STATISTIQUE ET ANALYSE DES DONNÉES

JEAN-MARIE MONNEZ

Convergence presque sûre de processus d'approximation stochastique dynamique

Statistique et analyse des données, tome 14, nº 2 (1989), p. 55-79

<http://www.numdam.org/item?id=SAD_1989__14_2_55_0>

© Association pour la statistique et ses utilisations, 1989, tous droits réservés.

L'accès aux archives de la revue « Statistique et analyse des données » implique l'accord avec les conditions générales d'utilisation (http://www.numdam.org/conditions). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme Numérisation de documents anciens mathématiques

http://www.numdam.org/

Statistique et Analyse des Données
1989 - Vol. 14 n° 2 - p. 55 - 79

CONVERGENCE PRESQUE SURE DE PROCESSUS D'APPROXIMATION STOCHASTIQUE DYNAMIQUE

Jean-Marie MONNEZ

Laboratoire de Mathématiques
Université de Nancy I - B.P. 239
54506 Vandoeuvre-les-Nancy cedex

Résumé : *Nous donnons dans cet article des théorèmes de convergence presque sûre de deux processus d'approximation stochastique dynamique généralisant celui de Robbins-Monro, lorsque le paramètre à estimer varie en fonction du temps. Nous nous plaçons dans le cas où l'espace des paramètres* $\mathbb{R}^k$ *est différent de celui des observations* $\mathbb{R}^p$. *Nous établissons un théorème général pour l'ensemble de l'étude dont la démonstration est basée sur l'utilisation d'un lemme de la théorie des martingales.*

Abstract : *We give in this paper almost sure convergence theorems of two dynamic stochastic approximation processes of the Robbins-Monro type, when the parameter to be estimated varies with time. We consider the case where the parameter space* $\mathbb{R}^k$ *and the observation space* $\mathbb{R}^p$ *are different. A general theorem for the whole study is established, the proof of which is achieved by using a martingale convergence lemma.*

Indices de classification STMA : 04-100.

1 - INTRODUCTION

Soit une fonction M de $\mathbb{R}$ dans $\mathbb{R}$. Soit l'équation $M(x) = 0$ de solution θ. On suppose que, pour tout x, on ne peut pas observer $M(x)$, mais seulement $M(x) + Z(x) = Y(x)$, $Z(x)$ étant un bruit aléatoire de moyenne nulle.
Pour estimer θ, Robbins et Monro [14] ont défini le processus stochastique (X_n) tel que pour $n \geq 1$:

$$X_{n+1} = X_n - a_n Y_n.$$

Manuscrit reçu le 22 novembre 1988, révisé le 22 février 1989

Y_n est une variable aléatoire réelle dont la loi de probabilité conditionnelle par rapport à la tribu du passé, T_n, engendrée par $X_1, Y_1, \ldots, Y_{n-1}$, est la loi de $Y(x_n)$, x_n étant la réalisation de X_n. En pratique, la réalisation de Y_n est une observation de $Y(x_n)$ indépendante à x_n fixé des précédentes. Cette hypothèse est dite "d'indépendance" ou "de Robbins-Monro". a_n est un nombre réel positif.

A la suite de Robbins et Monro, Kiefer et Wolfowitz [10] ont défini un processus du même type permettant d'estimer le zéro de la dérivée de M. Après ces deux articles de base, diverses généralisations de ces processus ont été étudiées. On peut, par exemple, considérer le modèle suivant :

$$X_{n+1} = X_n - a_n A_n (X_n, Y_n).$$

X_n est une variable aléatoire dans $\mathbb{R}^k$, Y_n une variable aléatoire dans $\mathbb{R}^p$, A_n une application mesurable de $\mathbb{R}^k \times \mathbb{R}^p$ dans $\mathbb{R}^k$.

Les différents modes de convergence de ces processus ont été démontrés. En particulier, en ce qui concerne la convergence presque sûre, on a utilisé très tôt la théorie des martingales (Blum [5], Gladyshev [8]). De nombreuses applications ont été mises en évidence, par exemple dans les domaines de la régression [1] et de l'estimation de paramètres [13]. La bibliographie sur ces sujets est abondante. On peut trouver un grand nombre de références dans [21], [12], [4].

Cependant, un certain nombre d'applications ont conduit à la définition d'algorithmes pour lesquels l'hypothèse d'indépendance, vue plus haut, n'est pas vérifiée. Par exemple, en filtrage adaptatif linéaire, on a étudié l'algorithme de Widrow :

$$X_{n+1} = X_n - a_n Z_n (Z'_n X_n - c_n).$$

c_n représente un signal de référence et Z_n un vecteur de signaux observés ; les couples (c_n, Z_n) ne sont pas indépendants. Différents articles sur ce sujet sont rassemblés dans [9]. Un autre exemple peut être pris dans le domaine de l'identification des systèmes linéaires ; Ljung considère pour le vecteur d'état du système, Y_n, une dynamique markovienne conditionnellement linéaire : $Y_n = A(X_n) Y_{n-1} + B(X_n) W_n$, où (W_n) est une suite de variables aléatoires indépendantes de même loi ; on peut en trouver l'étude dans [11]. Un exposé systématique sur ce type d'algorithmes, avec une large bibliographie, est fait dans [4].

Dans cet article, nous nous intéressons au cas où le paramètre θ à estimer varie dans le temps, en supposant cependant l'hypothèse de Robbins-Monro vérifiée. θ_n peut par exemple représenter le véritable état d'un système au temps n, que l'on cherche à estimer. Les processus correspondants sont dits d'approximation stochastique dynamique.

Le premier modèle que nous traitons a été étudié par Dupac dans [6], [7], puis en particulier par Uosaki dans [17], [18], [19]. On suppose que $\theta_n = f_n(\theta_{n-1}) + v_n$, f_n étant une application connue de $\mathbb{R}^k$ dans $\mathbb{R}^k$ et v_n un élément inconnu de $\mathbb{R}^k$. Nous donnons un théorème de convergence presque sûre du processus correspondant (théorème 1) adapté au cas où les espaces des paramètres ($\mathbb{R}^k$) et des observations ($\mathbb{R}^p$) sont différents, puis une généralisation de ce résultat (théorème 2).
Nous définissons ensuite un deuxième modèle, composition du précédent et d'un modèle étudié par Ruppert [16]. Soit $k = p$, $\theta_n = (\theta_n^1, \theta_n^2, \ldots, \theta_n^p)$. Soit p entiers $k_1, k_2, \ldots, k_p$. On suppose que, pour $i = 1, 2, \ldots, p$, $\theta_n^i = \langle\beta_n^i, U_n^i\rangle$, β_n^i étant un élément inconnu et U_n^i un élément connu au temps n de $\mathbb{R}^{k_i}$, et qu'il existe une application f_n^i de $\mathbb{R}^{k_i}$ dans $\mathbb{R}^{k_i}$ et un élément inconnu v_n^i de $\mathbb{R}^{k_i}$ tels que $\beta_n^i = f_n^i(\beta_{n-1}^i) + v_n^i$. Pour $p = 1$, $f_n^1(\beta) = \beta$, $v_n = 0$, on retrouve le modèle de Ruppert.
Nous donnons un théorème de convergence presque sûre de ce processus (théorème 3) comme corollaire du théorème 1.
L'origine de cette étude est dans [12]. La première forme du théorème 1 a été énoncée dans [2].

2 - ETUDE D'UN PREMIER MODELE D'APPROXIMATION STOCHASTIQUE DYNAMIQUE

2.1 - Le modèle

Soit la famille de variables aléatoires observables dans $\mathbb{R}^p$, $\{Y_n(x), n \in \mathbb{N}^*, x \in \mathbb{R}^k\}$. θ_n est zéro de $E[Y_n(x)]$. On suppose qu'il existe, pour tout $n \geq 2$, une application connue de $\mathbb{R}^k$ dans $\mathbb{R}^k$, f_n, et un élément inconnu de $\mathbb{R}^k$, v_n, tels que $\theta_n = f_n(\theta_{n-1}) + v_n$. (1)
Pour tout n, soit A_n^1 une application mesurable de $\mathbb{R}^k$ dans $\mathbb{R}^{k\times p}$. On définit le processus d'approximation stochastique (X_n) dans $\mathbb{R}^k$ tel que, pour $n \geq 1$:

$$X_n = f_n(X_{n-1}) - a_n A_n^1(f_n(X_{n-1}))Z_n .$$

Z_n est une variable aléatoire dans $\mathbb{R}^p$ dont la loi de probabilité conditionnelle par rapport à la tribu du passé T_n engendrée par $X_1, Z_1, \ldots, Z_{n-1}$ est la loi de $Y_n(f_n(x_{n-1}))$, x_{n-1}

étant la réalisation de X_{n-1}. On pose par convention : $f_1(x) = x$, $v_1 = 0$; on choisit au début X_o ; $a_n \in \mathbb{R}^+$.

De façon plus générale, en considérant la famille des variables aléatoires $Y_n(\dot{x})$ et une suite (θ_n) d'éléments de $\mathbb{R}^k$ vérifiant la relation (1), nous allons étudier la convergence du processus (X_n) défini par :

$$X_n = f_n(X_{n-1}) - a_n\, A_n(f_n(X_{n-1}), Z_n).$$

Z_n est une variable aléatoire définie comme ci-dessus, A_n une application mesurable de $\mathbb{R}^k \times \mathbb{R}^p$ dans $\mathbb{R}^k$.

On note dans la suite : $M_n(x) = E[A_n(x, Y_n(x))]$; $V_n = f_n(X_{n-1})$.

2.2 Théorème de convergence presque sûre

Théorème 1 :

On fait les hypothèses :

1i. $\forall\, n, \exists\, \gamma_n, \delta_n \in \mathbb{R}^+ : \forall\, x,\ \| f_{n+1}(x) - f_{n+1}(\theta_n) \|^2 \le (1 + \gamma_n)\, \| x - \theta_n \|^2 + \delta_n$.

ii. $\sum_1^\infty \gamma_n < \infty$; $\sum_1^\infty \delta_n < \infty$.

iii. $\forall\, 0 < \varepsilon < 1,\ \lim_{n\to\infty} \sup_{\{\varepsilon < \| x-\theta_n \| < \frac{1}{\varepsilon}\}} \| f_{n+1}(x) - x \| = 0.$

iv. $\sum_1^\infty \| v_n \| < \infty.$

2. $\exists\, d, e \in \mathbb{R}^+ : \forall\, n, \forall\, x,\ E[\| A_n(x;Y_n(x)) \|^2] \le d\, \| x - \theta_n \|^2 + e.$

3i. $\forall\, n, \forall\, x, \langle x - \theta_n, M_n(x) \rangle \ge 0.$

ii. $\exists\, q \in \mathbb{N}^*, \exists\, (n_\ell, \ell \ge 1) \subset \mathbb{N} :\ \forall \ell, n_{\ell+1} \ge n_\ell + q,\ \ \forall\, 0 < \varepsilon < 1,$

$\exists\, L(\varepsilon) \in \mathbb{N} : b(\varepsilon) > 0$ avec

$$b(\varepsilon) = \inf_{\ell > L(\varepsilon)} \ \inf_{\{\varepsilon < \| x-\theta_{n_\ell} \| < \frac{1}{\varepsilon}\}} \sum_{j \in I_\ell} \langle x - \theta_j, M_j(x) \rangle,\ I_\ell = \{n_\ell, n_\ell + 1, \dots, n_\ell + q - 1\}.$$

iii. $\forall \varepsilon > 0, \exists \eta > 0 : (\|x_1 - x_2\| < \eta) \Rightarrow (\sup_n \|M_n(x_1) - M_n(x_2)\| < \varepsilon)$.

4i. $\forall n,\ a_n > 0,\ \sum_1^{\infty} \min_{j \in I_\ell} a_j = +\infty$.

ii. $\sum_1^{\infty} a_n^2 < \infty$.

5. $E[\|X_0\|^2] < \infty$.

Sous les hypothèses 1, 2, 3, 4, 5, on a : $\|X_n - \theta_n\| \xrightarrow{p.s.} 0$.

Pour $q = 1$, on a $\|X_n - \theta_n\| \xrightarrow{p.s.} 0$ sous les hypothèses 1i, 1ii, 1iv, 2, 3i, 3ii, 4, 5. ■

Remarques :

1. L'hypothèse 3ii est une extension de l'hypothèse de stabilité classique où I_ℓ est réduit à un seul élément ($q = 1$) ; elle est mieux adaptée au cas où les espaces des paramètres ($\mathbb{R}^k$) et des observations ($\mathbb{R}^p$) sont différents, comme le montre l'exemple ci-dessous.

Soit un système dynamique où l'état θ_n au temps n, élément de $\mathbb{R}^k$, vérifie :

$$\theta_n = f_n(\theta_{n-1}) + v_n.$$

f_n est une application connue de $\mathbb{R}^k$ dans $\mathbb{R}^k$, v_n un élément inconnu de $\mathbb{R}^k$.

Au temps n, on fait l'observation aléatoire Y_n^* dans $\mathbb{R}^p$:

$$Y_n^* = B_n\, \theta_n + Z_n.$$

B_n est une matrice réelle $p \times k$, Z_n une variable aléatoire dans $\mathbb{R}^p$ d'espérance mathématique nulle.

Pour estimer (θ_n), on définit le processus d'approximation stochastique dynamique (X_n) dans $\mathbb{R}^k$:

$$X_n = f_n(X_{n-1}) - a_n\, B'_n(B_n\, f_n(X_{n-1}) - Y_n^*).$$

Dans ce cas, on considère la famille des variables aléatoires dans $\mathbb{R}^p$ $Y_n(x) = B_n\, x - Y_n^*(x \in \mathbb{R}^k)$; on a $E[Y_n(x)] = B_n(x - \theta_n)$, $A_n^1(x) = B'_n$, $\langle x - \theta_n, A_n^1(x)\, E[Y_n(x)]\rangle = \langle x - \theta_n, B'_n\, B_n(x - \theta_n)\rangle$.

Pour $p < k$, $rg(B'_n B_n) = rgB_n \leq p$; l'hypothèse 3ii, avec $q = 1$, n'est pas vérifiée. Par contre, elle peut l'être pour $qp \geq k$: on considère

$$\sum_{j\in I_\ell} <x - \theta_j, B'_jB_j (x - \theta_j)> = <x - \theta_{n_\ell}, \sum_{j\in I_\ell} B'_jB_j (x - \theta_{n_\ell})> + R_\ell \ ;$$

en supposant $\theta_n - \theta_{n+1} \to 0$ et $\|B_j\|$ uniformément majoré, 3ii est vérifiée si la plus petite valeur propre de $\sum_{j\in I_\ell} B'_jB_j$ est uniformément minorée.

2. Les hypothèses sur la suite de fonctions (f_n) sont vérifiées, par exemple, pour $f_n(x) = x + \frac{1}{n}$. On remarque que, dans ce cas, la suite (θ_n) n'est pas convergente ; on a bien un modèle dynamique.
On constate que, dans le cas $q = 1$, l'hypothèse restrictive 1iii sur la suite (f_n) est inutile. En outre, si les fonctions M_n vérifient l'hypothèse $\exists K > 0 : \forall j, \forall x, <x - \theta_j, M_j(x)> \geq K \|x - \theta_j\|^2$, il suffit que $\gamma_n = o(a_n)$ pour obtenir le résultat de convergence (on peut l'établir facilement à partir de (2) dans la démonstration).
Enfin, en utilisant un théorème de Dvoretzky généralisé par Venter [20], on peut montrer la convergence en ne faisant sur v_n que l'hypothèse $\|v_n\| = o(a_n)$ ([18], [2]). Ceci élargit dans le cas $q = 1$ le choix de θ_n ; par exemple, dans $\mathbb{R}$, $\theta_n = \lambda n^\beta+\rho$, avec $\beta > 0$ [7].
Mais l'étude faite ici est motivée par le cas $q > 1$.

3. On peut étudier la convergence du processus plus général :

$$X_n = f_n(X_{n-1}) - a_n A_n(f_n(X_{n-1}), Z_n) + \varepsilon_n.$$

ε_n est une variable aléatoire dans $\mathbb{R}^k$. Des conditions sur ε_n suffisantes pour assurer la convergence presque sûre de $X_n - \theta_n$ vers 0 sont 6 ou 6' [3]. (On pose $V_n = f_n(X_{n-1})$) .

6i. $\sum_1^\infty E[\| E[\varepsilon_n | T_n]\|^2]^{\frac{1}{2}} < \infty.$

ii. $\sum_1^\infty E[\|\varepsilon_n\|^2] < \infty.$

6'i. $\forall$ n, $\exists$ g_n, $h_n \in \mathbb{R}^+ : \| E[\varepsilon_n \mid T_n] \| \leq g_n \| V_n - \theta_n \| + h_n$ p.s.

$\sum_1^\infty g_n < \infty, \sum_1^\infty h_n < \infty.$

ii. $\forall$ n, $\exists$ p_n, $q_n \in \mathbb{R}^+ : E[\| \varepsilon_n \|^2 \mid T_n] \leq p_n \| V_n - \theta_n \|^2 + q_n$ p.s.

$\sum_1^\infty p_n < \infty, \sum_1^\infty q_n < \infty.$

Seules les parties a et b de la démonstration ci-dessous sont modifiées lorsque l'on fait ces hypothèses.

4. L'extension de ce théorème au cas du processus sous contraintes convexes est immédiate.

Soit K un convexe fermé non vide de $\mathbb{R}^k$. On suppose que, pour tout n, la fonction f_n est une application de K dans K et θ_n appartient à K. Soit P l'opérateur de projection sur K. On considère le processus :

$$X_n = P(f_n(X_{n-1}) - a_n A_n(f_n(X_{n-1}), Z_n)).$$

Sous les hypothèses du théorème 1, énoncées pour x appartenant à K, on a : $X_n - \theta_n \xrightarrow{p.s.} 0$. Pour s'en assurer, il suffit de reprendre la démonstration du théorème et d'utiliser dans les parties a et e la propriété de contraction de la projection.

Démonstration :

a) Sous l'hypothèse 1i, on a :

$$\begin{aligned}
\| f_{n+1}(x) - \theta_{n+1} \|^2 &= \| f_{n+1}(x) - f_{n+1}(\theta_n) - v_{n+1} \|^2 \\
&\leq \| f_{n+1}(x) - f_{n+1}(\theta_n) \|^2 (1 + 2\| v_{n+1} \|) + 2 \| v_{n+1} \| + \| v_{n+1} \|^2 \\
&\leq ((1 + \gamma_n) \| x - \theta_n \|^2 + \delta_n)(1 + 2\| v_{n+1} \|) + 2 \| v_{n+1} \| + \| v_{n+1} \|^2 \\
&\leq (1 + \nu_n) \| x - \theta_n \|^2 + \alpha_n
\end{aligned}$$

avec $1 + \nu_n = (1 + \gamma_n)(1 + 2 \| v_{n+1} \|)$

$\alpha_n = \delta_n (1 + 2 \| v_{n+1} \|) + 2 \| v_{n+1} \| + \| v_{n+1} \|^2.$

On étudie le processus :

$$V_n = f_n(X_{n-1})$$
$$X_n = V_n - a_n A_n(V_n, Z_n).$$

$$\| V_{n+1} - \theta_{n+1} \|^2 = \| f_{n+1}(X_n) - \theta_{n+1} \|^2 \leq (1+\nu_n) \| X_n - \theta_n \|^2 + \alpha_n$$
$$\leq (1+\nu_n) \| V_n - \theta_n - a_n A_n(V_n, Z_n) \|^2 + \alpha_n$$
$$\leq (1+\nu_n) \| V_n - \theta_n \|^2 - 2(1+\nu_n) \langle V_n - \theta_n, a_n A_n(V_n, Z_n) \rangle$$
$$+ (1+\nu_n) a_n^2 \| A_n(V_n, Z_n) \|^2 + \alpha_n.$$

On passe à l'espérance conditionnelle par rapport à T_n ; en utilisant la propriété de Robbins-Monro, on obtient :

$$E[\| V_{n+1} - \theta_{n+1} \|^2 | T_n] \leq (1 + \nu_n) \| V_n - \theta_n \|^2$$
$$+ (1 + \nu_n) a_n^2 E[\| A_n(V_n, Z_n) \|^2 | T_n] + \alpha_n - 2(1 + \nu_n) \langle V_n - \theta_n, a_n M_n(V_n) \rangle \quad \text{p. s.}$$

Sous l'hypothèse 2, on a :

$$E[\| A_n(V_n, Z_n) \|^2 \mid T_n] \leq d \| V_n - \theta_n \|^2 + e \quad \text{p.s.}$$

On en déduit :

$$E[\| V_{n+1} - \theta_{n+1} \|^2 | T_n] \leq (1 + \lambda_n) \| V_n - \theta_n \|^2 + \mu_n$$
$$- 2(1 + \nu_n) \langle V_n - \theta_n, a_n M_n(V_n) \rangle \quad \text{p.s.} \quad (2)$$

avec
$$1 + \lambda_n = (1 + \nu_n)(1 + d a_n^2)$$
$$\mu_n = \alpha_n + e(1 + \nu_n) a_n^2 .$$

Sous les hypothèses 1ii, 1iv, 4ii, on a $\sum_1^\infty \lambda_n < \infty, \sum_1^\infty \mu_n < \infty, \sum_1^\infty \nu_n < \infty$. Sous l'hypothèse 3i, on peut alors appliquer un lemme de la théorie des martingales, démontré par Robbins et Siegmund [15] :

Lemme :

Soit $(\Omega, \mathfrak{A}, P)$ un espace probabilisé, (T_n) une suite croissante de sous-tribus de $\mathfrak{A}$. Soit, pour tout n, A_n, B_n, C_n, D_n des variables aléatoires réelles T_n-mesurables, non négatives, intégrables telles que :

$$\forall n, \quad E[A_{n+1} | T_n] \leq A_n(1 + B_n) + C_n - D_n \quad \text{p. s.} ;$$
$$\sum_1^\infty B_n < \infty, \sum_1^\infty C_n < \infty \quad \text{p. s.}$$

Alors, la suite (A_n) converge presque sûrement vers une variable aléatoire réelle A finie et on a $\sum_1^\infty D_n < \infty$ p. s. ■

On en déduit que :

$$\exists\, T \text{ v. a. dans } \mathbb{R}^+ : \| V_n - \theta_n \| \xrightarrow{\text{p.s.}} T \ ;$$

$$\sum_1^\infty \langle V_n - \theta_n, a_n M_n(V_n) \rangle < \infty \quad \text{p. s.}$$

b) A partir de (2), en utlisant l'hypothèse 3i et en passant à l'espérance, on obtient :

$$E[\| V_{n+1} - \theta_{n+1} \|^2] \leq (1 + \lambda_n)\ E[\| V_n - \theta_n \|^2] + \mu_n.$$

En appliquant alors le lemme de Robbins et Siegmund, on conclut que :

$$\exists\, t \geq 0 : E[\| V_n - \theta_n \|^2] \to t.$$

c) Soit ω appartenant à l'intersection des ensembles de convergence définis. On suppose $T(\omega) \neq 0$. Alors :

$$\exists\, 0 < \varepsilon_1 < 1, \exists\, N(\omega, \varepsilon_1) : \forall\, n > N(\omega, \varepsilon_1),\ \varepsilon_1 < \| V_n(\omega) - \theta_n \| < \frac{1}{\varepsilon_1} \quad (3)$$

Sous l'hypothèse 3ii, on a alors :

$$\exists\, L(\omega, \varepsilon_1) : \forall\, \ell > L(\omega, \varepsilon_1),\ \sum_{j \in I_\ell} \langle V_{n_\ell}(\omega) - \theta_j, M_j(V_{n_\ell}(\omega)) \rangle \geq b(\varepsilon_1).$$

Par conséquent :

$$\exists\, m_\ell(\omega) \in I_\ell \ : \langle V_{n_\ell}(\omega) - \theta_{m_\ell(\omega)}, M_{m_\ell(\omega)}(V_{n_\ell}(\omega)) \rangle \geq \frac{b(\varepsilon_1)}{q}. \quad (4)$$

d) Pour $q = 1$, on a : $\forall\, \ell,\ m_\ell(\omega) = n_\ell$.

On en déduit, sous les hypothèses 3i et 4i :

$$\sum_1^\infty a_n \langle V_n(\omega) - \theta_n, M_n(V_n(\omega)) \rangle = +\infty.$$

Ceci est en contradiction avec la deuxième conclusion de la partie a. Donc, $T(\omega) = 0$.

e) Pour étudier le cas $q > 1$, on montre d'abord que

$\|V_{n+1}(\omega) - V_n(\omega)\| \to 0 \quad (n \to \infty)$.

On a :

$$V_{n+1} = f_{n+1}(X_n)$$
$$X_n = V_n - a_n A_n(V_n, Z_n).$$

Sous l'hypothèse 2 :

$$E\left[\sum_1^\infty a_n^2 \, \|A_n(V_n, Z_n)\|^2\right] \leq \sum_1^\infty a_n^2 (d\, E[\|V_n - \theta_n\|^2] + e).$$

Comme $E[\|V_n - \theta_n\|^2] \to t$ et $\sum_1^\infty a_n^2 < \infty$, on a :

$E\left[\sum_1^\infty a_n^2 \, \|A_n(V_n, Z_n)\|^2\right] < \infty$, ce qui implique : $a_n A_n(V_n, Z_n) \overset{p.s.}{\to} 0$.

Donc :

$$X_n - V_n \overset{p.s.}{\to} 0.$$

De (3), on déduit alors que :

$$\exists\, 0 < \varepsilon_2 < 1, \exists\, N_2(\omega, \varepsilon_2) : \forall\, n > N_2(\omega, \varepsilon_2) \;,\; \varepsilon_2 < \|X_n(\omega) - \theta_n\| < \frac{1}{\varepsilon_2}.$$

Sous l'hypothèse 1iii, ceci implique :

$$f_{n+1}(X_n(\omega)) - X_n(\omega) \to 0.$$

Par conséquent :

$$V_{n+1}(\omega) - V_n(\omega) = (f_{n+1}(X_n(\omega)) - X_n(\omega)) + (X_n(\omega) - V_n(\omega)) \to 0.$$

f) On supprime désormais l'écriture de ω. On considère la décomposition :

$$<V_{m_\ell} - \theta_{m_\ell}, M_{m_\ell}(V_{m_\ell})> = <V_{m_\ell} - V_{n_\ell}, M_{m_\ell}(V_{m_\ell})> + <V_{n_\ell} - \theta_{m_\ell}, M_{m_\ell}(V_{n_\ell})>$$
$$+ <V_{n_\ell} - \theta_{m_\ell}, M_{m_\ell}(V_{m_\ell}) - M_{m_\ell}(V_{n_\ell})>.$$

Soit $\varepsilon > 0$. Sous l'hypothèse 3iii, à ε correspond un η. Soit $\varepsilon' < \eta$. Puisque $\| V_{n+1} - V_n \| \to 0$ et $0 \leq m_\ell - n_\ell < q$, on a, à partir d'un certain rang :

$$\| V_{m_\ell} - V_{n_\ell} \| < \varepsilon' < \eta.$$

Sous l'hypothèse 3iii, ceci implique :

$$\| M_{m_\ell}(V_{m_\ell}) - M_{m_\ell}(V_{n_\ell}) \| < \varepsilon.$$

On en déduit :

$$\begin{aligned} |<V_{n_\ell} - \theta_{m_\ell}, M_{m_\ell}(V_{m_\ell}) - M_{m_\ell}(V_{n_\ell})>| &\leq \varepsilon \| V_{n_\ell} - \theta_{m_\ell} \| \\ &\leq \varepsilon (\| V_{n_\ell} - V_{m_\ell} \| + \| V_{m_\ell} - \theta_{m_\ell} \|) < \varepsilon (\varepsilon' + \frac{1}{\varepsilon_1}). \end{aligned}$$

En outre :

$$|<V_{m_\ell} - V_{n_\ell}, M_{m_\ell}(V_{m_\ell})>| \leq \varepsilon' \| M_{m_\ell}(V_{m_\ell}) \|.$$

Or :

$$\| M_n(x) \| = \| E[A_n(x, Y_n(x))] \| \leq E[\| A_n(x, Y_n(x)) \|^2]^{\frac{1}{2}} \leq (d \| x - \theta_n \|^2 + e)^{\frac{1}{2}}.$$

Donc :

$$\| M_{m_\ell}(V_{m_\ell}) \| \leq (d \| V_{m_\ell} - \theta_{m_\ell} \|^2 + e)^{\frac{1}{2}} < (\frac{d}{\varepsilon_1^2} + e)^{\frac{1}{2}}$$

et :

$$|<V_{m_\ell} - V_{n_\ell}, M_{m_\ell}(V_{m_\ell})>| < \varepsilon' (\frac{d}{\varepsilon_1^2} + e)^{\frac{1}{2}}.$$

De (4), on déduit alors que :

$$<V_{m_\ell} - \theta_{m_\ell}, M_{m_\ell}(V_{m_\ell})> > \frac{b(\varepsilon_1)}{q} - \varepsilon (\varepsilon' + \frac{1}{\varepsilon_1}) - \varepsilon' (\frac{d}{\varepsilon_1^2} + e)^{\frac{1}{2}}.$$

On peut choisir ε et ε' suffisamment petits pour que :

$$<V_{m_\ell} - \theta_{m_\ell}, M_{m_\ell}(V_{m_\ell})> > \frac{b(\varepsilon_1)}{2q}.$$

Sous les hypothèses 3i et 4i, on a alors :

$$\sum_n a_n <V_n - \theta_n, M_n(V_n)> = +\infty.$$

Ceci est en contradiction avec la deuxième conclusion de la partie a. Donc : $T(\omega) = 0$; $V_n - \theta_n \xrightarrow{p.s.} 0$. Comme $X_n - V_n \xrightarrow{p.s.} 0$, on a : $X_n - \theta_n \xrightarrow{p.s.} 0$.

2.3 - Généralisation du théorème

On peut généraliser les hypothèses du théorème 1 en remplaçant, pour tout n, la fonction $||x - \theta_n||^2$ par une fonction positive $F_n(x)$; on montre alors que $F_n(X_n) \xrightarrow{p.s.} 0$. Ce type de généralisation a été introduit par Blum [5] et étudié en particulier par Nevel'son et Has'minskii [13]. Ceci permet également de traiter le cas où l'équation $E[Y_n(x)] = 0$ admet plus d'une solution, θ_n étant remplacé, pour tout n, par un ensemble B_n de $\mathbb{R}^k$; on peut alors montrer que $d(X_n, B_n) \xrightarrow{p.s.} 0$, $d(x, B)$ désignant la distance de x à B.

Théorème 2 :

Soit $(F_n, n \geq 1)$ une suite de fonctions réelles définies dans $\mathbb{R}^k$, à valeurs positives ou nulles, continûment différentiables jusqu'à l'ordre deux ; soit, pour tout n, G_n le gradient de F_n, H_n le hessien de F_n.

D'après la formule de Taylor, on a :

$$F_n(x - a_n A_n(x, Y_n(x))) = F_n(x) - a_n <G_n(x), A_n(x, Y_n(x))> + a_n^2 K_n(x)$$

avec

$$K_n(x) = \frac{1}{2} <A_n(x, Y_n(x)), H_n(x - \mu_n(x) a_n A_n(x, Y_n(x))) A_n(x, Y_n(x))>, 0 < \mu_n(x) < 1.$$

On fait les hypothèses :

1i. $\forall\, n, \exists\, \gamma_n,\ \delta_n \in \mathbb{R}^k : \forall\, x,\ F_{n+1} \circ f_{n+1}(x) \le (1 + \gamma_n)\, F_n(x) + \delta_n.$

ii. $\sum_1^\infty \gamma_n < \infty\, ;\ \sum_1^\infty \delta_n < \infty.$

iii. $\forall\, 0 < \varepsilon < 1,\ \lim_{n\to\infty} \sup_{\{\varepsilon < F_n(x) < \frac{1}{\varepsilon}\}} \| f_{n+1}(x) - x \| = 0.$

iv. $\exists\, \alpha > 0, \forall\, \varepsilon > 0, \exists\, \eta > 0 : (\| x_1 - x_2 \| < \eta) \Rightarrow (\sup_n | F_n^\alpha(x_1) - F_n^\alpha(x_2) | < \varepsilon).$

2. $\exists\, d, e \in \mathbb{R}^k : \forall\, n, \forall\, x,\ E[K_n(x)] \le dF_n(x) + e.$

3i. $\forall\, n, \forall\, x,\ \langle G_n(x), M_n(x)\rangle \ge 0.$

ii. $\exists\, q \in \mathbb{N}^*, \exists\, (n_\ell,\ \ell \ge 1) \subset \mathbb{N} : \forall\, \ell,\ n_{\ell+1} \ge n_\ell + q,$

$\forall\, 0 < \varepsilon < 1,\ \exists\, L(\varepsilon) \in \mathbb{N} : b(\varepsilon) > 0$ avec

$$b(\varepsilon) = \inf_{\ell > L(\varepsilon)} \inf_{\{\varepsilon < F_{n_\ell}(x) < \frac{1}{\varepsilon}\}} \sum_{j \in I_\ell} \langle G_j(x), M_j(x)\rangle, \quad I_\ell = \{n_\ell, n_\ell + 1, \ldots, n_\ell + q - 1\}.$$

iii. $\forall\, \varepsilon > 0, \exists\, \eta > 0 : (\| x_1 - x_2 \| < \eta) \Rightarrow (\sup_n \| M_n(x_1) - M_n(x_2) \| < \varepsilon),$

$(\sup_n \| G_n(x_1) - G_n(x_2) \| < \varepsilon).$

iv. $\forall\, 0 < \varepsilon < 1,\quad G(\varepsilon) = \sup_n \sup_{\{\varepsilon < F_n(x) < \frac{1}{\varepsilon}\}} \| G_n(x) \| < \infty\, ;$

$\forall\, 0 < \varepsilon < 1,\quad V(\varepsilon) = \sup_n \sup_{\{\varepsilon < F_n(x) < \frac{1}{\varepsilon}\}} \| M_n(x) \| < \infty.$

4i. $\forall\, n,\ a_n > 0,\quad \sum_1^\infty \min_{j \in I_\ell} a_j = +\infty.$

ii. $\sum_1^\infty a_n^2 < \infty.$

5. $E[F_1(X_o)] < \infty.$

Sous les hypothèses 1, 2, 3, 4, 5, on a, dans l'ensemble

$\{a_n \| A_n(V_n, Z_n) \| \to 0\},\quad F_n(V_n) \xrightarrow{\text{p.s.}} 0.$

Pour $q = 1$, on a $F_n(V_n) \xrightarrow{\text{p.s.}} 0$ sous les hypothèses 1i, 1ii, 2, 3i, 3ii, 4, 5. ∎

Démonstration résumée :

On suit le plan de la démonstration du théorème 1.

a) Sous l'hypothèse 1i, on a :

$$F_{n+1}(V_{n+1}) \leq (1 + \gamma_n)\, F_n(X_n) + \delta_n.$$

On développe $F_n(X_n)$ par la formule de Taylor. On a alors sous l'hypothèse 2 :

$$E[F_{n+1}(V_{n+1}) \mid T_n] \leq (1 + \gamma_n)\,(1 + d\, a_n^2)\, F_n(V_n) + e\,(1 + \gamma_n)\, a_n^2 + \delta_n - (1 + \gamma_n)\, a_n < G_n(V_n), M_n(V_n)> \quad \text{p. s.}$$

On applique le lemme de Robbins et Siegmund : $\exists\, T$ v. a. dans $\mathbb{R}^+$:

$$F_n(V_n) \xrightarrow{\text{p.s.}} T \; ; \sum_{1}^{\infty} a_n < G_n(V_n), M_n(V_n)> < \infty \quad \text{p. s.}$$

b) On suppose $T(\omega) \neq 0$. On a alors sous l'hypothèse 3ii :

$$\exists\, 0 < \varepsilon_1 < 1, \exists\, L(\omega, \varepsilon_1) : \forall \ell > L(\omega, \varepsilon_1), \exists\, m_\ell(\omega) \in I_\ell :$$
$$< G_{m_\ell(\omega)}(V_{n_\ell}(\omega)), M_{m_\ell(\omega)}(V_{n_\ell}(\omega))> \geq \frac{b(\varepsilon_1)}{q}.$$

c) Pour $q = 1$, $m_\ell(\omega) = n_\ell$. On est alors en contradiction avec la deuxième conclusion de la partie a. Donc $T(\omega) = 0$.

d) Pour $q > 1$, on montre dans cette partie que $V_{n+1}(\omega) - V_n(\omega) \to 0$ lorsque ω appartient à l'ensemble $\{a_n \| A_n(V_n, Z_n) \| \to 0\}$.

Pour cela, on écrit que :

$$\| V_{n+1}(\omega) - V_n(\omega) \| \leq \| f_{n+1}(X_n(\omega)) - X_n(\omega) \| + \| V_n(\omega) - X_n(\omega) \|$$

et on utilise les hypothèses 1iii et 1iv.

e) On considère alors la décomposition :

$$< G_{m_\ell}(V_{m_\ell}), M_{m_\ell}(V_{m_\ell})> = < G_{m_\ell}(V_{m_\ell}) - G_{m_\ell}(V_{n_\ell}), M_{m_\ell}(V_{m_\ell})>$$
$$+ < G_{m_\ell}(V_{n_\ell}), M_{m_\ell}(V_{n_\ell})> + < G_{m_\ell}(V_{n_\ell}), M_{m_\ell}(V_{m_\ell}) - M_{m_\ell}(V_{n_\ell})>.$$

Sous les hypothèses 3iii et 3iv, comme $V_{m_\ell} - V_{n_\ell} \to 0$, on a à partir d'un certain rang :

$$<G_{m_\ell}(V_{m_\ell}), M_{m_\ell}(V_{m_\ell})> > \frac{b(\varepsilon_1)}{2q}.$$

On est alors en contradiction avec la deuxième conclusion de la partie a. Donc $T(\omega) = 0$.

3 - APPLICATION A L'ETUDE D'UN DEUXIEME MODELE D'APPROXIMATION STOCHASTIQUE DYNAMIQUE

3.1 - Le modèle

Soit la famille de variables aléatoires observables dans $\mathbb{R}^p$, $\{Y_n(x), n \in \mathbb{N}^*, x \in \mathbb{R}^p\}$.

On suppose l'existence, pour tout n et tout x, de $E[Y_n(x)] = M_n(x)$. On suppose que, pour tout n, l'équation $M_n(x) = 0$ admet la solution θ_n, de composantes $\theta_n^1, \theta_n^2, \dots, \theta_n^p$.

Soit p entiers $k_1, k_2, \dots, k_p$.

Pour $i = 1, 2, \dots, p$, pour tout n, on suppose l'existence de deux éléments de $\mathbb{R}^{k_i}$, β_n^i, inconnu, et U_n^i, connu au temps n, tels que $\theta_n^i = <\beta_n^i, U_n^i>$, $<\,,\,>$ désignant le produit scalaire euclidien usuel dans $\mathbb{R}^{k_i}$, et, pour $n \geq 2$, l'existence d'une application f_n^i de $\mathbb{R}^{k_i}$ dans $\mathbb{R}^{k_i}$, connue au temps n, et d'un élément v_n^i de $\mathbb{R}^{k_i}$, inconnu, tels que $\beta_n^i = f_n^i(\beta_{n-1}^i) + v_n^i$.

On note $Y_n^1(x), Y_n^2(x), \dots, Y_n^p(x)$ les variables aléatoires réelles marginales de $Y_n(x)$.

Soit (a_n) une suite de nombres réels positifs.

Pour $i = 1, 2, \dots, p$, on définit le processus (B_n^i) dans $\mathbb{R}^{k_i}$ par :

$$B_n^i = f_n^i(B_{n-1}^i) - a_n U_n^i Z_n^i$$

où pour tout n, Z_n^i est une variable aléatoire réelle dont la loi de probabilité conditionnelle par rapport à la tribu du passé est la loi de

$$Y_n^i(<f_n^1(b_{n-1}^1), U_n^1>, \dots, <f_n^p(b_{n-1}^p), U_n^p>)$$

$b^1_{n-1}, \ldots, b^p_{n-1}$ étant les réalisations respectives de $B^1_{n-1}, \ldots, B^p_{n-1}$. (On pose par convention : $\forall$ i, $f^i_1(b^i) = b^i$; $v^i_1 = 0$; on choisit au début $B^1_o, \ldots, B^p_o$).

$$\text{On a } B_n = \begin{pmatrix} B^1_n \\ B^2_n \\ \cdot \\ \cdot \\ \cdot \\ B^p_n \end{pmatrix} = \underbrace{\begin{pmatrix} f^1_n(B^1_{n-1}) \\ f^2_n(B^2_{n-1}) \\ \cdot \\ \cdot \\ \cdot \\ f^p_n(B^p_{n-1}) \end{pmatrix}}_{f_n(B_{n-1})} - a_n \underbrace{\begin{pmatrix} U^1_n & & & & \\ & U^2_n & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & \cdot \\ & & & & & U^p_n \end{pmatrix}}_{A^1_n} \underbrace{\begin{pmatrix} Z^1_n \\ Z^2_n \\ \cdot \\ \cdot \\ \cdot \\ Z^p_n \end{pmatrix}}_{Z_n}$$

$$B_n = f_n(B_{n-1}) - a_n A^1_n Z_n.$$

B_n est un vecteur aléatoire dans $\mathbb{R}^k$, avec $k = k_1 + k_2 + \ldots + k_p$. f_n est une application de $\mathbb{R}^k$ dans $\mathbb{R}^k$. A^1_n est une matrice $k \times p$. Z_n est un vecteur aléatoire dans $\mathbb{R}^k$.

On définit, pour tout n, le vecteur β_n dans $\mathbb{R}^k$: $\beta_n = (\beta^1_n, \beta^2_n, \ldots, \beta^p_n)'$.

On définit, pour tout n, pour $i = 1, 2, \ldots, p$, la variable aléatoire réelle $X^i_n = \langle B^i_n, U^i_n \rangle$, et, pour tout n, la variable aléatoire X_n dans $\mathbb{R}^p$ de composantes $X^1_n, X^2_n, \ldots, X^p_n$.

$\mathbb{R}^k$ et $\mathbb{R}^p$ sont munis du produit scalaire euclidien usuel.

3.2 - Théorème de convergence presque sûre

Théorème 3 :

On fait les hypothèses :

1i. $\forall$ n, $\exists\, \gamma_n, \delta_n \in \mathbb{R}^+$:

$\forall$ i, $\forall\, b^i \in \mathbb{R}^{k_i}$, $\| f^i_{n+1}(b^i) - f^i_{n+1}(\beta^i_n) \|^2 \leq (1 + \gamma_n) \| b^i - \beta^i_n \|^2 + \delta_n$.

ii. $\sum_1^\infty \gamma_n < \infty$; $\sum_1^\infty \delta_n < \infty$.

iii. $\forall\, i, \forall\, A > 0,\ \lim\limits_{n\to\infty}\ \sup\limits_{\{\|b-\beta_n\| < A\}} \|f^i_{n+1}(b^i) - b^i\| = 0.$

iv. $\forall\, i,\ \sum\limits_1^{\infty} \|v^i_n\| < \infty.$

v. $U = \sup\limits_{i,n} \|U^i_n\| < \infty.$

vi. $\exists\, q \in \mathbb{N}^*, \exists\, (n_\ell, \ell \geq 1) \subset \mathbb{N} : \forall \ell, n_{\ell+1} \geq n_\ell + q, \delta > 0,$ avec

$$\delta = \inf_{i,\ell} \lambda_{\min} \Big(\sum_{j\in I_\ell} U^i_j U^{i'}_j\Big),$$

$$I_\ell = \{n_\ell, n_\ell + 1, \dots, n_\ell + q - 1\}.$$

2. $\exists\, d, e \in \mathbb{R}^+ : \forall\, n, \forall\, x, E[\|Y_n(x)\|^2] \leq d\|x - \theta_n\|^2 + e.$

3i. $\forall\, n, \forall\, x, \langle x - \theta_n, M_n(x)\rangle \geq 0.$

ii. $\forall\, 0 < \varepsilon < 1,\ c(\varepsilon) > 0,$ avec

$$c(\varepsilon) = \inf_{\ell}\ \min_{j\in I_\ell}\ \inf_{\{\varepsilon < \|x-\theta_j\| < \frac{1}{\varepsilon}\}} \langle x - \theta_j, M_j(x)\rangle.$$

iii. $\forall\, \varepsilon > 0, \exists\, \eta > 0 : (\|x_1 - x_2\| < \eta) \Rightarrow (\sup\limits_n \|M_n(x_1) - M_n(x_2)\| < \varepsilon).$

4i. $\forall\, n, a_n > 0,\ \sum\limits_1^{\infty} \min\limits_{j\in I_\ell} a_j = +\infty.$

ii. $\sum\limits_1^{\infty} a_n^2 < \infty.$

5. $E[\|B_o\|^2] < \infty.$

Sous les hypothèses 1, 2, 3, 4, 5, on a :

$$B_n - \beta_n \xrightarrow{p.s.} 0\ ;\ X_n - \theta_n \xrightarrow{p.s.} 0.\ \blacksquare$$

Démonstration :

Soit, pour $i = 1, 2, \dots, p,\ b^i \in \mathbb{R}^{k_i}$.

Soit $b = (b^1, b^2, \dots, b^p)' \in \mathbb{R}^k$.

$\mathbb{R}^{k_1}, \mathbb{R}^{k_2}, \dots \mathbb{R}^{k_p}, \mathbb{R}^k, \mathbb{R}^p$ sont munis du produit scalaire euclidien usuel.

Soit $Y_n(\langle b^1, U^1_n\rangle, \dots, \langle b^p, U^p_n\rangle) = Y_{1n}(b).$

On considère la famille de variables aléatoires dans $\mathbb{R}^p$:

$$\{Y_{1n}(b),\ n \in \mathbb{N},\ b \in \mathbb{R}^k\}$$

et le processus (B_n) dans $\mathbb{R}^k$:

$$B_n = f_n(B_{n-1}) - a_n A_n^1 Z_n$$

tel que la loi de probabilité conditionnelle de Z_n par rapport à la tribu du passé soit la loi de $Y_{1n}(f_n(b_{n-1}))$, avec $f_n(b) = (f_n^1(b^1), \ldots, f_n^p(b^p))$.

On vérifie alors pour cette famille et ce processus les hypothèses du théorème 1.

Hypothèse 1i :

$\forall\, n, \exists\, \gamma_{1n}, \delta_{1n} \in \mathbb{R}^+ : \forall\, b, \| f_{n+1}(b) - f_{n+1}(\beta_n) \|^2 \leq (1 + \gamma_{1n}) \| b - \beta_n \|^2 + \delta_{1n}.$

Sous l'hypothèse 1i du théorème 3, on a :

$$\| f_{n+1}(b) - f_{n+1}(\beta_n) \|^2 = \sum_{i=1}^{p} \| f_{n+1}^i(b^i) - f_{n+1}^i(\beta_n^i) \|^2$$

$$\leq (1 + \gamma_n) \sum_{i=1}^{p} \| b^i - \beta_n^i \|^2 + p\, \delta_n \leq (1 + \gamma_n) \| b - \beta_n \|^2 + p\, \delta_n.$$

L'hypothèse 1i du théorème 1 est vérifiée avec $\gamma_{1n} = \gamma_n$, $\delta_{1n} = p\, \delta_n$.

Hypothèses 1ii, 1iii, 1iv :

$$\sum_1^\infty \gamma_{1n} < \infty\,;\ \sum_1^\infty \delta_{1n} < \infty\,;$$

$$\forall\, 0 < \varepsilon < 1,\ \lim_{n \to \infty} \sup_{\{\varepsilon < \| b - \beta_n \| < \frac{1}{\varepsilon}\}} \| f_{n+1}(b) - b \| = 0\,;$$

$$\sum_1^\infty \| v_n \| < \infty.$$

Sous les hypothèses 1ii, 1iii, 1iv du théorème 3, la vérification de ces hypothèses est immédiate.

Hypothèse 2 :

$$\exists\, d_1, e_1 \in \mathbb{R}^{k+} : \forall\, n, \forall\, b,\ E[\| A_n^1\, Y_{1n}(b) \|^2] \leq d_1\ \| b-\beta_n \|^2 + e_1.$$

Sous les hypothèses 1v et 2 du théorème 3, on a :

$$\begin{aligned} E[\| A_n^1 Y_{1n}\,(b) \|^2] &= E[\sum_{i=1}^{p} \| U_n^i \|^2 (Y_{1n}^i(b))^2] \\ &\leq U^2\, E[\sum_{i=1}^{p} (Y_{1n}^i(b))^2] = U^2\, E[\| Y_{1n}(b) \|^2] \\ &\leq U^2\, E[\| Y_n\,(x_n) \|^2] \end{aligned}$$

(avec $x_n = (<b^1, U_n^1>, \ldots, <b^p, U_n^p>)) \leq U^2\,(d \| x_n - \theta_n \|^2 + e)$.

Or :

$$E[\| A_n^1\, Y_{1n}(b) \|^2] = E[\sum_{i=1}^{p} \| U_n^i \|^2 (Y_{1n}^i(b))^2]$$

Donc :

$$E[\| A_n^1\, Y_{1n}(b) \|^2] \leq d\, U^4 \| b - \beta_n \|^2 + e\, U^2.$$

L'hypothèse 2 du théorème 1 est vérifiée avec $d_1 = d\,U^4$, $e_1 = e\,U^2$.

Hypothèse 3i :

$\forall\, n, \forall\, b,\ <b - \beta_n, A_n^1\, M_{1n}(b)> \geq 0$ avec $M_{1n}(b) = E[Y_{1n}(b)]$.

On a : $M_{1n}(b) = M_n(x_n)$.

On note $M_{1n}^1, M_{1n}^2, \ldots, M_{1n}^p$ (respectivement $M_n^1, M_n^2, \ldots, M_n^p$) les composantes de M_{1n} (respectivement M_n).

Sous l'hypothèse 3i du théorème 3, on a :

$$\begin{aligned} <b - \beta_n, A_n^1\, M_{1n}(b)> &= \sum_{i=1}^{p} <b^i - \beta_n^i, U_n^i\, M_{1n}^i(b)> = \sum_{i=1}^{p} <b^i - \beta_n^i, U_n^i> M_{1n}^i(b) \\ &= \sum_{i=1}^{p} (x_n^i - \theta_n^i)\, M_n^i(x_n) = <x_n - \theta_n, M_n(x_n)> \geq 0. \end{aligned}$$

L'hypothèse 3i du théorème 1 est vérifiée.

Hypothèse 3ii :

$\exists\, q \in \mathbb{N}^*, \exists\, (n_\ell\,,\ \ell \geq 1) \subset \mathbb{N} : \forall \ell\,,\ n_{\ell+1} \geq n_\ell + q,$
$\forall\, 0 < \varepsilon < 1,\ \ \exists\, L(\varepsilon) \in \mathbb{N}\ \ : b(\varepsilon) > 0$, avec

$$b(\varepsilon) = \inf_{\ell > L(\varepsilon)} \ \inf_{\{\varepsilon < \| b - \beta_{n_\ell} \| < \frac{1}{\varepsilon}\}} \ \sum_{j \in I_\ell} \langle b - \beta_j, A_j^1\, M_{1j}(b) \rangle,$$

$I_\ell = \{n_\ell, n_\ell + 1, \dots, n_\ell + q - 1\}.$

On a :

$$b(\varepsilon) = \inf_{\ell > L(\varepsilon)} \ \inf_{\{\varepsilon < \| b - \beta_{n_\ell} \| < \frac{1}{\varepsilon}\}} \ \sum_{j \in I_\ell} \langle x_j - \theta_j, M_j(x_j) \rangle.$$

On va montrer que, sous les hypothèses 1iii, 1iv, 1v et 1vi, on a :

$\forall\, 0 < \varepsilon < 1,\ \ \exists\, L(\varepsilon) :\ \ \forall \ell > L(\varepsilon),$

$$(\varepsilon < \| b - \beta_{n_\ell} \| < \frac{1}{\varepsilon}) \ \Rightarrow\ (\exists\, j \in I_\ell, \exists\, \varepsilon_2 \colon \varepsilon_2 < \| x_j - \theta_j \| < \frac{1}{\varepsilon_2}).$$

Comme $\| b - \beta_{n_\ell} \|^2 = \sum_{i=1}^{p} \| b^i - \beta^i_{n_\ell} \|^2$, on a :

$$(\| b - \beta_{n_\ell} \| > \varepsilon) \ \Rightarrow\ (\exists\, i \in \{1, 2, \dots, p\} : \| b^i - \beta^i_{n_\ell} \|^2 > \frac{\varepsilon^2}{p}). \quad (1)$$

Pour $j \in I_\ell$, pour $i = 1, 2, \dots, p$, on a :

$$x^i_j - \theta^i_j = \langle b^i - \beta^i_{n_\ell}, U^i_j \rangle + \langle \beta^i_{n_\ell} - \beta^i_j, U^i_j \rangle.$$

$$\| x^i_j - \theta^i_j | - | \langle b^i - \beta^i_{n_\ell}, U^i_j \rangle \| \ \leq\ | \langle \beta^i_{n_\ell} - \beta^i_j, U^i_j \rangle | \ \leq\ U \| \beta^i_{n_\ell} - \beta^i_j \|.$$

$$\| \beta^i_{n_\ell} - \beta^i_j \| \ \leq \| \beta^i_j - \beta^i_{j-1} \| + \dots + \| \beta^i_{n_\ell+1} - \beta^i_{n_\ell} \|$$

$$\leq \| f^i_j(\beta^i_{j-1}) - \beta^i_{j-1} \| + \dots + \| f^i_{n_\ell+1}(\beta^i_{n_\ell}) - \beta^i_{n_\ell} \| + \| v^i_j \| + \dots + \| v^i_{n_\ell+1} \|.$$

Soit : $\varepsilon_1 = \varepsilon \sqrt{\frac{\delta}{pq}}$. Sous les hypothèses 1iii et 1iv, à partir d'un certain rang $L(\varepsilon)$, on a :

$$\| \beta^i_{n_\ell} - \beta^i_j \| \leq \frac{\varepsilon_1}{2U}.$$

$$\| \, |x^i_j - \theta^i_j| - |<b^i - \beta^i_{n_\ell}, U^i_j>| \, \| \leq \frac{\varepsilon_1}{2}.$$

Donc :

$$|<b^i - \beta^i_{n_\ell}, U^i_j>| - \frac{\varepsilon_1}{2} \leq |x^i_j - \theta^i_j| \leq |<b^i - \beta^i_{n_\ell}, U^i_j>| + \frac{\varepsilon_1}{2}. \quad (2)$$

De (2), on déduit d'une part que :

$$\| x_j - \theta_j \|^2 = \sum_{i=1}^{p} (x^i_j - \theta^i_j)^2 \leq 2 \sum_{i=1}^{p} (<b^i - \beta^i_{n_\lambda}, U^i_j>)^2 + 2\,p\,\frac{\varepsilon_1^2}{4} \leq 2U^2 \| b - \beta_{n_\lambda} \|^2 + p\,\frac{\varepsilon_1^2}{2}$$

$$\leq 2\,\frac{U^2}{\varepsilon^2} + \frac{\delta}{q}\,\frac{\varepsilon^2}{2} \text{ pour } \| b - \beta_{n_\ell} \| < \frac{1}{\varepsilon}. \quad (3)$$

De (2), on déduit d'autre part que :

$$<b^i - \beta^i_{n_\ell}, U^i_j>^2 \leq (|x^i_j - \theta^i_j| + \frac{\varepsilon_1}{2})^2 \leq 2|x^i_j - \theta^i_j|^2 + \frac{\varepsilon_1^2}{2}.$$

Donc, pour $j \in I_\ell$:

$$(x^i_j - \theta^i_j)^2 \geq \frac{1}{2} <b^i - \beta^i_n, U^i_j>^2 - q\,\frac{\varepsilon_1^2}{4}.$$

On en déduit :

$$\sum_{j\in I_\ell} (x^i_j - \theta^i_j)^2 \geq \frac{1}{2} \sum_{j\in I_\ell} <b^i - \beta^i_{n_\ell}, U^i_j>^2 - q\,\frac{\varepsilon_1^2}{4}.$$

$$\geq \frac{1}{2} (b^i - \beta^i_{n_\ell})' \sum_{j\in I_\ell} U^i_j U^{i\,\prime}_j (b^i - \beta^i_{n_\ell}) - q\,\frac{\varepsilon_1^2}{4} \geq \frac{1}{2}\,\delta \| b^i - \beta^i_{n_\ell} \|^2 - q\,\frac{\varepsilon_1^2}{4}$$

sous l'hypothèse 1vi.

D'après (1), on a, pour le i défini tel que $\| b^i - \beta^i_{n_\ell} \|^2 > \frac{\varepsilon^2}{p}$:

$$\sum_{j\in I_\ell} (x^i_j - \theta^i_j)^2 \geq \frac{1}{2}\delta\frac{\varepsilon^2}{p} - q\frac{\varepsilon_1^2}{4} = \frac{1}{4}\delta\frac{\varepsilon^2}{p}.$$

Donc : $\exists\, j \in I_\ell$: $(x^i_j - \theta^i_j)^2 \geq \frac{1}{4}\delta\frac{\varepsilon^2}{pq}$.

Par conséquent :

$$\exists\, j \in I_\ell : \| x_j - \theta_j \|^2 \geq \frac{1}{4}\delta\frac{\varepsilon^2}{pq}. \quad (4)$$

D'après (3) et (4), pour $\varepsilon < \| b - \beta_{n_\ell} \| < \frac{1}{\varepsilon}$, on a :

$$\exists\, j \in I_\ell : \frac{1}{4}\frac{\delta}{pq}\varepsilon^2 \leq \| x_j - \theta_j \|^2 \leq 2\frac{U^2}{\varepsilon^2} + \frac{1}{2}\frac{\delta}{q}\varepsilon^2.$$

On en déduit :

$$\exists\, j \in I_\ell,\ \exists\, 0 < \varepsilon_2 < 1 : \ \varepsilon_2 < \| x_j - \theta_j \| < \frac{1}{\varepsilon_2}.$$

Sous l'hypothèse 3ii, on a :

$$\exists\, j \in I_\ell : \ <x_j - \theta_j, M_j(x_j)> \ \geq c(\varepsilon_2).$$

Sous l'hypothèse 3i on a :

$$\sum_{j\in I_\ell} <x_j - \theta_j, M_j(x_j)> \ \geq c(\varepsilon_2).$$

Par conséquent :

$$b(\varepsilon) = \inf_{\ell > L(\varepsilon)} \ \inf_{\{\varepsilon < \| b - \beta_{n_\ell} \| < \frac{1}{\varepsilon}\}} \ \sum_{j\in I_\ell} <x_j - \theta_j, M_j(x_j)> \ \geq c(\varepsilon_2) > 0.$$

L'hypothèse 3ii du théorème 1 est vérifiée.

Hypothèse 3iii :

$$\forall \varepsilon > 0, \exists \eta_1 > 0 : (\|b_1 - b_2\| < \eta_1) \Rightarrow (\sup_n \|A_n^1 M_{1n}(b_1) - A_n^1 M_{1n}(b_2)\| < \varepsilon).$$

Sous l'hypothèse 1v du théorème 3, on a :

$$\|A_n^1 M_{1n}(b_1) - A_n^1 M_{1n}(b_2)\|^2 = \|A_n^1(M_n(x_{1n}) - M_n(x_{2n}))\|^2$$
$$= \sum_{i=1}^{p} \|U_n^i\|^2 (M_n^i(x_{1n}) - M_n^i(x_{2n}))^2 \leq U^2 \|M_n(x_{1n}) - M_n(x_{2n})\|^2.$$

Sous l'hypothèse 3iii du théorème 3, pour $\varepsilon > 0$, on a :

$$\exists \eta > 0 : (\|x_{1n} - x_{2n}\| < \eta) \Rightarrow (\|M_n(x_{1n}) - M_n(x_{2n})\| < \frac{\varepsilon}{U}).$$

Or : $\|x_{1n} - x_{2n}\|^2 = \sum_{i=1}^{p} (<b_1^i - b_2^i, U_n^i>)^2 \leq U^2 \|b_1 - b_2\|^2.$

L'hypothèse 3iii du théorème 1 est vérifiée avec $\eta_1 = \frac{\eta}{U}$.

4 - CONCLUSION

Nous avons établi un théorème de convergence presque sûre d'un processus d'approximation stochastique dynamique étudié à l'origine par Dupac, avec une hypothèse de stabilité plus générale que l'hypothèse classique et mieux adaptée au cas où les espaces des paramètres et des observations sont différents. Puis nous avons donné une généralisation de ce résultat.

Nous avons défini un modèle d'approximation stochastique dynamique, composition du précédent et d'un modèle étudié par Ruppert ; nous avons démontré la convergence presque sûre du processus correspondant par application du théorème précédent.

BIBLIOGRAPHIE

[1] **Albert, A.E., Gardner, L.A.**, Stochastic Approximation and Nonlinear Regression, Research Monograph 42, M.I.T. Press, Cambridge, Massachusetts, 1967.

[2] **Assili, O.**, Régression séquentielle par approximation stochastique, Thèse de Doctorat de troisième cycle, Université de Nancy I, 1985.

[3] **Bennar, A.**, Approximation stochastique : convergence dans le cas de plusieurs solutions et étude de modèles de corrélations, Thèse de Doctorat de troisième cycle, Université de Nancy I, 1985.

[4] **Benveniste, A., Metivier, M., Priouret, P.**, Algorithmes adaptatifs et approximations stochastiques, Théorie et applications, Masson, 1987.

[5] **Blum, J.R.**, "Multidimensional Stochastic Approximation Methods", AMS,1954, Vol. 25, pp. 737-744.

[6] **Dupac, V.**, "A Dynamical Stochastic Approximation Method", AMS, 1965, Vol. 36, pp. 1695-1702.

[7] **Dupac, V.**, "Stochastic Approximations in the Presence of Trend", Czechoslovak Math. J., 1966, Vol. 16 (91), pp. 454-461.

[8] **Gladyshev, E.G.**, "On Stochastic Approximation", Theory of Probability and its Applications, 1965, Vol. 10, pp. 275-278.

[9] I.E.E.E. TRANSACTIONS ON INFORMATION THEORY, 1984, Vol. IT-30, n° 2.

[10] **Kiefer, J., Wolfowitz, J.**, "Stochastic Estimation of the Maximum of a Regression Function", AMS, 1952, Vol. 23, pp. 462-466.

[11] **Ljung, L., Soderstrom, T.**, Theory and Practice of Recursive Identification, M.I.T. Press, 1983.

[12] **Monnez, J.M.**, Etude d'un processus général multidimensionnel d'approximation stochastique sous contraintes convexes. Applications à l'estimation statistique, Thèse de Doctorat ès Sciences Mathématiques, Université de Nancy I, 1982.

[13] **Nevel'son, M.B., Has'minskii, R.Z.**, Stochastic Approximation and Recursive Estimation, Translation of Mathematical Monographs, Vol. 47, American Mathematical Society, Providence, Rhode Island, 1973.

[14] **Robbins, H., Monro, S.**, "A Stochastic Approximation Method", AMS, 1951, Vol. 22, pp. 400-407.

[15] **Robbins, H., Siegmund, D.**, "A Convergence Theorem for Nonnegative almost Supermartingales and some Applications", Optimizing Methods in Statistics, edited by J.S. Rustagi, Academic Press, New York, 1971, pp. 233-257.

[16] **Ruppert, D.**,"A New Dynamic Stochastic Approximation Procedure", AS, 1979, Vol. 7, n° 6, pp. 1179-1195.

[17] **Uosaki, K.**, "Some Generalizations of Dynamic Stochastic Approximation Processes", AS, 1974, Vol. 2, n° 5, pp. 1042-1048.

[18] **Uosaki, K.**, "State Estimation Scheme for Nonlinear Dynamical Systems Based on the Stochastic Approximation", Information Sciences, 1975, Vol. 9, pp. 133-149.

[19] **Uosaki, K.**, "An Asymptotically more Efficient Nonlinear State Estimation Scheme Based on the Stochastic Approximation", Int. J. Systems Sci., 1977, Vol. 8, n° 9, pp. 971-983.

[20] **Venter, J.H.**, "On Dvoretzky Stochastic Approximation Theorems", AMS, 1966, Vol. 37, pp. 1534-1544.

[21] **Wasan, M.T.**, Stochastic Approximation, Cambridge Tracts in Mathematics and Mathematical Physics, n° 58, Cambridge University Press, 1969.