SEPTIEME COLLOQUE SUR LE TRAITEMENT DU SIGNAL ET SES APPLICATIONS

NICE du 28 MAI au 2 JUIN 1979



# SUR L'APPLICATION DE LA METHODE DE LJUNG A L'ETUDE DE LA CONVERGENCE D'UN ALGORITHME DE FILTRAGE ADAPTATIF

Dominique PERRIOT-MATHONNA

THOMSON-CSF (DSE)

## RESUME

Les algorithmes stochastiques récursifs, tels que les algorithmes d'approximation stochastique, ont introduit, par leur existence, les problèmes de convergence en probabilité.

Jusqu'à maintenant, ces techniques d'analyse de convergence restaient peu nombreuses et relativement spécifiques.

Les principaux résultats dans ce domaine sont certainement ceux dûs à Robbins et Monro (1) , Kiefer et Wolfowitz (2) , Blum (3) et Dvoretzky (4) .

Récemment, Ljung (5) a proposé une approche très générale du problème qui semble prometteuse. La méthode consiste :

- à associer à l'algorithme stochastique une équation différentielle déterministe,
- à montrer que les propriétés de convergence peuvent s'étudier en terme de stabilité de l'équation différentielle associée.

Dans une première partie, on rappelle brièvement les bases des algorithmes d'approximation stochastique et l'on introduit les résultats principaux établis par Ljung.

La seconde partie est consacrée à l'application de la méthode de Ljung à l'algorithme de filtrage adaptatif de Hampton (7) .

Enfin, on discute les conditions de convergence trouvées par cette approche et on compare ces résultats avec ceux de Hampton.

## SUMMARY

Recursive stochastic algorithms such as stochastic approximation algorithms introduce problems of convergence in probability.

Until now, methods of convergence analysis were very peculiar and often specific.

Main results in this area are surely those due to Robbins and Monro (1) , Kiefer and Wolfowitz (2) , Blum (3) and Dvoretzky (4) .

Recently, L. Ljung (5) proposed a new and original method. The basic ideas are :

- associating with the stochastic scheme a deterministic differential equation,
- proving that convergence properties can be studied in terms of stability of this ordinary differential equation.

Fundamental ideas about stochastic approximation algorithms are first briefly reviewed. Then, the main results of Ljung's works are introduced.

These results are used to study convergence properties of an adaptive filtering scheme due to Hampton (7) . Finaly, we compare results obtained by the Ljung approach with the Hampton ones.

# SUR L'APPLICATION DE LA METHODE DE LJUNG A L'ETUDE DE LA CONVERGENCE D'UN ALGORITHME DE FILTRAGE ADAPTATIF

## 1 - INTRODUCTION

Les algorithmes d'approximation stochastique constituent à la fois une technique d'optimisation de critères stochastiques et une méthode d'estimation récursive à partir d'un terme correctif quelconque : la seule exigence est que l'algorithme génère une suite d'estimés qui converge vers le paramètre cherché.

Ces algorithmes peuvent se mettre sous la forme générale suivante :

$$z(t) = z(t-1) + \gamma(t)\, Q(t; z(t-1), \varphi(t)) \quad (1)$$

$$\varphi(t) = A(z(t-1))\varphi(t-1) + B(z(t-1))\, e(t) \quad (2)$$

avec :

$\{z(.)\}$ : suite de vecteurs colonne de dimension n ; ce sont les estimés du vecteur paramètre inconnu

$\{\gamma(.)\}$ suite infinie de scalaires positifs

$\varphi(.)$ vecteur d'observation au temps t de dimension m : c'est une variable aléatoire dépendant généralement des estimés précédents et d'une suite de vecteurs aléatoires de dimension r

$Q(.;.,.)$ fonction déterministe de $\mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^m$ dans $\mathbb{R}^n$

$A(.), B(.)$ fonctions matricielles de dimensions respectives $m \times m$ et $m \times r$

La fonction $Q$ et la suite $\{\gamma\}$ décrivent complètement un algorithme.

On s'intéresse donc aux conditions sous lesquelles $\{z(t)\}$ converge vers la valeur vraie cherchée $z$.

Robbins et Monro (1) sont certainement les premiers à avoir introduit la notion d'algorithme d'approximation stochastique en cherchant la racine d'une fonction de régression. Kiefer et Wolfowitz (2) proposent un algorithme stochastique itératif pour déterminer l'extrémum d'une même fonction et démontrent que leur algorithme converge en probabilité.

Blum (3) reprenant les algorithmes exposés en (1) et (2) démontre, sous des hypothèses moins strictes, que ces algorithmes convergent avec la probabilité 1 .

Une forme générale d'algorithme a été étudiée par Dvoretzky (4) : le schéma proposé converge au sens des moindres carrés et avec la probabilité 1 .

En particulier, ses travaux généralisent les algorithmes définis en (1) et (2) .

Ces résultats, même ceux de Dvoretzky, ne permettent pas cependant de recouvrir l'ensemble des schémas récursifs stochastiques et, en particulier, les schémas multi-dimensionnels.

Les récents travaux de Ljung (5) ont apporté une solution à l'ensemble de ces problèmes : détermination des points possibles de convergence, comportement asymptotique des algorithmes, preuve de convergence avec probabilité 1 .

L'idée fondamentale de cette méthode présentée dans (5) et (6) est de transposer l'étude de la convergence de schémas stochastiques récursifs généralement non-linéaires en une étude de stabilité d'une équation différentielle déterministe ordinaire.

Au schéma récursif (1) et (2) , Ljung associe l'équation différentielle

$$\frac{d}{d\tau} z(\tau) = f(z(\tau)) \quad (3)$$

avec :

- $f(z) = E\{Q(z, \bar{\varphi}(k, z))\}$

$\bar{\varphi}$ étant une variable aléatoire, approximation de $\varphi(t)$ en supposant les estimés successifs $z(.)$ localement constants

- $\tau$ paramétrant les estimés avec la relation :

$$\tau_t = \sum_{k=1}^{t} \gamma(k)$$

Sous un certain nombre de conditions :

- caractérisant la nature stochastique du bruit $e(.)$
- précisant les propriétés de régularité de $A(.)$ $B(.)$ et surtout de $Q(.;.,.)$
- décrivant les caractéristiques de la suite $\{\gamma(.)\}$

Ljung démontre quatre théorèmes qui peuvent se résumer par les propriétés suivantes :

i) $\{z(t)\}$ ne peut converger que vers des points stationnaires stables de (3) : ce résultat permet essentiellement de constater qu'il y a ou non possibilité de convergence.

ii) si $z(t)$ appartient infiniment souvent avec la probabilité 1 au domaine d'attraction d'un point stationnaire stable $z^*$ de (3), alors $\{z(t)\}$ converge vers $z^*$ avec la probabilité 1 quand $t \to \infty$ . Ce résultat et son corollaire permettent de démontrer la convergence avec la probabilité 1 .

iii) les trajectoires de (3) sont les chemins asymptotiques des estimés $z(.)$ générés par l'algorithme (1) et (2) .

## 2 - APPLICATION DE LA METHODE DE LJUNG A L'ALGORITHME DE HAMPTON

### 2.1 - L'algorithme de filtrage adaptatif de Hampton

Hampton dans (7) présente un algorithme de filtrage de Kalman adaptatif utilisant une méthode d'approximation stochastique.

Ce schéma permet d'estimer récursivement la matrice de gain d'un filtre de Kalman en présence de statistiques de bruits inconnues à priori.

Considérons le système linéaire discret supposé complètement gouvernable et observable décrit par l'équation d'état et l'équation de mesure :

$$\left.\begin{aligned} x_{k+1} &= \phi x_k + B u_k + w_k \\ y_k &= H x_k + v_k \end{aligned}\right\} \quad (4)$$

Les bruits sont supposés blancs, décorrélés, à moyennes nulles.

L'algorithme de filtrage adaptatif de Hampton peut être décrit par :

$$\left.\begin{aligned} x_{k|k} &= x_{k|k-1} + \hat{K}_k (y_k - H x_{k|k-1}) \\ x_{k+1|k} &= \phi x_{k|k} + B u_k \end{aligned}\right\} \quad (5)$$

avec :

$$\hat{K}_{k+1} = \hat{K}_k + \gamma_{k+1} \left[\Delta_{k+m} - \hat{K}_k \nu_k\right] \nu_k^T \omega_{k+1} \quad (6)$$

$$\omega_{k+1} = \frac{k+1}{k}\left[\omega_k - \omega_k \nu_k (k + \nu_k^T \omega_k \nu_k)^{-1} \nu_k^T \omega_k\right] \quad (7)$$

où, à l'itération $k+1$ :

$\hat{K}_{k+1}$ est l'estimé courant de la matrice de gain du filtre

$\gamma_{k+1}$ est le terme de rang $k+1$ de la suite harmonique $\left\{\gamma_k = \frac{1}{k}\right\}$

$\nu_{k+1}$ est le vecteur d'innovation généré par le filtre $\nu_{k+1} \triangleq y_{k+1} - H x_{k+1|k}$

$\omega_{k+1}$ est le terme de rang $k+1$ d'une suite matricielle de pondération

$\Delta_{k+m}$ est un vecteur de dimension n (n = dimension du vecteur d'état) défini par :

$$\Delta_{k+m} = \begin{cases} (H\phi)^{-1} \nu_{k+1} & \text{si } H\phi \text{ est inversable} \\ \begin{bmatrix} H\phi^m \\ \vdots \\ H\phi^2 \\ H\phi \end{bmatrix}^{-1} \begin{bmatrix} d_{k+m} \\ \vdots \\ d_{k+2} \\ d_{k+1} \end{bmatrix} & \text{si } H\phi \text{ n'est pas inversable} \end{cases}$$

avec :

$m$ : indice d'observabilité du système

$d_{k+i}$ : résidu prédit défini par

$$d_{k+i} \triangleq y_{k+i} - H x_{k+i|k-1}$$

Cet algorithme et ses propriétés caractéristiques sont étudiés en détail dans (8) .

2.2 - Analyse de la convergence de l'algorithme par la méthode de Ljung

Elle nécessite, en premier lieu, la vérification des conditions évoquées au paragraphe 1 : cet aspect traité dans (8) ne sera pas rappelé ici.

On montre que l'on peut associer à (1) , (2) les équations différentielles déterministes :

$$\frac{d}{d\tau} K^T(\tau) = \omega(\tau)\, h(K(\tau)) \quad (8)$$

$$\frac{d}{d\tau} \omega^{-1}(\tau) = G(K(\tau)) - \omega^{-1}(\tau) \quad (9)$$

avec :

$$h(K) = \lim_{n\to\infty} E\left\{\nu_n(K)\nu_n^T(K)(K_{opt} - K)^T D(z^{-1})^T\right\} \quad (10)$$

$$G(K) = \lim_{n\to\infty} E\left\{\nu_n(K)\nu_n^T(K)\right\}$$

où :

$\nu_n(K)$ est l'innovation générée par le filtre au nième pas d'itération quand celui-ci est mis en oeuvre avec une matrice de gain K

$K_{opt}$ est la matrice de gain de Kalman optimale

$D(z^{-1})$ est une fonction de transfert dépendant seulement du système réel et du gain optimal

$$D(z^{-1}) = -K_{opt} H \left[I - \phi(I - K_{opt}H) z^{-1}\right]^{-1} \phi z^{-1} + I \quad (11)$$

On notera que la paramétrisation en $\tau$ est équivalente à une indexation par k avec :

$$\tau_k = \sum_{n=1}^{k} \gamma_n$$

2.2.1 - Etude de la stabilité

Cette étude peut être effectuée en utilisant la fonction de Lyapunov :

$$V(K, \omega^{-1}) = tr\, (K - K_{opt})\, \omega^{-1} (K - K_{opt})^T \quad (12)$$

où "$tr$" désigne l'opérateur "trace".

Comme $D(z^{-1})$ est un opérateur déterministe, (10) peut encore s'écrire :

$$h(K) = G(K)(K_{opt} - K)^T D(z^{-1})^T \quad (13)$$

En utilisant (8), (9) et (13), on démontre que :

$$\frac{d}{d\tau} V(K,\omega^{-1}) = -tr\,(K - K_{opt}) G(K) (K - K_{opt})^T \left[D(z^{-1}) + D(z^{-1})^T - I\right] - tr\,(K - K_{opt})\, \omega^{-1} (K - K_{opt})^T \quad (14)$$

Comme $\omega^{-1}$ est toujours définie positive (matrice de covariance), (14) sera strictement négative à condition que

$D(z^{-1}) + D(z^{-1})^T - I$ soit strictement positive réelle

ce qui entraine :

$D(z^{-1}) - \frac{I}{2}$ strictement positif réel (15)

2.2.2 - Application du corollaire de Ljung

L'existence de la fonction de Lyapunov (12) pour les équations différentielles (8) et (9) telle que:

$$\frac{d}{d\tau} V(K,\omega^{-1}) = \frac{d}{dz} V(K,\omega^{-1}) . \frac{dz}{d\tau} = V_z'(K,\omega^{-1}) . f(z) \leq 0$$

assure la seconde condition du corollaire.

La première condition relative au fait que $\varphi(t)$ soit borné est vérifiée avec :

$$\varphi_k = \left(x_{k+1|k}^T,\ u_k^T,\ y_k^T\right)^T$$

Le corollaire assure soit convergence avec probabilité 1 dans un ensemble $D_c$ soit existence d'un point d'accumulation sur la frontière d'un ensemble de définition $D_s$ .

- Convergence dans $D_c$ :

L'ensemble $D_c$ est défini par :

$$V_z' f(z) = 0 \Rightarrow \left\{\begin{aligned} h(K) &= 0 \\ G(K) &= \omega^{-1} \end{aligned}\right\} \Rightarrow (K - K_{opt})\nu_n(K) \equiv 0$$

SUR L'APPLICATION DE LA METHODE DE LJUNG A L'ETUDE
DE LA CONVERGENCE D'UN ALGORITHME DE FILTRAGE ADAPTATIF

donc il y a convergence avec probabilité 1 vers les points stationnaires de (8) et (9) et $\{\hat{K}_n\}$ converge avec la probabilité 1 quand n tend vers l'infini dans :

$$D_c = \{K / (K - K_{opt})\,\gamma_n(K) \equiv 0\}$$

On peut alors démontrer (Cf. réf. 8 ) que :

$$K \in D_c \Rightarrow K.\gamma_n(K) \equiv K_{opt}.\gamma_n(K_{opt}) \qquad (16)$$

La relation (16) signifie que l'algorithme récursif (6) , (7) génère une suite d'estimés $\{\hat{K}_n\}$ qui converge avec la probabilité 1 vers une valeur permettant d'obtenir la représentation entrée-sortie optimale du filtre de Kalman :les termes d'estimation $\hat{K}_n.\gamma_n(\hat{K}_n)$ convergeront vers l'optimum $K_{opt}.\gamma_n(K_{opt})$ quand n tend vers l'infini.

2.2.3 - Comportement de l'algorithme quand K est à l'extérieur de $D_s$ ou proche de la frontière

Rappelons que $D_s$ est défini par :

$D_s = \{K \mid \phi(I - KH)$ ait toutes ses valeurs propres à l'intérieur du cercle unité $\}$

Cet ensemble est donc connu de l'utilisateur et l'on peut toujours imaginer une modification de l'algorithme pour que $K \in D_s$

Cependant, si $\hat{K}_k$ est proche d'une zone d'instabilité :

$$\hat{K}_k.\gamma_k \gg \Delta_{k+m}$$

et (6) peut être approché par :

$$\hat{K}^T_{k+1} \simeq \hat{K}^T_k + \frac{\omega_{k+1}}{k+1}\gamma_k\gamma_k^T\hat{K}^T_k = \left[I - \frac{\omega_{k+1}}{k+1}\gamma_k\gamma_k^T\right]K_k^T$$

On démontre facilement que cette application est contractante ; $\hat{K}_{k+1}$ ne sera pas infiniment proche de $\hat{K}_k$ et l'algorithme ne pourra converger dans cette région.

2.2.4 - Non existence d'un point d'accumulation en-dehors ou sur la frontière de $D_s$

La démonstration de cette proposition s'effectue d'une manière analogue à celle du théorème I de Ljung (Cf. réf. 5 ).

i) On démontre que la fonction de Lyapunov (12) est décroissante en-dehors ou sur la frontière de $D_s$ .

En utilisant la condition de positivité trouvée (15) , on peut en effet prouver que :

$$tr\,(\hat{K}_{k+1} - K_{opt})\,\omega^{-1}_{k+1}(\hat{K}_{k+1} - K_{opt})^T < tr\,(K_k - K_{opt})\,\omega^{-1}_k(K_k - K_{opt})^T$$

ii) Ljung établit dans 5 qu'une telle fonction de Lyapunov est partout décroissante sauf aux points d'accumulation.

Donc, s'il existait un point d'accumulation à l'extérieur ou sur la frontière de $D_s$ , ii) serait vérifiée, ce qui serait contradictoire avec i).

2.3 - Résultats obtenus

L'étude menée permet de démontrer, dans le cas général, les résultats suivants :

i) si $\phi$ est stable, les estimés $\hat{K}_n$ appartiendront à $D_s$ infiniment souvent

ii) sous la condition

$$D(z^{-1}) - \frac{I}{2} \qquad \text{strictement réelle positive}$$

avec $D(z^{-1})$ défini par (11) , l'algorithme (6) et(7) génère une suite d'estimés $\{\hat{K}_n\}$ qui converge, avec la probabilité 1, quand n tend vers l'infini dans un ensemble $D_c$ tel que :

$$D_c = \{K \mid (K - K_{opt})\,\gamma_n(K) \equiv 0\}$$

Comme les innovations ne sont jamais identiquement nulles, $\{\hat{K}_n\}$ converge vers $K_{opt}$ avec la probabilité 1.

2.4 - Cas des processus à sortie scalaire

L'étude qui précède a été effectuée dans un cas général. Dans le cas de l'utilisation de l'algorithme de Hampton face à un processus multidimensionnel ou non, mais à sortie scalaire, l'algorithme est du type gradient.

L'analyse de stabilité peut être alors menée avec la fonction de Luapunov

$$V(K,\omega^{-1}) = tr\,(K - K_{opt})(K - K_{opt})^T$$

et, on aboutit à une condition de positivité moins stricte :

$$\frac{d}{d\tau}V(K,\omega^{-1}) < 0 \qquad \text{si } D(z^{-1}) \text{ est strictement réel positif}$$

L'application du corollaire et l'étude 2.2.3 sont identiques ; la non-existence d'un point d'accumulation à l'extérieur ou à la frontière de $D_s$ peut être démontrée sous la seule condition

$$D(z^{-1}) \qquad \text{strictement positif réel}$$

3 - DISCUSSION ET CONCLUSIONS

3.1 - Sur la nature de la condition de positivité - $D(z^{-1})$ strictement positif réel est une condition implicite

On peut, en effet, démontrer que la condition $D(z^{-1})$ strictement positif réel est implicite et qu'elle n'entraîne aucune restriction particulière tant sur l'algorithme que sur le processus.

SUR L'APPLICATION DE LA METHODE DE LJUNG A L'ETUDE DE LA CONVERGENCE D'UN ALGORITHME DE FILTRAGE ADAPTATIF

En effet,

$$(K-K_{opt})\,h(K) = -(K-K_{opt})\,G(K)\,(K-K_{opt})^T D(z^{-1})^T$$

or :

$$(K-K_{opt})\,h(K) = (K-K_{opt}) \lim_{n\to\infty} E\{\gamma_n(\Delta_{n+m}-K.\gamma_n)^T\}$$

$$= \lim_{n\to\infty} E\{(K-K_{opt})\gamma_n[\Delta_{n+m}-K_{opt}\gamma_n-(K-K_{opt})\gamma_n]^T\}$$

En utilisant la condition nécessaire et suffisante d'optimalité du filtre

$$E\{\gamma_n[\Delta_{n+m}-K_{opt}.\gamma_n]^T\}=0 \iff K=K_{opt}$$

on déduit directement que :

$$(K-K_{opt})\,h(K) = -(K-K_{opt})\,G(K)\,(K-K_{opt})^T D(z^{-1})^T$$

est définie négative sauf aux points de convergence où :

$$h(K)=0 \iff (K-K_{opt})\gamma_n = 0$$

ce qui prouve le caractère réel positif de $D(z^{-1})$

- Caractère nécessaire et suffisant de la condition de positivité

On peut remarquer que $D(z^{-1})$ strictement positif réel est, dans le cas d'un système à sortie scalaire, une condition nécessaire et suffisante de stabilité des équations différentielles (8) et (9) .
Par contre, on note que l'étude de stabilité dans le cas général où nous l'avons présenté, ne nécessite pas forcément la condition $D(z^{-1})-\frac{I}{2}$ strictement réel positif.

3.2 - Comparaisons des résultats obtenus par la méthode de Ljung avec ceux de Hampton

Dans ⑦ , Hampton démontre, dans le seul cas de systèmes mono-dimensionnels, que l'algorithme converge avec la probabilité 1 ; pour cela, il utilise un théorème dû à Gladyshev ⑨ qui, appliqué dans ce cas, impose les quatre conditions :

i) $\sum_{n=1}^{\infty} \gamma_n \omega_n = \infty$

ii) $\sum_{n=1}^{\infty} (\gamma_n\omega_n)^2 < \infty$

iii) $(\hat{K}_n-K_{opt})\, E\{\nabla_K \ell\,(\Delta_{n+1}-\hat{K}_n\gamma_n)\} \geq 0$
pour $|\hat{K}_n-K_{opt}|<\varepsilon$ et $\forall\varepsilon$

iv) $\exists r>0 \;:\; E\{[\nabla_K \ell\,(\Delta_{n+1}-\hat{K}_n\gamma_n)]^2\} \leq r\,(1+\hat{K}_n)^2$

avec :

$$\ell(\Delta_{n+1}-\hat{K}_n\gamma_n) = \frac{1}{2}(\Delta_{n+1}-\hat{K}_n\gamma_n)^T(\Delta_{n+1}-\hat{K}_n\gamma_n)$$

$\nabla_K \ell$ = gradient de $\ell$ par rapport à $K$

Ces conditions sont vérifiées sans introduire de restriction.

On peut les comparer aux conditions de Ljung.

- Conditions (i) et (ii) :

Elles sont analogues aux conditions de Ljung sur la suite $\{\gamma_n\}$ : celle de Ljung est moins restrictive puisqu'il suffit que

$$\exists p>1 \;:\; \sum_{n=1}^{\infty} \gamma_n^p < \infty$$

- Condition (iii) :

On peut l'interpréter comme une condition de stabilité pour l'équation différentielle :

$$\frac{d}{d\tau} z = f(z)$$

avec la fonction de Lyapunov :

$$V(z) = (z - z_{opt})^2$$

ce qui conduit à :

$(z-z_{opt})f(z)\leq 0$ analogue à $(\hat{K}_n-K_{opt})\,E\{\gamma_n(\Delta_{n+1}-\hat{K}_n\gamma_n)^T\}\leq 0$

- Condition (iv) :

Elle exprime que le terme d'actualisation des estimés ne croît pas plus vite qu'une fonction quadratique en $K$ .
On peut rapprocher cette condition de la première condition "$\varphi$ bornée" du corollaire de Ljung. Si $Q(t;z,\varphi)$ croît trop rapidement avec $z$, il peut arriver qu'au cours des itérations, $\hat{K}_n \notin D_S$ ou même $\hat{K}_n \to \infty$ . C'est pourquoi nous avons démontré que si $\hat{K}_n \notin D_S$ ou est à la frontière de $D_S$, l'algorithme permet l'évolution des estimés pour retourner dans $D_S$ .
Cette comparaison met en lumière les traits communs des méthodes d'analyse ainsi que les résultats qu'elles fournissent :
- dans le cas de processus mono-dimensionnels, les deux méthodes permettent de montrer, sans aucune condition, qu'il y a convergence avec la probabilité 1,
- dans le cas général, Hampton ne démontre pas la convergence de son algorithme. Par contre, la méthode de Ljung permet d'établir la convergence avec probabilité 1 sous une condition :

$$D(z^{-1})-\frac{I}{2} \quad \text{strictement positif réel}$$

Cette condition semble impossible à vérifier puisqu'elle nécessite la connaissance de la matrice de gain optimale $K_{opt}$ que l'on cherche à estimer.
Des études sont actuellement en cours pour essayer de démontrer cette condition directement en utilisant la notion de positivité.

La méthode de Ljung que nous avons appliquée ici est certainement la première méthode générale d'analyse de convergence des algorithmes stochastiques récursifs . Cependant, cette étude et certains résultats de simulation semblent montrer que le problème consistant à trouver des conditions nécessaires et suffisantes de convergence reste encore très ouvert.

## BIBLIOGRAPHIE


1. ROBBINS H. ET MONRO S "A stochastic approximation method" Annals of Math. Stat., vol 22,pp.400-407 (1951)

2. KIEFER J. ET WOLFOWITZ J. "Stochastic estimation of the maximum of a regression function"- Annals of Math. Stat., vol 23, pp. 462-466 (1952)

3. BLUM J.R. "Approximation methods which converge with probability one" Annals of Math. Stat., vol 25,pp. 382-386 (1954)

4. DVORETZKY A. "On stochastic approximation" Proceedings 3rd Berkeley Symposium on Math. Stat. and Probability, vol 1, pp.39-56 (1956)

5. LJUNG L. "Analysis of recursive stochastic algorithms" - IEEE Transactions on Automatic Control, AC-22, pp. 551-575 (1977)

6. LJUNG L. "Strong convergence of a stochastic approximation algorithm" The Annals of Statistics, vol 6, pp.680-696 (1978)

7. HAMPTON R.L.T. "Stochastic algorithms for self - adaptive filtering and prediction" NASA 03-002-006 semi-annual Report - Part A-N71-34180 (1971)

8. PERRIOT-MATHONNA D. Thèse de Docteur Ingènieur à publier (Université de Paris Sud - Orsay)

9. GLADYSHEV E.G. "On stochastic approximation" Theory of Probability and its applications, vol 10, n°2, pp. 275-278 (1965)