ANNALES DE L'I. H. P., SECTION B

CATHERINE BOUTON

Approximation gaussienne d'algorithmes stochastiques à dynamique markovienne

Annales de l'I. H. P., section B, tome 24, nº 1 (1988), p. 131-155

<http://www.numdam.org/item?id=AIHPB_1988__24_1_131_0>

Ann. Inst. Henri Poincaré,

Vol. 24, n° 1, 1988, p. 131-155. *Probabilités et Statistiques*

Approximation gaussienne d'algorithmes stochastiques à dynamique markovienne

par

Catherine BOUTON

Laboratoire de Probabilités, Université Paris-VI,
Tour 56, C. 56-66, 3ᵉ ét., 4, place Jussieu, 75252 Paris Cedex

RÉSUMÉ. — Nous étudions l'approximation diffusion pour des algorithmes à pas constant et à pas décroissant. Nos hypothèses s'appliquent à une classe générale d'algorithmes. De plus, nous ne supposons ni la continuité ni l'existence de moments du terme d'adaptation, ce qui permet de traiter des exemples intéressants dans la pratique.

Mots clés : Algorithmes stochastiques, adaptation récursive, approximation gaussienne.

ABSTRACT. — We study the diffusion approximation of a stochastic algorithm of which the gain is either constant or going to zero. We work under hypothesis which are satisfied for a large class of stochastic algorithms. Moreover, we don't assume the continuity or the existence of moments of the adaptation term, so that theorems apply to interesting practical examples.

I. INTRODUCTION

Nous nous intéressons, dans cet article, à des algorithmes de la forme suivante :

$$\theta_{n+1}=\theta_n-\gamma_{n+1}\,h(\theta_n)-\gamma_{n+1}\,g(\theta_n,\ Y_{n+1}) \tag{1}$$

Classification A.M.S. : 60 G 35.

où $\theta_n \in \mathbb{R}^d$, (Y_n) est une suite de variables aléatoires définies sur un même espace de probabilité $(\Omega, \mathscr{F}, P)$, à valeurs dans $\mathbb{R}^k$, (γ_n) est une suite constante ou décroissante de réels strictement positifs, h est une fonction définie sur $\mathbb{R}^d$ et lipschitzienne sur un ouvert D de $\mathbb{R}^d$, et $g(\theta, y)$ une fonction mesurable de $\mathbb{R}^d \times \mathbb{R}^k$ dans $\mathbb{R}^d$.

L'étude récente de ces algorithmes se présente à beaucoup d'égards comme la version discrète de problèmes de moyenne et de fluctuation pour des équations différentielles aléatoires (*cf.* Khas'minsky [7], Freidlin et Wentzell [4], Geman [6]). Un algorithme du type (1) est considéré comme perturbation aléatoire d'un algorithme déterministe. Son comportement asymptotique est comparé au comportement asymptotique de l'algorithme déterministe et on étudie les fluctuations du deuxième par rapport au premier. S'il est vrai que l'étude du processus à temps discret défini par un algorithme à pas constant reste très analogue à l'étude du modèle différentiel, l'étude des algorithmes à pas non constant introduit des difficultés techniques propres et des termes nouveaux dans les résultats asymptotiques. Les techniques de majoration permettant d'appliquer les méthodes de martingales sont différentes. De nombreuses études ont été faites dans cette direction (*cf.* Kushner [10], Ljung [12]). Kushner a élaboré une théorie très générale, s'appliquant à une très grande classe d'algorithmes, mais il utilise des hypothèses de moments dont on sait qu'elles ne sont pas vérifiées dans des situations du type de celles considérées par Ljung. Récemment, en ce qui concerne l'étude asymptotique, Métivier et Priouret ont repris l'étude dans un cadre contenant les algorithmes stochastiques considérés par Ljung et dans un contexte d'hypothèses portant directement sur la dynamique du processus « conduisant » l'algorithme. Il s'agit en l'occurrence d'une chaîne de Markov en « rétroaction avec le paramètre » (*voir* plus loin) (*cf.* [15]).

Nous nous intéressons ici au problème de la fluctuation dans le cadre des algorithmes étudiés par Métivier et Priouret.

Nous considérons des algorithmes à dynamique markovienne, c'est-à-dire que nous supposons l'existence d'une famille de probabilités de transition sur $\mathbb{R}^k$ $\{\pi_\theta(y, dx), y \in \mathbb{R}^k, \theta \in \mathbb{R}^d\}$ telle que :

$$P(Y_{n+1} \in B / \mathscr{F}_n) = \pi_{\theta_n}(Y_n, B)$$

pour tout B borélien de $\mathbb{R}^k$, $\mathscr{F}_n$ désignant la tribu engendrée par $\{\theta_i, Y_i, i \leqq n\}$.

Les hypothèses sont toutes formulées en termes des transitions π_θ de la chaîne Y_n, et notamment en fonction de la « régularité » des solutions de l'équation de Poisson associée, et non en termes des propriétés de mélange du processus $(Y_n)_{n \geqq 0}$. Ces deux types d'hypothèses ne sont d'ailleurs pas entièrement comparables. Il est à remarquer que nos hypothèses s'appliquent en particulier à des algorithmes pour lesquels le terme d'adaptation est discontinu [Kushner et Huang (*cf.* [8]) ont étudié le problème de la fluctuation pour des algorithmes à pas décroissant, avec des hypothèses de mélange et de moments sur Y_n, et de différentiabilité du terme d'adaptation par rapport à θ].

De façon précise, nos résultats peuvent être résumés de la façon suivante :

Associons à l'algorithme (1) l'équation différentielle :

$$\frac{d\bar{\theta}}{dt}(t) = -h(\bar{\theta}(t)). \tag{2}$$

Nous supposons qu'il existe pour tout x de D une unique solution de (2), $\bar{\theta}(x, t)$, de condition initiale x, définie pour tout $t \geqq 0$ et restant dans D.

Notons $t_n = \sum_{i=1}^{n} \gamma_i$, $m(t) = \sup\{k \geqq 0 \mid t_k \leqq t\}$, $\|\gamma\| = \sup_{i \geqq 0} \gamma_i$

(1.1) Posons $\theta_t = \theta_n$ pour $t \in [t_n, t_{n+1}[$. θ_t est un processus à trajectoires cad-lag.

Sous certaines conditions de régularité sur les noyaux π_θ, sous l'hypothèse de l'existence d'une unique mesure invariante Γ_θ pour chaque chaîne π_θ, et en supposant $g(\theta, y)$ d'intégrale nulle par rapport à Γ_θ, M. Métivier et P. Priouret ont démontré des résultats concernant la convergence vers zéro de l'écart $\|\theta_t - \bar{\theta}(x, t)\|$ (*cf.* [13], [14]).

Nous démontrons dans cet article un théorème d'approximation diffusion, sous des hypothèses analogues à celles introduites dans [13].

(1.2) Posons $U_n = \dfrac{\theta_n - \bar{\theta}(x, t_n)}{\sqrt{\gamma_n}}$ et $U_t = U_n$ pour $t \in [t_n, t_{n+1}[$.

Les variables U_n vérifient la relation :

$$\begin{aligned} U_{n+1} = {} & (1 + \alpha_{n+1}\gamma_{n+1}) U_n - \sqrt{\gamma_{n+1}}\, g(\theta_n, Y_{n+1}) \\ & - \sqrt{\gamma_{n+1}}\,(h(\theta_n) - h(\bar{\theta}(x, t_n)) - \gamma_{n+1}\sqrt{\gamma_{n+1}}\, \eta_n \end{aligned} \tag{1.3}$$

où $\alpha_{n+1} = \left(\sqrt{\dfrac{\gamma_n}{\gamma_{n+1}^3}} - \dfrac{1}{\gamma_{n+1}}\right)$ et où η_n est uniformément borné pour $n \leqq m(T)$, pour tout $T > 0$.

Nous montrons que le processus U_t converge en loi vers une diffusion gaussienne de loi initiale nulle, lorsque $\|\gamma\|$ tend vers zéro, la condition initiale (x, y) étant fixée (*cf.* théorème 1, § 2).

Dans le cas particulier d'algorithmes à pas décroissant, convergeant presque sûrement vers un point θ^* asymptotiquement stable pour (2), nous montrons un autre théorème d'approximation diffusion, concernant l'écart $(\theta_n-\theta^*)$:

Posons $U_n^*=\dfrac{\theta_n-\theta^*}{\sqrt{\gamma_n}}$

$$(1.4)\qquad \begin{cases} U_t^*=U_n^* & \text{pour}\quad t\in[t_n,\,t_{n+1}[\\ U_n^{*N}=U_{t+t_N}^* & \text{pour}\quad N\in\mathbb{N}. \end{cases}$$

Nous montrons la convergence faible des processus U_t^{*N} pour N tendant vers l'infini, vers une diffusion gaussienne stationnaire (*cf.* Théorème 2, § 2). La convergence faible des variables U_n^* dans $\mathbb{R}^d$ vers la loi initiale de cette diffusion est alors une conséquence immédiate.

Notre travail se présente donc comme une contribution à l'étude des algorithmes dans un contexte markovien, faisant apparaître clairement le rôle joué par les méthodes de martingales et par l'équation de Poisson associée à la chaîne de Markov contrôlée conduisant le processus, contexte à la fois général et très près des applications.

L'article est organisé comme suit : dans le paragraphe 2 nous énonçons les deux théorèmes d'approximation diffusion de cet article. Nous démontrons le théorème 1 dans le paragraphe 3, puis le théorème 2 dans le paragraphe 4.

2. ÉNONCÉ DES THÉORÈMES

2.1. Théorème d'approximation diffusion pour des algorithmes à pas constant ou décroissant

Le problème posé dans ce paragraphe est l'étude du comportement du processus U_t (défini par 1.1) lorsque $\|\gamma\|$ tend vers zéro, la condition initiale (x, y) étant fixée dans $D\times\mathbb{R}^k$. Nous considérons donc une suite $(\gamma_i^N)_{N\geq 0}$ de pas constants ou décroissants, telle que $\|\gamma^N\|=\sup_{i\geq 0}\gamma_i^N$ tende vers zéro, et $(x, y)\in D\times\mathbb{R}^k$ fixé.

Remarque. — Pour étudier un algorithme à pas constant nous prendrons une suite γ^N de réels positifs tendant vers zéro, chaque suite γ_i^N étant constante et égale à γ^N. Pour l'étude d'un algorithme à pas décroissant $(\gamma_n)_{n \geqq 0}$, nous considérerons les pas (γ_n^N) où $\gamma_n^N = \gamma_{n+N}$.

Soit (θ_n^N) l'algorithme (1) correspondant au pas (γ_n^N) et de condition initiale (x, y). Notons θ_t^N et U_t^N les processus correspondants.

Avant d'énoncer le théorème, nous introduisons les hypothèses suivantes.

(H.1) π_θ admet, pour tout $\theta \in \mathbb{R}^d$, une unique probabilité invariante Γ_θ, et

$$\int g(\theta, y)\, d\Gamma_\theta(y) = 0$$

(H.2) La fonction h est de classe $\mathscr{C}^2$ sur D.

Soit Q un compact quelconque de l'ouvert D.

(H.3) Il existe $\rho > 0$, et $K_1(Q)$ telles que :

$$\sup_{\theta \in Q} \| g(\theta, y) \| \leqq K_1(Q)(1 + |y|)^\rho.$$

D'autre part, si φ est une fonction définie sur $\mathbb{R}^d$, on note :

$$[\varphi]_q = \sup_{y \neq z} \frac{\| \varphi(y) - \varphi(z) \|}{\| y - z \| (1 + |y|^q + |z|^q)}.$$

L'hypothèse suivante nous permettra d'avoir des propriétés de moments pour la suite Y_n (*cf.* [13]) :

(H.4) Il existe $p \geqq 2$, r entier, $r \geqq 1$ et des constantes $\alpha_1(Q) < 1$, et $M_1(Q)$ telles que :

(i) $\sup_{\theta \in Q} \int \pi_\theta^r(y, dz) |z|^p \leqq \alpha_1(Q) |y|^p + M_1(Q)$.

(ii) Pour tout $q \leqq p-1$, il existe $K_2(Q)$ tel que, pour tout $\theta \in Q$,

$$|\pi_\theta \varphi(y) - \pi_\theta \varphi(z)| \leqq K_2(Q) [\varphi]_q \, |y - z| (1 + |y|^q + |z|^q)$$

(iii) Pour tout $q < p-1$, il existe $K_3(Q)$ tel que pour $\theta \in Q$, $\theta' \in Q$

$$|\pi_\theta \varphi(y) - \pi_{\theta'} \varphi(y)| \leqq K_3(Q) [\varphi]_q \, |\theta - \theta'| (1 + |y|^{q+1})$$

En ce qui concerne les suites γ_n^N, nous supposons :

(H.5) (i) pour tout N, $\sum_{i=0}^{+\infty} \gamma_i^N = +\infty$, $\gamma_{i+1}^N \leqq \gamma_i^N$, $i \geqq 0$.

(ii) $\gamma_1^N = \sup_{i \geqq 0} \gamma_i^N$ converge vers 0, pour $N \to +\infty$.

(iii) Les suites : $\alpha_n^N = \dfrac{1}{\gamma_{n+1}^N} \left(\sqrt{\dfrac{\gamma_n^N}{\gamma_{n+1}^N}} - 1 \right)$

convergent pour tout $n \geqq 0$, lorsque N tend vers l'infini vers une limite $\bar{\alpha}$ indépendante de n, et uniformément en n.

(Nous poserons $\|\alpha\| = \sup_{i, \mathrm{N}} |\alpha_i^{\mathrm{N}}|$.)

Remarque. — 1. Pour un algorithme à pas constant $\alpha_n^{\mathrm{N}} = 0$, et $\bar{\alpha} = 0$.

2. Pour un algorithme à pas décroissant, $(\gamma_n^{\mathrm{N}}) = (\gamma_{n+\mathrm{N}})$, et la condition est vérifiée si $\alpha_n = \sqrt{\dfrac{\gamma_n}{\gamma_{n+1}^3}} - \dfrac{1}{\gamma_{n+1}}$ converge vers $\bar{\alpha}$.

Par exemple : pour $\gamma_n = \dfrac{1}{n^\alpha}$, $0 < \alpha < 1$, $\bar{\alpha} = 0$; pour $\gamma_n = \dfrac{1}{n}$, $\bar{\alpha} = \dfrac{1}{2}$.

(H.6) Il existe des fonctions v_θ, définies sur R^k, pour tout $\theta \in \mathrm{R}^d$, des réels a et b positifs, et des constantes $\mathrm{K}_4(\mathrm{Q})$ et $\mathrm{K}_5(\mathrm{Q})$ vérifiant :

(i) $(\mathrm{I} - \pi_\theta) v_\theta(y) = g(\theta, y)$;

(ii) $\sup_{\theta \in \mathrm{Q}} |v_\theta(y)| \leqq \mathrm{K}_4(\mathrm{Q})(1 + |y|^a)$;

(iii) $\sup_{\substack{\theta \in \mathrm{Q} \\ \theta' \in \mathrm{Q}}} |\pi_\theta v_\theta(y) - \pi_{\theta'} v_{\theta'}(y)| \leqq \mathrm{K}_5(\mathrm{Q}) |\theta - \theta'| (1 + |y|^b)$.

Posons pour $i, j \in \{1, \ldots, d\}$

$$\mathrm{G}_{ij}(\theta, y) = (g(\theta, y))_i (g(\theta, y))_j$$

et

$$\mathrm{F}_{ij}(\theta, y) = (g(\theta, y))_i (\pi_\theta v_\theta(y))_j$$

où l'indice i représente la i-ième coordonnée du vecteur de R^d correspondant.

(H.7) Nous supposons les fonctions :

$$a_{ij}(\theta) = \int \mathrm{F}_{ij}(\theta, y)\, d\Gamma_\theta(y)$$

et

$$b_{ij}(\theta) = \int \mathrm{G}_{ij}(\theta, y)\, d\Gamma_\theta(y)$$

localement lipschitziennes dans D.

Et de manière analogue à (H.6), nous supposons l'existence de fonctions u_θ^{ij}, de réels positifs a_1, b_1 et de constantes $\mathrm{K}_6(\mathrm{Q})$, $\mathrm{K}_7(\mathrm{Q})$ [respectivement ω_θ^{ij}, a_2, b_2, et $\mathrm{K}_8(\mathrm{Q})$, $\mathrm{K}_9(\mathrm{Q})$] tels que les conditions correspondantes à (i), (ii) et (iii) de (H.6), où $g(\theta, y)$ est remplacée par $(\mathrm{F}_{ij}(\theta, y) - a_{ij}(\theta))$ (respectivement $\mathrm{G}_{ij}(\theta, y) - b_{ij}(\theta)$)) soient vérifiées.

(H.8) p vérifie :

$$p \geqq \rho + \sup(\rho + a, \rho + 2b, a_1, a_2, b_2, 2\rho, b_1)$$

et

$$p > \sup(2a, a_1, a_2).$$

Posons $A_{ij}(\theta) = b_{ij}(\theta) + a_{ij}(\theta) + a_{ji}(\theta)$ et $A(\theta)$ la matrice d'ordre d $(A_{ij}(\theta))_{1 \leqq i \leqq d,\, 1 \leqq j \leqq d}$. Remarquons que $A(\theta)$ est symétrique positive, pour $\theta \in \mathbb{R}^d$, puisque :

$$A_{ij}(\theta) = \int (v_\theta)_i (v_\theta)_j - (\pi_\theta v_\theta)_i (\pi_\theta v_\theta)_j \, d\Gamma_\theta.$$

Soit

$$L_\theta \varphi(z) = \varphi'(z) . (\bar{\alpha} I - h'(\theta)) z + \frac{1}{2} \sum_{1 \leqq i, j \leqq d} A_{ij}(\theta) \frac{\partial^2 \varphi}{\partial x_i \partial x_j}.$$

Soit $\mathbb{D}_d$ l'espace des fonctions cad-lag de $\mathbb{R}^+$ dans $\mathbb{R}^d$, muni de la topologie de Skorokhod, (ξ_t) le processus canonique sur $\mathbb{D}_d$, et $\bar{\mathscr{F}}_t$ la filtration associée.

THÉORÈME 1. — *Sous les hypothèses* (H.1)-(H.8), *et pour toute condition initiale* $(x, y) \in D \times \mathbb{R}^k$, *la suite de processus* $(U_t^N)_{N \geqq 0}$ *converge faiblement dans* $\mathbb{D}_d$ *vers la diffusion gaussienne associée à* $L_{\bar{\theta}(x, t)}$ *et de loi initiale nulle.* (*Soit* $\tilde{P}_x$ *la loi de cette diffusion.*)

Remarque 1. — Précisons la convergence par rapport à la condition initiale (x, y) : Si $Q \times K$ est un compact de $D \times \mathbb{R}^k$, $\varphi \in \mathscr{C}_K^\infty$ alors :

$$\lim_{N \to +\infty} \sup_{(x, y) \in Q \times K} \left| E(\varphi(U_t^{N, x, y}) - \tilde{P}_x(\varphi(\xi_t)) \right| = 0$$

où $U_t^{N, x, y}$ est le processus correspondant à l'algorithme de condition initiale (x, y), et de pas (γ_n^N).

Remarque 2. — 1. Dans le cas Robbins-Monro, $\Gamma_\theta = \pi_\theta$ et $A(\theta)$ est égale à la matrice de covariance de $g(\theta, y)$ pour la loi Γ_θ.

2. Lorsque les fonctions :

$$v_\theta(y) = \sum_{n=0}^{+\infty} \pi_\theta^n (g(\theta, y))$$

sont définies et vérifient les conditions (i), (ii), (iii) de l'hypothèse (H.6), la matrice $A(\theta)$ est égale à :

$$A(\theta) = \sum_{n=-\infty}^{+\infty} R_\theta(n)$$

où $R_\theta(n) = E[g(\theta, Y_k^\theta)\, {}^t g(\theta, Y_{k+n}^\theta)]$ pour $k \geqq \sup(0, -n)$, où Y_k^θ représente la chaîne de Markov de loi initiale Γ_θ, contrôlée par π_θ.

2.2. Théorème d'approximation diffusion pour un algorithme à pas décroissant convergeant presque sûrement vers un point asymptotiquement stable de (2)

Soit un algorithme stochastique (1) à pas décroissant (γ_n), de condition initiale fixée $(x_0, y_0) \in \mathrm{D} \times \mathbb{R}^k$. Le problème est d'étudier la convergence faible des processus $\mathrm{U}_t^{*\mathrm{N}}$ définis par (1.4). Introduisons les hypothèses supplémentaires :

(H.9) (θ_n) converge presque sûrement vers $\theta^* \in \mathrm{D}$.

(H.10) La suite γ_n vérifie :

(i) $\gamma_n \downarrow 0$, $\sum_{n=0}^{+\infty} \gamma_n = +\infty$, et il existe $\alpha > 1$ tel que $\sum_{n \geqq 0} \gamma_n^\alpha < +\infty$.

(ii) La suite $\alpha_{n+1} = \sqrt{\dfrac{\gamma_n}{\gamma_{n+1}^3}} - \dfrac{1}{\gamma_{n+1}}$ converge vers $\bar{\alpha}$.

(H.11) Le point θ^* de D vérifie :

(i) $h(\theta^*)=0$, $\bar{\alpha}\mathrm{I} - h'(\theta^*)$ a toutes ses valeurs propres de partie réelle strictement négative;

(ii) pour $x \in \mathrm{D}$, $x \neq \theta^*$, $h(x).(x-\theta^*) > 0$, et il existe une boule compacte B de centre θ^*, de D telle que pour $x \in \mathrm{B}$: $(h(x).(x-\theta^*)) \geqq \delta \|x-\theta^*\|^2$, où $\delta > \bar{\alpha}$.

L'hypothèse (H.11) implique que $\|x-\theta^*\|^2$ est une fonction de Lyapounov stricte sur D, le point θ^* est donc asymptotiquement stable de l'équation différentielle (2).

THÉORÈME 2. — *On suppose que les hypothèses* (H.1)-(H.4), (H.6)-(H.11) *sont satisfaites. Alors la suite de processus* $\mathrm{U}_t^{*\mathrm{N}}$ *converge en loi dans* $\mathbb{D}_d$ *vers la diffusion gaussienne :*

$$\xi_t = e^{\bar{\mathrm{H}}t}\left(\xi_0 + \int_0^t e^{-\bar{\mathrm{H}}s} \mathrm{C}(\theta^*)\, d\mathrm{B}_s\right)$$

et de loi initiale de la loi normale $\mathrm{N}(0, \Sigma)$, *où* $\mathrm{C}(\theta^*)$ *est une racine carrée positive de* $\mathrm{A}(\theta^*)$, $\bar{\mathrm{H}} = (\bar{\alpha}\,\mathrm{I} - h'(\theta^*))$ *et*

$$\Sigma = \int_0^{+\infty} \exp(\bar{\mathrm{H}}\,t)\,\mathrm{A}(\theta^*)\exp({}^t\bar{\mathrm{H}}\,t)\,dt.$$

Le théorème 2 implique immédiatement le résultat suivant sur la vitesse de convergence de l'algorithme :

COROLLAIRE. — *Si les hypothèses* (H.1)-(H.4), (H.6)-(H.11) *sont satisfaites alors la suite de variables* $\mathrm{U}_n^* = \theta_n - \theta^*/\sqrt{\gamma_n}$ *converge en loi dans* $\mathbb{R}^d$ *vers* $\mathrm{N}(0, \Sigma)$.

Remarques. — 1. Dans le cas traité par Sacks et Gladyshev $A(\theta^*)$ est la matrice de covariance de $g(\theta^*, y)$ pour Γ_{θ^*}, et on retrouve les résultats de vitesse de convergence obtenus dans [5] et [17].

2. Si d'une part les fonctions $v_\theta = \sum_{n=0}^{+\infty} \pi_\theta^n g(\theta, y)$ sont définies et vérifient les conditions de (H.6), et si d'autre part $g(\theta^*, Y_k)$ est une suite stationnaire :

$$A(\theta^*) = \sum_{k=-\infty}^{+\infty} R_{\theta^*}(n)$$

où $R_{\theta^*}(n) = E[g(\theta^*, Y_k).{}^t g(\theta^*, Y_{k+n})]$.

On retrouve ainsi la forme des résultats de Kushner et Huang (*cf.* [8]).

3. DÉMONSTRATION DU THÉORÈME 1

Pour étudier la convergence en loi des processus U_t^N, nous montrons la compacité faible de la suite, et nous identifions la limite comme unique solution du problème de martingale associé au générateur $L_{\bar\theta(x,t)}$ et de loi initiale nulle.

Pour des raisons techniques, nous étudions la compacité faible des processus U_t^N au paragraphe 3.2 (proposition 3.2). Au paragraphe 3.1, nous considérons pour un algorithme de pas (γ), le terme :

$$\varphi(U_t) - \varphi(U_0) - \int_0^t L^{(\gamma)}(\bar\theta(x, s))(U_{s-})\, ds$$

où φ est une fonction de classe C^∞ à support compact et $L^{(\gamma)}(\theta)$ est un générateur défini au paragraphe 3.0. Nous montrons dans la proposition 3.1 que ce terme est la somme d'une martingale et d'un reste proche de 0 lorsque $\|\gamma\|$ est proche de 0. Enfin, au paragraphe 3.3 nous terminerons la démonstration en utilisant le fait que $L^{(\gamma)}(\theta)$ est proche de $L(\theta)$ lorsque $\|\gamma\|$ est proche de 0.

3.0. Notations

Nous notons $E_{\gamma, x, y}$ et $P_{\gamma, x, y}$ (respectivement $E_{N, x, y}$ et $P_{N, x, y}$) l'espérance et la probabilité lorsque les processus ou les variables à intégrer sont relatifs à l'algorithme de pas $\gamma=(\gamma_n)$ [respectivement (γ_n^{N})] et de condition initiale (x, y).

Nous énonçons dans la proposition 3.0 ci-dessous les résultats impliqués par (H.4), sur les moments des variables Y_n et sur les temps d'arrêts $\tau_{\mathrm{Q}} \wedge \nu_\varepsilon$, où :

$$\tau_{\mathrm{Q}}=\inf\{n \geqq 0 \mid \theta_n \notin \mathrm{Q}\}$$
$$\nu_\varepsilon=\inf\{n > 0 \mid |\theta_n-\theta_{n-1}| > \varepsilon\}.$$

PROPOSITION 3.0. — *Supposons les hypothèses* (H.1) *et* (H.4) *vérifiées. Alors pour tout compact* Q *de* D, *il existe une constante* C_{Q}, *et un réel* $\varepsilon_0 > 0$, *dépendant de* Q *et des données tels que, pour tout* $(x, y) \in \mathrm{Q} \times \mathrm{R}^k$, *pour toute suite* γ *décroissante et positive, et pour tout* $1 \leqq q \leqq p$, *on ait : si* $\varepsilon < \varepsilon_0$:

(i) $\sup_{n \geqq 0} \mathrm{E}_{\gamma, x, y}(\|\mathrm{Y}_n\|^q \cdot 1_{\tau_{\mathrm{Q}} \wedge \nu_\varepsilon \geqq n}) \leqq \mathrm{C}_{\mathrm{Q}}(1+|y|^q)$.

De plus, les temps d'arrêt $\tau_{\mathrm{Q}} \wedge \nu_\varepsilon$ *vérifient :*

(ii) *Soit* $\mathrm{T} > 0$, *et* Q_0 *un compact de* Q *tel que :*

$$\{\bar{\theta}(x, t)/x \in \mathrm{Q}_0,\ t \leqq \mathrm{T}\} \subsetneq \mathrm{Q}.$$

Alors pour tout $(x, y) \in \mathrm{Q}_0 \times \mathrm{R}^k$

$$\lim_{\mathrm{N} \to +\infty} \mathrm{P}_{\mathrm{N}, x, y}(\tau_{\mathrm{Q}} \wedge \nu_\varepsilon \leqq m(\mathrm{T}))=0$$

et la convergence a lieu uniformément pour $(x, y) \in \mathrm{Q}^0 \times \mathrm{K}$, *pour tout* K *compact de* R^k.

(Pour la démonstration, *cf.* [13]).

Nous fixons $(x, y) \in \mathrm{D} \times \mathrm{R}^k$, la condition initiale de l'algorithme (1). Dans toute la démonstration T désigne un réel positif fixé, et Q un compact de D contenant strictement la trajectoire $\{\bar{\theta}(x, t),\ t \leqslant \mathrm{T}\}$.

Nous choisissons pour toute la démonstration ε, $\varepsilon > 0$, de sorte que la condition (*i*) de la proposition 3.0 soit vérifiée pour Q et ε.

Nous notons alors $\tau=\tau_{\mathrm{Q}} \wedge \nu_\varepsilon$. Remarquons que t_τ $\left(\text{où } t_n=\sum_{i=1}^{n} \gamma_i\right)$ est alors un temps d'arrêt par rapport à $\mathscr{F}_t=\sigma\{\theta_s, \mathrm{Y}_s, s \leqq t\}$.

Pour tout processus A_t mesurable par rapport à $\mathscr{F}_t$, nous noterons $\mathrm{A}_t=\mathrm{A}_t \cdot 1_{\{t_\tau \geqq t\}}$.

Notons $\|f\|_{\mathrm{Q}}=\sup_{x \in \mathrm{Q}} \|f(x)\|$ pour f continue sur D.

Associons à la suite (γ) le générateur :

$$L^{(\gamma)}(\theta)\,\varphi(z)=\varphi'(z)(\alpha_t I-h'(\theta))\,z+\frac{1}{2}\sum_{1\leqq i,\,j\leqq d} A_{ij}(\theta)\frac{\partial^2\varphi(z)}{\partial x_i\,\partial x_j}$$

où $\alpha_t=\alpha_n$ sur $[t_n, t_{n+1}[$.

3.1. Étude de

$$\varphi(U_t)-\varphi(U_0)-\int_0^t L^{(\gamma)}(\bar\theta(x, s))\,\varphi(U_s-)\,ds$$

Pour $\varphi\in\mathscr{C}_K^\infty$ et pour l'algorithme (1) de pas $(\gamma)=(\gamma_n)$ donné, de condition initiale (x, y) et vérifiant, pour une constante c :

$$(*)\qquad \gamma_n\leqq c,\qquad \alpha_n\leqq c\quad \text{pour tout } n.$$

LEMME 3.1. — *Soit $\psi\in\mathscr{C}^\infty(\mathbb{R}^d, \mathbb{R})$ bornée, telle que ψ' soit bornée. Il existe pour tout $q>1$, $a_1\,q\leqq p$, une constante C dépendant des données, de Q, T, q, c mais non de (γ_n) telle que : pour tout $(i, j)\in\{1, \ldots, d\}^2$*

$$\sum_{n=0}^{m(t)-1}\gamma_{n+1}\,\psi(U_n)\,F_{ij}(\theta_n, Y_{n+1})=\int_0^t\psi(U_s-)\,a_{ij}(\theta_s-)\,ds+M_{ij}(t)+\mathscr{R}_{ij}(t)$$

où $M_{ij}(t\wedge t_\tau)$ est une martingale par rapport à $\mathscr{F}_t$ et

$$E(\sup_{t\leqq T}\|\tilde{\mathscr{R}}_{ij}(t)\|)\leqq C.\|\gamma\|^u,\quad \textit{où}\quad u=\left(1-\frac{1}{q}\right)\wedge\frac{1}{2}.$$

Remarque. — Le lemme est aussi vrai pour $G_{ij}(\theta, y)$ en remplaçant a_1 par a_2.

Démonstration. — Pour simplifier les notations, supposons $d=1$, $F_{11}(\theta, y)=F(\theta, y)$, $a_{1,1}(\theta)$, $u_{1,1}(\theta)=u(\theta)$. Grâce à l'hypothèse (H.7), nous décomposons :

$$\sum_{n=0}^{m(t)-1}\gamma_{n+1}\,\psi(U_n)\,F(\theta_n, Y_{n+1})=A_t+B_t+C_t\tag{3.0}$$

où

$$A_t=\sum_{n=0}^{m(t)-1}\gamma_{n+1}\,\psi(U_n)\,a(\theta_n)$$

$$B_t=\sum_{n=0}^{m(t)-1}\gamma_{n+1}\,\psi(U_n)\,(u_{\theta_n}(Y_{n+1})-\pi_{\theta_n}u_{\theta_n}(Y_n))$$

$$C_t = \sum_{n=0}^{m(t)-1} \gamma_{n+1} \psi(U_n)(\pi_{\theta_n} u_{\theta_n}(Y_n) - \pi_{\theta_n} u_{\theta_n}(Y_{n+1}).$$

Nous remarquons que $B_{t \wedge t_\tau}$ est une martingale par rapport à $\mathscr{F}_t$.

D'autre part :

Posons $A_t - \int_0^t \psi(U_s-) a(\theta_s-) ds = \mathscr{R}_t^1$.

Il vient immédiatement :

$$\text{(3.1)} \qquad E(\sup_{t \leqq T} \| \tilde{\mathscr{R}}_t^1 \|) \leqq \| \gamma \| . \| \psi \| . \| a \|_Q.$$

Pour étudier le terme C_t, nous le décomposons de la façon suivante :

$$C_t = C_t^1 + C_t^2$$

où

$$C_t^1 = \sum_{n=0}^{m(t)-1} \gamma_{n+1} \psi(U_n)(\pi_{\theta_n} u_{\theta_n}(Y_n) - \pi_{\theta_{n+1}} u_{\theta_{n+1}}(Y_{n+1}))$$

et

$$C_t^2 = \sum_{n=0}^{m(t)-1} \gamma_{n+1} \psi(U_n)(\pi_{\theta_{n+1}} u_{\theta_{n+1}}(Y_{n+1}) - \pi_{\theta_n} u_{\theta_n}(Y_{n+1})).$$

L'hypothèse analogue à (H.6) (iii), vérifiée par u_θ, b_1, permet d'obtenir facilement, puisque $\rho + b_1 \leqq p$:

$$\text{(3.2)} \qquad E(\sup_{t \leqq T} \| \tilde{C}_t^2 \|) = O(\| \gamma \|).$$

Nous décomposons alors le terme C_t^1 :

$$\begin{aligned} C_t^1 = & \sum_{n=0}^{m(t)-1} (\gamma_{n+1} - \gamma_n) \psi(U_n) \pi_{\theta_n} u_{\theta_n}(Y_n) \\ & + \sum_{n=0}^{m(t)-1} \gamma_n (\psi(U_n) - \psi(U_{n-1})) \pi_{\theta_n} u_{\theta_n}(Y_n) \\ & + \gamma_1 \psi(U_0) \pi_{\theta_0} u_{\theta_0}(Y_0) - \gamma_{m(t)} \psi(U_{m(t)}) \pi_{\theta_{m(t)}} u_{\theta_{m(t)}}(Y_{m(t)}) \\ = & C_t^{1,1} + C_t^{1,2} + C_t^{1,3}. \end{aligned}$$

Il est clair que (puisque $p \geqq a_1$)

$$\text{(3.3)} \qquad E(\sup_{t \leqq T} \| \tilde{C}_t^{1,1} \|) = O(\| \gamma \|).$$

D'autre part; puisque $a_1 q \leqq p$

$$(3.4) \qquad \mathrm{E}(\sup_{t \leqq \mathrm{T}} \| \tilde{\mathrm{C}}_t^{1,3} \|) \leqq 2 \| \psi \| \mathrm{E}\left(\sum_{n=0}^{m(t)-1} \gamma_n^q | \pi_{\theta_n} u_{\theta_n}(\mathrm{Y}_n) |^q \right)^{1/q}$$
$$\leqq \| \gamma \|^{1-(1/q)} (2 \| \psi \| . \mathrm{K}(\mathrm{T}, \mathrm{Q}, q)(1+|y|^{a_1}))$$

où $\mathrm{K}(\mathrm{T}, \mathrm{Q}, q)$ est une constante dépendant de Q, T, q et des données.

Enfin, pour étudier le terme $\mathrm{C}_t^{1,2}$, remarquons que la relation (1.3) vérifiée par les variables U_n, et l'hypothèse (H.3) entraînent :

$$\| \psi(\mathrm{U}_n) - \psi(\mathrm{U}_{n-1}) \| 1_{n \leqq m(\mathrm{T}) \wedge \tau} \leqq \sqrt{\|\gamma\|} . \| \psi' \| \mathrm{C}(\mathrm{Q}, c)(1+|\mathrm{Y}_n|)^p 1_{\tau \geqq n}.$$

Il s'ensuit que :

$$(3.5) \qquad \mathrm{E}(\sup_{t \leqq \mathrm{T}} \| \tilde{\mathrm{C}}_t^{1,2} \|) = o(\sqrt{\|\gamma\|}).$$

Le lemme est finalement impliqué par (3.1) à (3.5) en prenant pour M_t le terme B_t, et pour $\mathscr{R}_t$, $\mathrm{C}_t + \mathscr{R}_t^1$. □

LEMME 3.2. — *Soit* $\psi \in \mathscr{C}^\infty(\mathbb{R}^d, \mathbb{R}^d)$ bornée, telle que ψ' *soit à support compact. Pour* $q > 2$, *tel que* $aq \leqq p$, $\dfrac{a_1 q}{2} \leqq p$ *existe une constante* $\mathrm{C}(\mathrm{Q}, \mathrm{T}, q, c)$ *ne dépendant pas de* (γ_n) *telle que :*

$$\mathrm{J}_t = \sum_{n=0}^{m(t)-1} \sqrt{\gamma_{n+1}} \langle \psi(\mathrm{U}_n), g(\theta_n, \mathrm{Y}_{n+1}) \rangle$$
$$= -\int_0^t \sum_{1 \leqq i, j \leqq d} \psi'_{i,j}(\mathrm{U}_{s^-}) a_{ij}(\theta_s) \, ds + \mathrm{N}_t + \varepsilon_t$$

où $\mathrm{N}_{t \wedge t_\tau}$ *est une martingale par rapport à* $\mathscr{F}_t$ *et*

$$\mathrm{E}(\sup_{t \leqq \mathrm{T}} \| \tilde{\varepsilon}_t \|) \leqq \| \gamma \|^u \mathrm{C},$$

où $u = \dfrac{1}{2} - \dfrac{1}{q}$ $\left(\psi'_{i,j} = \dfrac{\partial \psi_i}{\partial x_j}, \ \psi_i \right.$ *étant la i-ième coordonnée de* $\left. \psi \right)$.

Démonstration. — L'hypothèse (H.6) nous permet de décomposer :

$$\mathrm{J}_t = \mathrm{N}_t^1 + \mathrm{D}_t$$

où

$$\mathrm{N}_t^1 = \sum_{n=0}^{m(t)-1} \sqrt{\gamma_{n+1}} \langle \psi(\mathrm{U}_n), v_{\theta_n}(\mathrm{Y}_{n+1}) - \pi_{\theta_n} v_{\theta_n}(\mathrm{Y}_n) \rangle$$

et

$$D_t = \sum_{n=0}^{m(t)-1} \sqrt{\gamma_{n+1}} \langle \psi(U_n), \pi_{\theta_n} v_{\theta_n}(Y_n) - \pi_{\theta_n} v_{\theta_n}(Y_{n+1}) \rangle.$$

Remarquons que $N^1_{t \wedge t_\tau}$ est une martingale par rapport à $\mathcal{F}_t$. Décomposons D_t de façon analogue à C_t dans la démonstration précédente.

$$D_t = D^1_t + D^2_t$$

où

$$D^1_t = \sum_{n=0}^{m(t)-1} \sqrt{\gamma_{n+1}} \langle \psi(U_n), \pi_{\theta_n} v_{\theta_n}(Y_n) - \pi_{\theta_{n+1}} v_{\theta_{n+1}}(Y_{n+1}) \rangle$$

et de la même manière, grâce à l'hypothèse (H.6) (iii) :

$$\text{(3.6)} \qquad E(\sup_{t \leqq T} \| \tilde{D}^2_t \|) = O(\sqrt{\| \gamma \|}).$$

De la même façon que pour C^1_t, nous décomposons :

$$D^1_t = D^{1,1}_t + D^{1,2}_t + D^{1,3}_t$$

où

$$D^{1,1}_t = \sum_{n=1}^{m(t)-1} (\sqrt{\gamma_{n+1}} - \sqrt{\gamma_n}) \langle \psi(U_n), \pi_{\theta_n} v_{\theta_n}(Y_n) \rangle$$

$$D^{1,2}_t = \sum_{n=1}^{m(t)-1} \sqrt{\gamma_n} \langle \psi(U_n) - \psi(U_{n-1}), \pi_{\theta_n} v_{\theta_n}(Y_n) \rangle.$$

Et de façon analogue à (3.3) et (3.4), on obtient :

$$\text{(3.7)} \qquad E(\sup_{t \leqq T} | \tilde{D}^{1,1}_t |) = O(\sqrt{\| \gamma \|})$$

$$\text{(3.8)} \qquad E(\sup_{t \leqq T} | \tilde{D}^{1,3}_t |) = O(\| \gamma \|^{(1/2)-(1/q)}).$$

D'autre part la relation (1.3) entraîne la majoration (où $K = \sup_{x \in S} \| x \|$, avec S support de ψ') :

$$\| \psi(U_n) - \psi(U_{n-1}) + \psi'(U_{n-1}) \sqrt{\gamma_n} g(\theta_{n-1}, Y_n) \| 1_{\tau \wedge m(T) \geqq n}$$
$$\leqq \| \gamma \| . (\| \psi' \| + \| \psi'' \|) C(Q, K, c)(1 + | Y_{n+1} |^{2p}) 1_{\tau \geqq n}.$$

Ceci permet de décomposer le terme :

$$D^{1,2}_t = - \sum_{n=0}^{m(t)-1} \gamma_{n+1} \sum_{1 \leqq i, j \leqq d} \psi'_{ij}(U_n) g_i(\theta_n, Y_{n+1}) (\pi_{\theta_n} v_{\theta_n})_j (Y_{n+1}) + D^{1,2,2}_t$$

où $D_t^{1,2,2}$ vérifie :

(3.9) $$E(\sup_{t \leqq T} |D_t^{1,2,2}|) = O(\sqrt{\|\gamma\|}).$$

En appliquant alors le lemme 3.1 aux fonctions ψ'_{ij}, et $F_{ij}(\theta, y)$ et pour $q' = q/2$, nous obtenons le lemme 3.2, avec $N_t = N_t^1 + \sum_{1 \leqq i, j \leqq d} M_{ij}(t)$. □

PROPOSITION 3.1. — *Soit* $\varphi \in \mathscr{C}^\infty(\mathbb{R}^d, \mathbb{R})$, *bornée et telle que* φ'' *soit à support compact. Il existe alors deux processus* $M_t(\varphi)$ *et* $R_t(\varphi)$, $\mathscr{F}_t$*-mesurables tels que :*

(i) $\varphi(U_t) - \varphi(U_0) = \int_0^t L^{(\gamma)}(\bar{\theta}(x, s)\, \varphi(U_s-)\, ds + M_t(\varphi) + \mathscr{R}_t(\varphi)$.

(ii) $M_{t \wedge t_\tau}(\varphi)$ *est une martingale par rapport à* $\mathscr{F}_t$.

(iii) $E(\sup_{t \leqq T} \|\tilde{\mathscr{R}}_t\|) \leqq \|\gamma\|^u . C$

où $u = \frac{1}{2} - \frac{1}{q}$ *pour* $q > 2$, $aq \leqq p$, $a_1 q \leqq 2p$, $a_2 q \leqq 2p$, *et* C *est une constante dépendant des données de* Q, T, c, x, y *mais non du pas* (γ_n) *de l'algorithme.*

Démonstration. — En appliquant la formule de Taylor à l'ordre 3, on obtient :

$$\varphi(U_t) - \varphi(U_0) = \sum_{n=0}^{m(t)-1} \varphi'(U_n).(U_{n+1} - U_n)$$
$$+ \frac{1}{2} \sum_{n=0}^{m(t)-1} \varphi''(U_n).(U_{n+1} - U_n)^2 + J_t^3$$
$$= J_t^1 + J_t^2 + J_t^3$$

où

(3.10) $$E(\sup_{t \leqq T} \|\tilde{J}_t^3\|) = O(\sqrt{\|\gamma\|}).$$

L'expression (1.3) implique d'une part :

$$J_t^1 = \int_0^t \varphi'(U_s-)((\alpha_s - h'(\bar{\theta}(x, s))\, U_s-)\, ds$$
$$- \sum_{n=0}^{m(t)-1} \sqrt{\gamma_{n+1}}\, \varphi'(U_n).g(\theta_n, Y_{n+1}) + J_t^{1,3}$$

où

(3.11) $$E(\sup_{t \leqq T} \|\tilde{J}_t^{1,3}\|) = O(\sqrt{\|\gamma\|})$$

et d'autre part :

$$J_t^2 = \frac{1}{2} \sum_{n=0}^{m(t)-1} \gamma_{n+1} \varphi''(U_n)(g(\theta_n, Y_{n+1}), g(\theta_n, Y_{n+1})) + J_t^{2,2}$$

où

(3.12) $$E(\sup_{t \leqq T} \| \tilde{J}_t^{2,2} \|) = O(\sqrt{\|\gamma\|}).$$

En appliquant alors le lemme 3.2 à la fonction $\psi = \varphi'$ pour J_t^1, puis, pour J_t^2 le lemme 3.1 avec les fonctions $\psi_{ij} = \dfrac{\partial^2 \varphi}{\partial x_i \partial x_j}$, $G_{ij}(\theta, y)$, on obtient finalement la proposition. □

3.2. Compacité faible de la suite U_t^N

Nous considérons encore l'algorithme (1) de pas (γ_n) et de condition initiale (x, y). La proposition (3.1) implique en particulier [pour $\varphi(x) = x$] que :

(3.13) $$U_t = \int_0^t (\alpha_s I - h'(\bar{\theta}(x, s)) U_{s^-} ds + M_t + \mathscr{R}_t$$

où

$$M_t = \sum_{n=0}^{m(t)-1} \sqrt{\gamma_{n+1}} (v_{\theta_n}(Y_{n+1}) - \pi_{\theta_n} v_{\theta_n}(Y_n))$$

et $E(\sup_{t \leqq T} |\tilde{\mathscr{R}}_t|) < C \|\gamma\|^u$, pour $u > 0$, comme dans la proposition 3.1.

Posons $V_t = \displaystyle\int_0^t (\alpha_s I - h'(\bar{\theta}(x, s)) U_{s^-} ds.$

Montrons les lemmes suivants :

LEMME 3.3. — *Soit $2 < q$, tel que $aq \leqq p$, alors $\forall t > 0$*

$$E(\sup_{\substack{s < T \\ 0 \leqq s-t \leqq \delta}} \| M_{t \wedge t_\tau} - M_{s \wedge t_\tau} \|^q) \leqq C(\gamma + \delta)^{q/2}$$

C étant une constante dépendant de q, Q, x, y et des données.

Démonstration. — Posons

$$\Delta\tilde{M}_{n+1} = (v_{\theta_n}(Y_{n+1}) - \pi_{\theta_n} v_{\theta_n}(Y_n)) 1_{\tau > n}.$$

On a $[M]_{t \wedge t_\tau} = \displaystyle\sum_{n=0}^{m(t)-1} \gamma_{n+1} \|\Delta\tilde{M}_{n+1}\|^2.$

Pour $q > 2$, tel que $aq \leqq p$, $M_{t \wedge t_\tau}$ a des moments d'ordre q. En appliquant alors l'inégalité de Burkhölder, il existe une constante C_q, telle que pour tout $t < T$:

$$(3.14)\quad E\Big(\sup_{\substack{s \leqq T\\ 0 \leqq s-t \leqq \delta}} \| M_{t \wedge t_\tau} - M_{s \wedge t_\tau} \|^q\Big) \leqq C_q E([M]_{(t+\delta) \wedge t_\tau} - [M]_{t \wedge t_\tau})^{q/2}.$$

D'autre part, l'inégalité de Hölder appliquée avec p' et $q' = q/2$, $\frac{1}{p'} + \frac{1}{q'} = 1$, entraîne :

$$(3.15)\quad \sum_{n=m(t)}^{m(t+\delta)-1} \gamma_{n+1} \| \Delta \tilde{M}_{n+1} \|^2 \leqq \Big(\sum_{n=m(t)}^{m(t+\delta)-1} \gamma_{n+1} \Big)^{1/p'} \Big(\sum_{n=m(t)}^{m(t+\delta)-1} \gamma_{n+1} \| \Delta \tilde{M}_{n+1} \|^{2q'} \Big)^{1/q'}.$$

En remarquant que $\sum_{n=m(t)}^{m(t+\delta)-1} \gamma_{n+1} \leqq \delta + \gamma$, on obtient le lemme. □

LEMME 3.4. — *Il existe une constante* C, *dépendant de* Q, T, c, *des données, de* (x, y) *mais indépendante de la suite* (γ_n), *telle que :*

$$E(\sup_{t \leqq T} \| \tilde{U}_t \|) \leqq C.$$

Démonstration. — La relation (3.13) implique :

$$(3.16)\quad E(\sup_{t \leqq T} \| \tilde{U}_t \|) \leqq E(\sup_{t \leqq T} \| \tilde{M}_t \|) + E(\sup_{t \leqq T} \| \tilde{R}_t \|) + (c+L) \int_0^T E(\sup_{t \leqq T} \| \tilde{U}_s \|)\, ds.$$

Pour obtenir le lemme 3.4, il suffit alors d'appliquer l'inégalité de Doob à la martingale M_t, puis le lemme de Gronwall. □

Considérons maintenant la suite de processus U_t^N, associée à la suite de pas (γ_n^N).

PROPOSITION 3.2. — *La suite* U_t^N *est faiblement compacte dans l'espace de Skrokhod* $\mathbb{D}_d$.

Démonstration. — Nous utilisons le critère de tension rappelé ci-dessous (*cf.* théorème 15.5 de [2]) :

THÉORÈME. — *Soit* X^n *une suite de processus à valeurs dans* $\mathbb{D}_d$, *telle que :*

(i) *La suite* $X^n(0)$ *est tendue dans* $\mathbb{R}^d$.

(ii) *Pour tout* T *positif, pour tous* α *et* η *positifs, il existe* δ, $0<\delta<T$, *et un entier* n_0 *tels que : pour* $n \geqq n_0$

$$P\Big(\sup_{\substack{0 \leqq t \leqq T-\delta \\ t \leqq s \leqq t+\delta}} |X_s^n - X_t^n| > \alpha\Big) \leqq \eta$$

Alors la suite X^n *est tendue dans* $\mathbb{D}_d$ *et si* $\tilde{\mathbb{P}}$ *est limite faible d'une sous-suite, alors* $\tilde{\mathbb{P}}(\mathbb{C}^d)=1$.

Remarquons que ce critère est additif. D'une part la suite $U^N(0)$ est nulle, donc tendue. D'autre part, en considérant (3.13), on peut décomposer :

$$U_t^N = (\tilde{V}_t^N + \tilde{M}_t^N + \tilde{R}_t^N) + U_t^N \cdot 1_{(t \geqq t_{\tau N})}.$$

La condition (ii) du théorème 15.5 est satisfaite pour la suite $(\tilde{V}_t^N)$, grâce au lemme 3.4, pour la suite $(\tilde{M}_t^N)$ grâce au lemme 3.3, et au théorème 8.3 de [2] (qui est aussi valable dans $\mathbb{D}_d$), et pour la suite $(\tilde{R}_t^N)$ grâce à 3.13. La proposition 3.0 (ii) concernant les temps d'arrêt $\tau_Q \wedge \nu_\varepsilon$ permet alors de conclure. □

3.3. Fin de la démonstration

L'hypothèse (H.6) entraîne que, pour $\varphi \in \mathscr{C}_K^\infty$:

$$\sup_{\substack{z \in \mathbb{R}^d, \\ t \geqq 0}} \left| L^{(\gamma^N)}(\bar{\theta}(x, t))\,\varphi - L(\bar{\theta}(x, t)\,\varphi \right|(z) \tag{3.17}$$

tend vers zéro lorsque N tend vers l'infini.

Soit $\tilde{\mathbb{P}}$ la limite faible d'une sous-suite convergente $U_t^{N_k}$. D'après la démonstration de la proposition 3.2, $\tilde{\mathbb{P}}(\mathbb{C}_d)=1$.

Posons $M_t(\varphi) = \varphi(\xi_t) - \varphi(\xi_0) - \int_0^t L_{\bar{\theta}(x,s)}\,\varphi(\xi_s)\,ds.$

Pour toute fonction ψ, définie sur $\mathbb{D}_d$, $\bar{\mathscr{F}}_s$ mesurable, continue et bornée, la fonction :

$$(M_t(\varphi) - M_s(\varphi))\,\psi$$

est continue $\tilde{\mathbb{P}}$. p. s., et bornée sur $\mathbb{D}_d$.

D'où :

$$\lim_{k \to +\infty} \tilde{E}^{N_k}((M_t(\varphi) - M_s(\varphi))\,\psi) = \tilde{E}((M_t(\varphi) - M_s(\varphi))\,\psi)$$

où $\tilde{E}^{N_k}$ désigne l'espérance par rapport à la loi de U^{N_k} et $\tilde{E}$ par rapport à $\tilde{\mathbb{P}}$.

Mais, la proposition 3.1, permet d'affirmer, en utilisant de plus le résultat (ii) de la proposition (3.0) sur les temps d'arrêt $\tau_Q \wedge \nu_\varepsilon$, et le résultat (3.17) sur la convergence des générateurs $L^{(\gamma^N)}_{\bar{\theta}(x,t)}$ que :

$$\lim_{N \to +\infty} \tilde{E}^N(M_t(\varphi) - M_s(\varphi))\,\psi = 0.$$

Il s'ensuit que $\tilde{\mathbb{P}}$ est la solution du problème de martingale associé à $L_{\bar{\theta}(x,t)}$, et de loi initiale nulle. Le théorème 1 résulte alors de l'unicité de la solution de ce problème de martingale, et de la compacité faible de la suite U^N_t (proposition 3.2). □

4. DÉMONSTRATION DU THÉORÈME 2

Dans toute la démonstration, (θ_n) est un algorithme (1) à pas décroissant (γ_n), de condition initiale $(x_0, y_0) \in D \times \mathbb{R}^k$, convergeant presque sûrement vers θ^*.

Choisissons $\varepsilon > 0$, pour que la proposition 3.0 soit satisfaite pour la boule compacte B introduite dans (H.11). Soit $\tau = \tau_B \wedge \nu_\varepsilon$ le temps d'arrêt associé.

Posons

$$\tau^N_B = \inf\{i \geqq N \mid \theta_i \notin B\}$$

$$\nu^N_\varepsilon = \inf\{i \geqq N \mid |\theta_i - \theta_{i-1}| > \varepsilon\}.$$

La convergence presque sûre de θ_n vers θ^* implique :

$$\lim_{N \to +\infty} P(\tau^B_B \wedge \nu^N_\varepsilon < +\infty) = 0. \tag{4.1}$$

Nous noterons $\tau^N = \tau^N_B \wedge \nu^N_\varepsilon$.

D'autre part nous posons $(\gamma^N_n) = (\gamma_{n+N})$. Soit $\theta^{N, x, y}_t$ les processus définis par (1.1), (1.2) associés à l'algorithme de pas (γ^N_n) et de condition initiale (x, y).

Nous notons $E_{N,x,y}$, $\mathbb{P}_{N,x,y}$ pour signifier que tous les processus ou variables à intégrer sont indexés par N, x, y [donc relatifs à l'algorithme de pas (γ_n^N) et de condition initiale (x, y)].

Remarquons que, pour tout $n_0 \geqq 0$:

$$U_t^{*N} = (U_t^{N,\theta_N,Y_N} + \bar{U}_t^N)\, 1_{\tau^{n_0} > N} + U_t^{*N}\, 1_{\tau^{n_0} \leqq N} \tag{4.2}$$

où

$$\bar{U}_t^N = \frac{\bar{\theta}(\theta_N, t_n^N) - \theta^*}{\sqrt{\gamma_n^N}} \qquad \text{pour} \quad t \in [t_n^N, t_{n+1}^N[$$

$\left(\text{avec } t_n^N = \sum_{i=N}^{n+N} \gamma_i\right)$.

4.1. Étude de U_t^{N,θ_N,Y_N}

LEMME 4.1. — *La suite Y_n est tendue dans $\mathbb{R}^k$.*

Démonstration. — Grâce à la proposition 3.0, il existe C tel que, pour tout $N \geqq 0$; $x \in B$, $y \in \mathbb{R}^k$:

$$\sup_{n \geqq 0} E_{N,x,y}(|Y_n|^p\, 1_{\tau \geqq n}) \leqq C(1 + |y|^p)$$

(θ_n, Y_n) étant une chaîne de Markov, on a pour $N \geqq 0$, $n \geqq N$, $n = N + m$:

$$P(|Y_n|^p . 1_{\tau^N \geqq n} > a) = \int P_{N,x,y}(|Y_m|^p\, 1_{\tau \geqq m} > a)\, dP_{\theta_N, Y_N}(x, y)$$

[où P_{θ_N, Y_N} est la loi du couple (θ_N, Y_N)]. Y_N étant finie presque sûrement, on obtient, par convergence dominée :

$$\lim_{a \to +\infty} \sup_{n \geqq N} P(|Y_n|^p . 1_{\tau^N \geqq n} > a) = 0.$$

Le lemme résulte alors de (4.1). □

PROPOSITION 4.1. — *La suite de processus $(U_t^{N,\theta_N,Y_N}) . 1_{\theta_N \in B}$ converge faiblement dans $\mathbb{D}_d$ vers la diffusion gaussienne de générateur L_{θ^*} et de loi initiale nulle.*

Démonstration. — Soit $\varphi \in \mathscr{C}_K^\infty(\mathbb{R}^d, \mathbb{R})$, et soit $B' \subset B$. Comme dans le théorème 1, nous notons $\tilde{\mathbb{P}}_x$ la loi de la diffusion gaussienne de générateur $L(\bar{\theta}(x, t))$, $\tilde{E}_x$ l'espérance pour cette loi. Nous avons :

$$|E(\varphi(U_t^{N,\theta_N,Y_N}) . 1_{\theta_N \in B}) - \tilde{E}_{\theta^*}(\varphi(\xi_t))|$$

$$\leq \int_{B\times\mathbb{R}^k} |E(\varphi(U_t^{N,x,y})) - \tilde{E}_x(\varphi(\xi_t))| \, dP_{\theta_N, Y_N}(x, y)$$

$$+ \int_{B'\times\mathbb{R}^k} |\tilde{E}_x(\varphi(\xi_t)) - \tilde{E}_{\theta^*}(\varphi(\xi_t))| \, dP_{\theta_N, Y_N}(x, y)$$

$$+ 2\|\varphi\| . P(\theta_N \notin B').$$

Le premier terme tend vers zéro, lorsque N tend vers l'infini, grâce à la remarque 1 du théorème 1, et grâce à la tension de la suite Y_N.

D'autre part, il est facile de voir que :

$$\lim_{x\to\theta^*} \sup_{t\leq T} |L_{\bar{\theta}(x,t)}\varphi - L_{\theta^*}\varphi| = 0.$$

Ceci implique la convergence faible de $\tilde{\mathbb{P}}_x$ vers $\tilde{\mathbb{P}}_{\theta^*}$, pour $x \to \theta^*$, grâce au théorème 11.1.4 de Strook et Varadhan (*cf.* [18]). Par suite on peut choisir la boule B′, de centre θ^* de sorte que le deuxième terme soit arbitrairement petit, pour tout $N \in \mathbb{N}$. B′ étant ainsi choisi, le dernier terme tend vers zéro, pour $N \to +\infty$. □

4.2. Étude de la suite $\bar{U}_t^N 1_{\tau^{n_0} > N}$, pour $n_0 \geq 0$

Il est facile de voir que $\bar{U}_t^N$ vérifie la relation sur $\{\theta_N \in B\}$:

(4.3) $$\bar{U}_t^N = \bar{U}_0^N + \int_0^t (\bar{\alpha} I - h'(\theta^*)) \bar{U}_{s^-}^N \, ds + \bar{\mathcal{R}}_t$$

où pour tout $T > 0$, il existe une constante $C(B, \varepsilon, T)$, indépendante de (γ_n^N), telle que :

$$E(\sup_{t\leq T} \|\bar{\mathcal{R}}_t\| 1_{\theta_N \in B}) \leq C(\sqrt{\|\gamma\|^N} + \sup_{n\geq 0} \|\alpha_n^N - \bar{\alpha}\|).$$

D'autre part, remarquons que : sur $\{\theta_N \in D\}$

$$\bar{U}^N(0) = U^{*N}(0) = \frac{\theta_N - \theta^*}{\sqrt{\gamma_N}}.$$

Pour étudier la compacité faible de $\bar{U}_t^N . 1_{\tau^{n_0} > N}$, nous allons montrer que U_N^* est tendue dans $\mathbb{R}^d$.

PROPOSITION 4.2. — $\sup_{n\geq 0} E(\|U_n^*\|^2 1_{\tau > n}) < +\infty.$

Remarque. — Il s'ensuivra que, pour tout $n_0 \geq 0$.

$$\lim_{a\to+\infty} (\sup_{n\geq n_0} P(\|U_n^*\|^2 . 1_{\tau^{n_0} > n} > a)) = 0.$$

puis, par convergence presque sûre de θ_n (4.1), que U_n^* est tendue dans $\mathbb{R}^d$.

D'autre part, la tension de la suite $\bar{U}^N(0)$, et la relation (4.3) entraîneront par application du lemme de Gronwall le corollaire suivant :

COROLLAIRE 4.2. – *Pour tout* $n_0 \geqq 0$, *la suite de processus* $\bar{U}_t^N \,.\, 1_{\tau^{n_0} > N}$ *est faiblement compacte dans* $\mathbb{D}_d$.

Démonstration de la proposition 4.2. – Nous admettons le lemme suivant, déduit de (H.10) (*cf.* [3]).

LEMME 4.3. – *Soit* $N_0 = \inf \{n \geqq 0 \mid 1 - 2\delta\gamma_n > 0\}$. *Soit* $\delta > \bar{\alpha}$.

(i) *Il existe une constante* K, *telle que :*

$$\prod_{k=N_0}^{n} (1 - 2\delta\gamma_k) \leqq K\gamma_n \qquad pour \quad n \geqq N_0.$$

(ii) *Il existe* $N \geqq N_0$, *et une constante* K_1 *telle que : pour* $n \geqq N$

$$\sum_{k=N_0}^{n} \prod_{j=k+1}^{n} (1 - 2\delta\gamma_{j+1}) \gamma_{k+1}^2 \leqq K_1 \gamma_{n+1}.$$

(iii) *La suite, pour* $n \geqq N$:

$$u_{k,n} = \gamma_{k+1} \prod_{j=k+1}^{n} (1 - 2\delta\gamma_{j+1})$$

est croissante par rapport à k, pour $k \geqq N$ *et n fixé.*

Pour la démonstration de la proposition 4.2, supposons $\theta^* = 0$, ce qui ne restreint pas la généralité. En prenant la norme au carré, puis l'espérance dans (1), on obtient :

$$\begin{aligned} &E(\|\theta_{n+1}\|^2 1_{\tau > n}) \\ &\qquad \leqq E(\|\theta_n\|^2 1_{\tau > n}) - 2\gamma_{n+1} E(\langle \theta_n, g(\theta_n, Y_{n+1}) \rangle 1_{\tau > n}) \\ &\qquad\qquad + \gamma_{n+1}^2 E(\|g(\theta_n, Y_{n+1})\|^2 1_{\tau > n}) \\ &\qquad\qquad + \gamma_{n+1}^2 E(2h(\theta_n) . g(\theta_n, Y_{n+1}) 1_{\tau > n}). \end{aligned}$$

Puis, en utilisant la minoration de $h(x) . (x - \theta^*)$ par $\delta \|x\|^2$ (H.11), en majorant les moments des deux derniers termes, et en itérant l'inégalité précédente, on obtient :

$$(4.4) \quad E(\|\theta_{n+1}\|^2 1_{\tau > n}) \leqq \prod_{k=N}^{n} (1 - 2\gamma_{k+1}\delta) E(\|\theta_N\|^2 1_{\tau_B \wedge \nu_\varepsilon > N})$$

$$+\sum_{k=N}^{n}\gamma_{k+1}^{2}\prod_{j=k+1}^{n}(1-2\delta\gamma_{j+1})(K(1+|y_0|^{2p})$$

$$-2E\left(\sum_{k=N}^{n}u_{k,n}\langle\theta_k, g(\theta_k, Y_{k+1})\rangle 1_{\tau>k}\right)$$

où K est une constante dépendant de B, ε et des données. Le lemme 4.3 permet d'affirmer que les deux premiers termes sont en $O(\gamma_n)$. En utilisant l'hypothèse (H.6), nous remarquons que le troisième terme est égal à $-2E(A_n)$, où :

$$A_n=\sum_{k=N}^{n}u_{k,n}\theta_k.\Delta Z_{k+1}1_{\tau>k}$$

avec

$$\Delta Z_{k+1}=\pi_{\theta_k}v_{\theta_k}(Y_k)-\pi_{\theta_k}v_{\theta_k}(Y_{k+1}).$$

Soit

$$\Delta Z_{k+1}=\Delta Z_{k+1}^{(1)}+\Delta Z_{k+1}^{(2)}$$

où

$$\Delta Z_{k+1}^{(1)}=\pi_{\theta_k}v_{\theta_k}(Y_k)-\pi_{\theta_k}v_{\theta_k}(Y_{k+1})$$

et

$$\Delta Z_{k+1}^{(2)}=\pi_{\theta_{k+1}}v_{\theta_{k+1}}(Y_{k+1})-\pi_{\theta_k}v_{\theta_k}(Y_{k+1}).$$

Soit A_n^1 et A_n^2 les sommes analogues à A_n, relatives à $\Delta Z_{k+1}^{(1)}$ et $\Delta Z_{k+1}^{(2)}$. L'hypothèse (H.6) (iii) permet de voir aisément que $E(\|A_n^{(2)}\| 1_{\tau>n})$ est un $O(\gamma_n)$. D'autre part, le terme A_n^1 peut se décomposer, grâce à la transformation de Abel :

$$A_n^1=\sum_{k=N+1}^{n}(u_{k,n}\theta_k 1_{\tau>k}-u_{k-1,n}\theta_{k-1}1_{\tau>k-1})\pi_{\theta_k}v_{\theta_k}(Y_k)+A_n^{1,2}$$

où le terme $E(\|A_n^{1,2}\| 1_{\tau>n})$ est un $O(u_{N,n}+u_{n,n})$ et par conséquent un $O(\gamma_n)$ en appliquant le lemme 4.3. Puis, pour le terme $A_n^{1,2}$ nous utilisons la décomposition :

$$\begin{aligned}(u_{k,n}\theta_k 1_{\tau>k}&-u_{k-1,n}\theta_{k-1}1_{\tau>k-1})\\ &=(u_{k,n}-u_{k-1,n})\theta_{k-1}1_{\tau>k-1}-u_{k,n}\theta_{k-1}1_{\tau=k}\\ &\qquad+u_{k,n}\gamma_{k+1}(h(\theta_{k-1})+g(\theta_{k-1}, Y_{k-1}))1_{\tau>k}.\end{aligned}$$

La suite $u_{k,n}$ est croissante en k, et $u_{n,n}=\gamma_{n+1}$, chacune des espérances des termes correspondant à cette décomposition est en $O(\gamma_n)$, grâce au lemme 4.3. D'où la proposition 4.2. □

PROPOSITION 4.3. — *Soit $n_0 \geqq 0$. Toute valeur d'adhérence $\tilde{\mathbb{P}}$ de la suite de processus $\bar{U}_t^N . 1_{\tau^{n_0} > N}$ est telle que :*

$$\xi_t = e^{(\bar{\alpha} I - h'(\theta^*))\, t} . \xi_0 \; \tilde{\mathbb{P}} . p. s.$$

Démonstration. — Pour $\varphi \in \mathscr{C}_K^\infty$, la relation suivante est vérifiée :

$$\varphi(\bar{U}_t^N) - \varphi(\bar{U}_0^N) = \int_0^t (\bar{\alpha} I - h'(\theta^*))\, \varphi'(U_{s^-}^N) . U_{s^-}^N \, ds + E_t^N$$

où $\lim_{N \to +\infty} E(\sup_{t \leqq T} \| E_t^N \| \, 1_{\tau^{n_0} > N}) = 0.$

On conclut alors comme dans la troisième étape du paragraphe 3. □

4.3. Fin de la démonstration

Soit $n_0 \geqq 0$, nous considérons la décomposition donnée par (4.2). La suite produit $(U^{N, \theta_N, Y_N}, \bar{U}_t^N)\, 1_{\tau^{n_0} > N}$ est tendue grâce aux propositions 4.1 et 4.2 et pour toute sous-suite convergente la somme $(U_t^{N, \theta_N, Y_N} + \bar{U}_t^N)\, 1_{\tau^{n_0} > N}$ converge vers une diffusion gaussienne de générateur L_{θ^*}. L'entier n_0 pouvant être pris arbitrairement grand, on en déduit que la suite U_t^{*N} est compacte dans $\mathbb{D}_d$, et que toute valeur d'adhérence est une diffusion gaussienne de générateur L_{θ^*}. Mais les processus U_t^{*N} étant des translatés $U_{t+t_N}^{*0}$, toute valeur d'adhérence est stationnaire. Or l'hypothèse (H.11) (i) assure que la diffusion gaussienne de générateur $L(\theta^*)$ admet une loi initiale stationnaire $N(0, \Sigma)$. On en déduit le théorème 2. □

RÉFÉRENCES

[1] A. BENVENISTE, *Introduction à la méthode de l'équation différentielle moyenne pour l'étude des algorithmes récursifs, exemples*, C.N.R.S., *Outils et Modèles mathématiques pour l'Automatique, l'Analyse des Systèmes, et le Traitement du Signal*, vol. **1**, 1981.

[2] P. BILLINGSLEY, *Convergence of Probability Measures*, Wiley, 1968.

[3] C. BOUTON, Approximation gaussienne d'algorithmes stochastiques à dynamique markovienne, *Thèse de 3e cycle*, éditée par l'École Polytechnique.

[4] M. I. FREIDLIN et A. D. WENTZELL, *Random Perturbations of Dynamical Systems*, Springer, 1984.

[5] E. G. GLADYSHEV, On Stochastic Approximation, *Theory of Probability and its Applications*, vol. **10**, p. 275-278.

[6] S. GEMAN, *Approximate Solution of Random Equations*, Bharucha-Reid, 1979.

[7] R. Z. KHAS'MINSKY, On Stochastic Processes Defined by Differential Equations with a Small Parameter, *Theory Prob. Appl.*, vol. **II**, 1966, p. 211-222.

[8] H. J. KUSHNER et H. HUANG, Rates of convergence for stochastic approximation type algorithms, *S.I.A.M. J. Control*, vol. **17**, n° 5, 1979.

[9] H. J. KUSHNER, A Martingale Method for the Convergence of a Sequence of Processes to a Jump-Diffusion Process, *Z.W.*, vol. **53**, 1980, p. 209-219.

[10] H. J. KUSHNER, *Approximation and Weak Convergence Methods for Random Processes*, M.I.T. Press, Cambridge, 1984.

[11] L. LJUNG et T. SODERSTRÖM, *Theory and Practice of Recursive Identification*, M.I.T. Press, 1983.

[12] L. LJUNG, Analysis of Recursive Stochastic Algorithms, *I.E.E.E. Trans. on Autom. Control*, vol. **AC 22**, n° 4, 1977.

[13] M. METIVIER et P. PRIOURET, Théorèmes de convergence presque sûre pour une classe d'algorithmes à pas décroissants, *Probability Theory* (à paraître).

[14] M. METIVIER et P. PRIOURET, Convergence avec probabilité $(1-\varepsilon)$ d'algorithmes stochastiques et application à l'égaliseur aveugle, *Ann. des Télécommunications*, t. **41**, n° 5-6, 1986.

[15] M. METIVIER et P. PRIOURET, Application of a Kushner and Clark Lemma to General Classes of Stochastic Algorithms, *I.E.E.E. Trans. Inf. Theory*, vol. **IT-30**, 1984, p. 140-150.

[16] G. C. PAPANICOLAOU et N. KOLHER, Asymptotic Theory of Mixing Stochastic Ordinary Differential Equations, *Comm. Pure Appl. Math.*, vol. **27**, 1974, p. 641-668.

[17] SACKS, Asymptotic Distribution of Stochastic Approximation Procedures, *Ann. Math. Stat.*, vol. **29**, 1958, p. 373-405.

[18] D. W. STROOK et S. R. S. VARADHAN, *Multidimensionnal Diffusion Processus*, Springer-Verlag, Berlin, 1979.

(*Manuscrit reçu le* 19 *juin* 1986)
(*corrigé le* 3 *juillet* 1987.)