

$$\begin{aligned}
&\int L_{1,t}^{(i)}\left[\ln\left(\frac{m_t^{(i)}(x_i\,|\,\theta)}{\mu_{\theta+h_te_i}(x_i)}\right)\right] m_t^{(i)}(x_i,\theta)\lambda(\mathrm{d}x_i)\Lambda(\mathrm{d}\theta) \\
&= \int\left(\int L_{1,t}^{(i)}\left[\ln\left(\frac{m_t^{(i)}(x_i\,|\,\theta)}{\mu_{\theta+h_te_i}(x_i)}\right)\right] m_t^{(i)}(x_i\,|\,\theta)\lambda(\mathrm{d}x_i)\right) n_t(\theta)\Lambda(\mathrm{d}\theta) \\
&\leqslant \int\left(k_2\int \ln\left(\frac{m_t^{(i)}(x_i\,|\,\theta)}{\mu_{\theta+h_te_i}(x_i)}\right) m_t^{(i)}(\mathrm{d}x_i\,|\,\theta)\right) n_t(\theta)\Lambda(\mathrm{d}\theta) \\
&= -k_2 I_t^{(i)}.
\end{aligned}$$

Intéressons-nous au second terme du membre de droite de (5). Notons $m_t^{(i)}(\mathrm{d}\theta\,|\,x_i)$ la projection sur N de la probabilité conditionnelle sous m_t, sachant que la coordonnée sur $M^{(i)}$ vaut x_i. Alors $m_t^{(i)}(\theta\,|\,x_i) = m_t^{(i)}(x_i,\theta)(p_t^{(i)}(x_i))^{-1}$ où $p_t^{(i)}$ est la projection de

m_t sur $M^{(i)}$. On a

$$L_{2,t}\left[\ln\left(\frac{m_t^{(i)}(x_i|\theta)}{\mu_{\theta+h_te_i}(x_i)}\right)\right] = L_{2,t}[\ln(m_t^{(i)}(\theta|x_i))] + L_{2,t}\left[\ln\left(\frac{1}{\mu_{\theta+h_te_i}(x_i)n_t(\theta)}\right)\right.$$
$$\left. + L_{2,t}[\ln(p_t^{(i)}(x_i))\right]$$
$$= L_{2,t}[\ln(m_t^{(i)}(\theta|x_i))] - L_{2,t}[\ln(\mu_{\theta+h_te_i}(x_i)n_t(\theta))] .$$

Cependant, par définition de l'opérateur $L_{2,t}$,

$$\int L_{2,t}[\ln(m_t^{(i)}(\theta|x_i))]\, m_t(\mathrm{d}x, \mathrm{d}\theta)$$
$$= \frac{1}{2}\int \sum_{1\leqslant j\leqslant n} \frac{\partial^2}{\partial\theta_j^2} \ln(m_t^{(i)}(\theta|x_i))m_t^{(i)}(x_i,\theta)\,\lambda(dx_i)\Lambda(\mathrm{d}\theta)$$
$$- \beta_t\int \sum_{1\leqslant j\leqslant n} l_{h_t}^{(j)}(x_0,x_j,\theta)\frac{\partial}{\partial\theta_j}\ln(m_t^{(i)}(\theta|x_i))m_t(\mathrm{d}x,\mathrm{d}\theta)$$
$$= -2\sum_{1\leqslant j\leqslant n}\int\left[\frac{\partial}{\partial\theta_j}\sqrt{m_t^{(i)}(\theta|x_i)}\right]^2\Lambda(\mathrm{d}\theta)p_t^{(i)}(\mathrm{d}x_i)$$
$$- \beta_t\sum_{1\leqslant j\leqslant n}\int l_{h_t}^{(j)}(x_0,x_j,\theta)\frac{\partial}{\partial\theta_j}\ln(m_t^{(i)}(\theta|x_i)]m_t(\mathrm{d}x,\mathrm{d}\theta) .$$

L'inégalité de Cauchy–Schwarz permet de borner le dernier terme du membre de droite:

$$\left|\beta_t\sum_{1\leqslant j\leqslant n}\int l_{h_t}^{(j)}(x_0,x_j,\theta)\frac{\partial}{\partial\theta_j}\ln[m_t^{(i)}(\theta|x_i)]m_t(\mathrm{d}x,\mathrm{d}\theta)\right|$$
$$\leqslant \beta_t\sum_{1\leqslant j\leqslant n}\sqrt{\int[l_{h_t}^{(j)}(x_0,x_j,\theta)]^2 m_t(\mathrm{d}x,\mathrm{d}\theta)}$$
$$\times\sqrt{\int\left[\frac{\partial}{\partial\theta_j}\ln[m_t^{(i)}(\theta|x_i)]\right]^2 m_t(\mathrm{d}x,\mathrm{d}\theta)}$$
$$\leqslant \beta_t\sum_{1\leqslant j\leqslant n}K(h_t^{-1}\vee 1)\sqrt{\int\left[\frac{\partial}{\partial\theta_j}\ln[m_t^{(i)}(\theta|x_i)]\right]^2 m_t^{(i)}(\theta|x_i)p_t^{(i)}(x_i)\lambda(\mathrm{d}x_i)\Lambda(\mathrm{d}\theta)}$$
$$= 2\beta_tK(h_t^{-1}\vee 1)\sum_{1\leqslant j\leqslant n}\sqrt{\int\left[\frac{\partial}{\partial\theta_j}\sqrt{(m_t^{(i)}(\theta|x_i)}\right]^2 p_t^{(i)}(x_i)\lambda(\mathrm{d}x_i)\Lambda(\mathrm{d}\theta)}$$
$$\leqslant \frac{1}{2}\beta_t^2 n[K(h_t^{-1}\vee 1)]^2 + 2\sum_{1\leqslant j\leqslant n}\int\left[\frac{\partial}{\partial\theta_j}\sqrt{(m_t^{(i)}(\theta|x_i)]}\right]^2 p_t^{(i)}(\mathrm{d}x_i)\Lambda(\mathrm{d}\theta)$$

où on a posé

$$K = \max_{1\leqslant j\leqslant n}\left\|\frac{\partial}{\partial\theta_j}f\right\|_\infty + 2\|f\|_\infty .$$

On a donc finalement

$$\int L_{2,t}[\ln(m_t^{(i)}(\theta\,|\,x_i))]\,m_t(\mathrm{d}x,\mathrm{d}\theta) \leqslant \frac{1}{2}\,K^2 n\beta_t^2(h_t^{-2} \vee 1).$$

D'autre part, décomposons le terme restant à évaluer en

$$L_{2,t}[\ln(\mu_{\theta+h_t e_i}(x_i)n_t(\theta))] = L_{2,t}[\ln(\mu_{\theta+h_t e_i}(x_i)\nu_{\beta_t}(\theta))] + L_{2,t}\left[\ln\left(\frac{n_t(\theta)}{\nu_{\beta_t}(\theta)}\right)\right].$$

On a vu dans la démonstration de la proposition 2 que

$$-\int L_{2,t}\left[\ln\left(\frac{n_t(\theta)}{\nu_{\beta_t}(\theta)}\right)\right]m_t(\mathrm{d}x,\mathrm{d}\theta) = -\frac{\mathrm{d}J_t}{\mathrm{d}t} + 2\,\frac{\mathrm{d}\beta_t}{\mathrm{d}t}\int F(\theta)(n_t(\theta) - \nu_{\beta_t}(\theta))\Lambda(\mathrm{d}\theta)$$
$$\leqslant -\frac{\mathrm{d}J_t}{\mathrm{d}t} + 4\,\|F\|_\infty\left|\frac{\mathrm{d}\beta_t}{\mathrm{d}t}\right|$$

et il est clair, en utilisant les estimées qui précédent la proposition 2, qu'il existe une constante $K' \geqslant 0$, indépendante de $(x_i,\theta) \in M^{(i)} \times N$ et de $t \geqslant 0$, telle que l'on ait

$$|L_{2,t}[\ln(\mu_{\theta+h_t e_i}(x_i))]| \leqslant K'(\beta_t \vee 1)(h_t^{-1} \vee 1),$$
$$|L_{2,t}[\ln\nu_{\beta_t}(\theta))]| \leqslant K'\beta_t(\beta_t \vee 1)(h_t^{-1} \vee 1).$$

On en déduit que pour tout $t > 0$,

$$\frac{\mathrm{d}I_t^{(i)}}{\mathrm{d}t} \leqslant \frac{k_1}{n+1}\left(\left|\frac{\mathrm{d}h_t}{\mathrm{d}t}\right| + \left|\frac{\mathrm{d}\beta_t}{\mathrm{d}t}\right| + (\beta_t^2 \vee 1)(h_t^{-2} \vee 1)\right) - \frac{\mathrm{d}J_t}{\mathrm{d}t} - k_2\gamma_t^{-1}I_t^{(i)}$$

avec $k_1 = (n+1)\max(A;\,4\|f\|_\infty;\,\frac{1}{2}K^2 n + 2K')$, d'où le résultat annoncé, en sommant ces inégalités pour $0 \leqslant i \leqslant n$. □

Nous sommes maintenant en mesure de montrer que sous les hypothèses du théoréme 1,

$$\lim_{t\to+\infty} J_t = 0. \tag{6}$$

Tout d'abord, puisque l'on suppose a priori que $\lim_{t\to+\infty}\max(\beta_t^{-1};\gamma_t;h_t) = 0$, on peut se ramener au cas où pour tout $t \geqslant 0$ on a $\beta_t \geqslant 1$, $\gamma_t \leqslant 1$ et $h_t \leqslant 1$, quitte à n'étudier l'algorithme qu'à partir de l'instant t_0 où la condition précédente est satisfaite pour tout $t \geqslant t_0$. Ainsi les propositions 2 et 3 montrent que pour tout $t > 0$,

$$\frac{\mathrm{d}J_t}{\mathrm{d}t} \leqslant c_1\left(\left|\frac{\mathrm{d}\beta_t}{\mathrm{d}t}\right| + (\beta_t h_t)^2 + \beta_t^2 h_t^{-2} I_t\right) - c_2\beta_t^{-p}\exp(-c(F)\beta_t)J_t,$$
$$\frac{\mathrm{d}I_t}{\mathrm{d}t} \leqslant k_1\left(\left|\frac{\mathrm{d}h_t}{\mathrm{d}t}\right| + \left|\frac{\mathrm{d}\beta_t}{\mathrm{d}t}\right| + \beta_t^2 h_t^{-2}\right) - (n+1)\,\frac{\mathrm{d}J_t}{\mathrm{d}t} - k_2\gamma_t^{-1}I_t. \tag{7}$$

Puis, toujours pour la même raison, on peut supposer que pour tout $t \geqslant 0$, l'expression

$$k_2 \gamma_t^{-1} + c_1(n+1)\beta_t^2 h_t^{-2} - c_2 \beta_t^{-p} \exp(-c(F)\beta_t)$$

est strictement positive. On peut alors poser, pour tout $t \geqslant 0$,

$$\kappa_t = \frac{c_1 \beta_t^2 h_t^{-2}}{k_2 \gamma_t^{-1} + c_1(n+1)\beta_t^2 h_t^{-2} - c_2 \beta_t^{-p} \exp(-c(F)\beta_t)}.$$

Faisons l'hypothèse supplémentaire (H) que κ_t converge vers 0 quand t tend vers l'infini et qu'il existe un $t_1 > 0$ tel que l'application $\kappa.$ soit décroissante sur $[t_1, +\infty[$. On peut alors aussi supposer que t_1 est tel que pour tout $t \geqslant t_1$, $\kappa_1 < (n+1)^{-1}$.

Définissons pour $t \geqslant 0$,

$$K_t = J_t + \kappa_t I_t.$$

Il suffit de trouver des conditions sur les évolutions $\beta.$, γ et h qui assurent que $\lim_{t \to +\infty} K_t = 0$, car ceci entraîne (6) (en effet, d'après l'inégalité de Jensen, les $I_t^{(i)}$ (et donc I_t) et J_t sont des quantités positives).

L'application $K.$ est dérivable et pour tout $t \geqslant t_1$,

$$\begin{aligned}
\frac{dK_t}{dt} &= \frac{dJ_t}{dt} + \frac{d\kappa_t}{dt} I_t + \kappa_t \frac{dI_t}{dt} \\
&\leqslant \frac{dJ_t}{dt} + \kappa_t \frac{dI_t}{dt} \\
&\leqslant (1-(n+1)\kappa_t)\frac{dJ_t}{dt} + k_1 \kappa_t \left(\left|\frac{dh_t}{dt}\right| + \left|\frac{d\beta_t}{dt}\right| + \beta_t^2 h_t^{-2} \right) - k_2 \kappa_t \gamma_t^{-1} I_t \\
&\leqslant [c_1(1-(n+1)\kappa_t) + k_1 \kappa_t] \left|\frac{d\beta_t}{dt}\right| + k_1 \kappa_t \left|\frac{dh_t}{dt}\right| \\
&\quad + c_1(1-(n+1)\kappa_t)\beta_t^2 h_t^2 + k_1 \kappa_t \beta_t^2 h_t^{-2} \\
&\quad + (c_1(1-(n+1)\kappa_t)\beta_t^2 h_t^{-2} - k_2 \kappa_t \gamma_t^{-1}) I_t \\
&\quad - c_2(1-(n+1)\kappa_t)\beta_t^{-p} \exp(-c(F)\beta_t) J_t.
\end{aligned}$$

Cependant, κ_t a été défini de manière que

$$\begin{aligned}
c_1(1-(n+1)\kappa_t)\beta_t^2 h_t^{-2} - k_2 \kappa_t \gamma_t^{-1} &= -c_2 \kappa_t \beta_t^{-p} \exp(-c(F)\beta_t) \\
&\leqslant -c_2 \kappa_t (1-(n+1)\kappa_t) \beta_t^{-p} \exp(-c(F)\beta_t)
\end{aligned}$$

et on obtient donc que l'application $t \mapsto K_t$ satisfait pour $t \geqslant t_1$, l'inégalité différentielle

$$\frac{dK_t}{dt} \leqslant A_t - B_t K_t$$

où

$$A_t = [c_1(1-(n+1)\kappa_t) + k_1\kappa_t]\left|\frac{\mathrm{d}\beta_t}{\mathrm{d}t}\right| + k_1\kappa_t\left|\frac{\mathrm{d}h_t}{\mathrm{d}t}\right|$$

$$+ c_1(1-(n+1)\kappa_t)\beta_t^2 h_t^2 + k_1\kappa_t\beta_t^2 h_t^{-2},$$

$$B_t = c_2(1-(n+1)\kappa_t)\beta_t^{-p}\exp(-c(F)\beta_t).$$

Une condition suffisante pour obtenir $\lim_{t\to+\infty} K_t = 0$, sachant que $K_.$, $A_.$ et $B_.$ sont des applications positives, est:

$$\int^{+\infty} B_t\,\mathrm{d}t = +\infty, \qquad \lim_{t\to+\infty}\frac{A_t}{B_t} = 0.$$

Prenons des évolutions particulières du type

$$\beta_t = k^{-1}\ln(t), \quad \gamma_t = t^{-a}, \quad h_t = t^{-b}$$

(pour t suffisamment grand) où k, a et b sont des réels strictement positifs, et voyons ce que les conditions précédentes imposent à ces constantes.

Il est facile de vérifier que (H) est satisfaite si et seulement si $a > 2b$.

D'autre part, la divergence de l'intégrale $\int^{\infty} B_t\,\mathrm{d}t$ est équivalente à $\alpha < 1$, où on a posé $\alpha = c(F)k^{-1}$, et la condition $\lim_{t\to+\infty} A_t B_t^{-1} = 0$ est notamment satisfaite si

$$\alpha < 1, \quad \alpha < 2b, \quad \alpha < \max(a;2b) - 4b,$$

i.e.

$$\alpha < \min(1;2b;a-4b).$$

Il est dans notre intérêt de choisir a et b de telle sorte que les trois termes qui apparaissent dans le minimum ci-dessus soient égaux. En effet, cela revient à choisir les valeurs minimales de a et b (pour l'avantage que l'on a à avoir a minimal, voir la remarque ci-dessous) qui permettent d'obtenir la meilleure condition sur α (c'est -à-dire $\alpha < 1$) quant à la plus rapide des décroissances admissibles pour la température partielle β_t^{-1}.

D'où le choix de $a = 3$ et de $b = \frac{1}{2}$ dans le théorème 1. □

Remarques. (a) Les calculs précédents ne permettent pas de retrouver que dans le cas du cercle ($n = 1$) il existe une constante $k > 0$ telle que si on prend pour t assez grand

$$\beta_t = k^{-1}\ln(t), \quad \gamma_t = \ln^{-p}(t), \quad h_t = \ln^{-q}(t)$$

avec $q > 0$ et $p > 2 + 4q$, alors la conclusion du théorème 1 est encore satisfaite pour l'algorithme (2), cf. (Miclo, 1994).

Or, le fait de devoir faire croître γ_t^{-1} avec des taux beaucoup plus rapides que ceux présentés ci-dessus peut être génant d'un point de vue pratique, car cela signifie qu'il faut accélérer d'autant plus la diffusion dégénérée $(X_t^{(0)}, X_t^{(1)}, \ldots, X_t^{(n)})_{t\geq 0}$ interagissant avec l'algorithme de recuit $(\Theta_t)_{t\geq 0}$.

(b) Il est possible d'adapter la méthode précédente à la situation dégénérée où les opérateurs L_θ ne sont plus nécessairement elliptiques, dumoins si les probabilités invariantes μ_θ (dont on n'est plus assuré de l'unicité) peuvent être choisies comme limites (uniformes en un certains sens en $\theta \in N$), quand $\varepsilon > 0$ tend vers 0, de mesures invariantes $\mu_{\theta,\varepsilon}$ associées à des opérateurs non dégénérés de la forme $L_{\theta,\varepsilon}\cdot = L_\theta \cdot + \varepsilon\hat{L}\cdot$, où $\hat{L}$ est un opérateur auxiliaire de diffusion, pouvant être lui aussi dégénéré. Dans l'algorithme que l'on considére, il faut alors introduire une évolution supplémentaire, $\varepsilon : \mathbb{R}_+ \to \mathbb{R}_+^*$, servant à approximer L_θ par L_{θ,ε_t} pour t grand. Ainsi dans (Miclo, 1996) on a montré, en reprenant certaines estimées de (Miclo, 1994) pour la dépendance en ε des quantités qui interviennent dans la preuve, qu'il suffit de prendre pour t assez grand $\varepsilon_t = \ln^{-1}(\ln(t))$, avec par ailleurs les mêmes conditions sur les autres évolutions que dans le théorème 1, pour assurer la convergence escomptée (mais on peut obtenir la même conclusion par exemple avec $\beta_t = k^{-1}\ln(t)$, $\gamma_t = t^{-a}$, $h_t = t^{-b}$ et $\varepsilon_t = r\ln^{-1}(t)$, si les réels strictement positifs a, b, k et r satisfont certaines conditions, ce qui permet d'avoir des températures partielles β et ε du même ordre).

Remerciements

Je tiens tout particulièrement à remercier le referee et l'editeur Peter Jagers de m'avoir suggéré d'utiliser les mêmes browniens directeurs sur les $n+1$ premiers facteurs de l'espace des phases $M^{n+1} \times N$. Initialement, on ne s'était intéressé qu'à l'algorithme décrit par (2) et on ne considérait que deux entropies partielles: J_t et celle donnée par

$$I_t = \int \ln\left(\frac{m_t(x_0, x_1, \ldots, x_n \mid \theta)}{\mu_\theta(x_0)\mu_{\theta+h_te_1}(x_1), \ldots, \mu_{\theta+h_te_n}(x_n)}\right) m_t(\mathrm{d}x_0, dx_1, \ldots, \mathrm{d}x_n, \mathrm{d}\theta).$$

Par une méthode similaire à celle présentée ici, on vérifiait que ces deux quantités satisfont un système différentiel identique à (7) et on obtenait donc les mêmes résultats (et notamment les mêmes conditions de convergence).

Références

R. Holley, S. Kusuoka and D. Stroock, Asymptotics of the spectral gap with applications to the theory of simulated annealing. J. Funct. Anal. 83 (1989) 333–347.

R. Holley and D. Stroock, Annealing via Sobolev inequalities, Comm. Math. Phys. 115 (1988) 553–569.

N. Ikeda and S. Watanabe, Stochastic Differential Equations and Diffusion Processes (North-Holland, Amsterdam, 1981).

L. Miclo, Recuit simulé sans potentiel sur une variété riemannienne compacte, Stochastics Stochastics Rep. 41 (1992) 23–56.

L. Miclo, Un algorithme de recuit simulé couplé avec une diffusion, Stochastics Stochastics Rep. 46 (1994) 193–268.

L. Miclo, Recuit simulé partiel: Le cas dégénéré, Prépublication 10–96, Laboratoire de Statistique et Probabilités, Université Paul Sabatier (Toulouse, France, 1996).

D. Stroock, The Malliavin calculus, a functional approach, J. Funct. Anal. 44 (1981) 212–257.

S. Taniguchi, Malliavin's stochastic calculus of variations for manifold-valued Wiener functionals and its applications, Z. Wahrs. Verwandte Gebiete 65 (1983) 269–290.