Etude asymptotique d'algorithmes de poursuite

BERNARD DELYON and ANATOLI JUDITSKY
IRISA, Campus de Beaulieu, 35042 Rennes France

RÉSUMÉ

Cet article discute du problème de la poursuite de paramètres de régression animés d'un mouvement aléatoire. Nous étudions les propriétés asymptotiques de certains algorithmes en présence de petit bruit. Toute l'étude est basée sur un théorème-limite central pour un type assez général d'équation aux différences.

ABSTRACT

The paper addresses the problem of tracking random drifting parameters of linearregression system. We study the asymptotic properties of several estimation algorithms in the limit of slow drift. The basic tool is the central limit theorem for a class of stochastic difference equations established under weak conditions on disturbances and observations. The estimates of the rate of convergence obtained in the paper allow us to develop the asymptotically optimal algorithms.

1 Introduction

Nous considérons ici les algorithmes de la forme

$$y_t = \theta_t^T \varphi_t + e_t.$$

où e_t est un bruit aléatoire, $\varphi_t \in \mathbf{R}^N, y_t \in \mathbf{R}$ sont les entrées et sorties observées et φ_t est un vecteur de paramètres inconnu, variable dans le temps. On suppose que θ_t est aléatoire et varie lentement, i.e.

$$\theta_t = \theta_{t-1} + \gamma w_t \tag{1}$$

où γ est un petit paramètre.

L'identification du système conduit généralement à un algorithme de la forme suivante (cf [6]):

$$\hat{\theta}_t = \hat{\theta}_{t-1} + \mu L_t(y_t - \varphi_t^T \hat{\theta}_{t-1}), \quad \hat{\theta}_t \in \mathbf{R}^N, \tag{2}$$

Chaque méthode se caractérise par la manière dont le gain vectoriel L_t est réestimé. μ est un réel positif. L'article se consacre à l'étude asymptotique (quand $\mu \to 0$ et $\gamma \to 0$) de la distribution stationaire de l'erreur d'estimation $\hat{\theta}_t - \theta_t$. Plus precisément, notre objectif est de montrer que pour diverses méthodes, que la distribution de l'erreur converge vers une loi normale de variance calculable.

Soit $\Gamma = \Gamma^T > 0$. On considérera les méthodes de réestimation suivantes:

- Moindres carrés de Widrow (LMS):

$$L_t = \Gamma\varphi_t \quad \text{ou} \quad L_t = \Gamma\varphi_t/(1 + \mu\varphi_t^T\Gamma\varphi_t)$$

- Moindres carrés normalisés (LMS):

$$L_t = \Gamma\varphi_t/(1 + |\varphi_t|^2)$$

- Moindres carrés récursifs avec facteur d'oubli (RLS):

$$\begin{aligned} R_t &= (1-\mu)R_{t-1} + \mu\varphi_t\varphi_t^T, \ R_0 = \varrho I, \ \varrho > 0 \\ L_t &= R_t^{-1}\varphi_t \end{aligned}$$

- Moindres carrés stabilisés ($0 < \rho < 1$)

$$\begin{aligned} L_t &= \Gamma_t\varphi_t \\ \Gamma_t &= (\mu\rho^{-1}R_t + \Gamma^{-1})^{-1} \\ R_t &= (1-\rho)R_{t-1} + \rho\varphi_t\varphi_t^T, \quad R_0 = 0 \end{aligned}$$

Remarques: Ces algorithmes sont bien connus (cf. [6]) à l'exception des moindres carrés stabilisés. Ce dernier a été conçu comme une version régularisée de MS et RLS à la fois (cf [4]).

Le point théorique principal de l'article est le théorème-limite central pour les équations aux différences developpé en section 2. Il donne une approximation en *horizon infini* pour la distribution asymptotique de l'erreur $\hat{\theta}_t - \theta_t$ de l'équation (2). Nous donnons ensuite un guide pour l'application de ce resultat aux algorithmes d'estimation. En section 3 nous considérons les algorithmes décrits ci-dessus et on donne des conditions d'obtention du théorème-limite central ainsi que les expressions de la covariance asymptotique de l'erreur.

2 Resultat fondamental

2.1 Théorème-limite central

On se donne un espace de probabilité $\{\Omega, \mathcal{F}, P\}$. Considérons le processus (Δ_t), $\Delta_t \in \mathbf{R}^N$ satisfaisant l'équation:

$$\Delta_t = (I - \mu P_t)\Delta_{t-1} + \mu\zeta_t \quad \Delta_0 \in \mathbf{R}^N \tag{3}$$

où $\mu \leq \mu_0$ est un réel positif, (P_t) et (ζ_t) sont des processus aléatoires, à valeurs dans $\mathbf{R}^{N\times N}$ et $\mathbf{R}^N$ respectivement. Nous allons énoncer (sans preuve, faute de place) le théorème-limite pour la solution Δ_t de (3). On autorise les processus P_t et ζ_t à dependre de μ, mais on suppose qu'on peut les décomposer de la façon suivante:

$$\begin{aligned} P_t &= P_t^0 + P_t' && (4) \\ \zeta_t &= \zeta_t^0 + \zeta_t' && (5) \end{aligned}$$

P_t^0 and ζ_t^0 ne dépendent pas de μ et P_t' et ζ_t' satisferont des hypothèses particulières. Supposons qu'existe un processus strictement stationnaire (P_t^0, ζ_t^0, X_t) and $\mathcal{F}_t = \sigma\{\ldots, P_t^0, \zeta_t^0, X_t\}$, tels que $P_t', \zeta_t' \in \mathcal{F}_t$. Les hypothèses sont:

[A1] $P_t^0 = (P_t^0)^T \geq 0$, $EP_1^0 > 0$, $E|P_1^0|^2 < \infty$.

[A2] (P_t^0) and (ζ_t^0) sont des processus strictement stationnaires ergodiques.

[A3] $\sum_{s=0}^{\infty}(E|E(\zeta_s^0|\mathcal{F}_0)|^2)^{1/2} < \infty$

[A4] $\lim_{\mu\to 0}\mu^{-1/2}\sup_t E(|\zeta'_t|) = 0$

[A5] Il existe un processus stationnaire ergodique p'_t satisfaisant:

$$|P'_t| \le p'_t, \quad \lim_{\mu_0\to 0} E(\sup_{\mu\le\mu_0}[p'_t]^2) = 0$$

$$\sup_t E([p'_t]^2) < \infty$$

Théorème 1 *Sous les hypothèses A1–A5, le système (3) avec toute condition initiale déterministe Δ_0 satisfait*

$$\lim_{\mu\to 0}\lim_{t\to\infty}\mu^{-1/2}\Delta_t = \mathcal{N}(0,V) \quad \textit{en loi} \tag{6}$$

i.e. $\mu^{-1/2}\Delta_t$ converges en loi vers la variable Gaussienne de moyenne nulle et de covariance V, où V est la solution de l'équation de Liapounov

$$BV + VB = S_0 \tag{7}$$

avec $B = EP_1^0$ et $S_0 = E\zeta_0^0\zeta_0^{0T} + \sum_{i=1}^{\infty} E\zeta_0^0\zeta_i^{0T} + E\zeta_i^0\zeta_0^{0T}$

Notons que les hypothèses A2, A3 garantissent l'existence de $S_0 < \infty$.
Le sens precis de la limite ci-dessus est

$$\lim_{\mu\to 0}\limsup_{t\to\infty}|E[f(\mu^{-1/2}\Delta_t)] - E[f(\mathcal{N}(0,V))]| = 0$$

pour toute fonction f continue bornée.

Remarques: Ce résultat sera appliqué à $\Delta_t = \hat{\theta}_t - \theta_t$. Notons que les hypothèses du théorème donnent une approximation *en horizon infini*. Ceci constitue une difference majeure avec les resultats habituels d'approximation, où l'on prouve la convergence faible sur un interval fini d'ordre γ^{-1}. Comparons le théorème 1 avec des résultats analogues comme le théorème 4.15, partie 2 de [1] ou le résultat correspondant de [5]. En fait ces résultats demandent implicitement que les trajectoires $(\hat{\theta}_t)$ de l'équation (3) soient bornées. Par ailleurs, la condition sur la dépendance de la suite (P_t) dans le théorème 1 est plus explicite et moins restrictive que celle, par exemple, du théorème 4.15 de [1] puisque l'ergodicité seule est réclamée.

Comparé avec les résultats obtenus dans [3], [13], [8]-[10] cet article ne considère pas les moments de $\hat{\theta}_t - \theta_t$ (qui peuvent très bien être infinis) mais s'intéresse à la limite en loi de ces processus. L'avantage notable de cette approche est la faiblesse des hypothèses imposées dans le théorème. Remarquons que les conclusions du théorème n'impliquent pas la convergence en loi de $\mu^{-1/2}\Delta_t$ pour μ fixé; il est simplement dit que la distance entre la distribution de $\mu^{-1/2}\Delta_t$ et la Gaussienne tend vers zéro.

2.2 Utilisation du théorème 1

L'équation (2) conduit à la récursion suivante pour l'erreur d'estimation $\delta_t = \theta_t - \hat{\theta}_t$

$$\delta_t = (I - \mu L_t\varphi_t^T)\delta_{t-1} + \mu L_t e_t - \gamma w_t \tag{8}$$

et si l'on pose

$$\begin{aligned}
\delta_t^{(1)} &= (I - \mu L_t\varphi_t^T)\delta_{t-1}^{(1)} + \sqrt{\mu}L_t e_t, \quad \delta_0^{(1)} = \delta_0 \\
\delta_t^{(2)} &= (I - \mu L_t\varphi_t^T)\delta_{t-1}^{(2)} - \sqrt{\mu}w_t, \quad \delta_0^{(2)} = 0 \\
\Delta_t^T &= \sqrt{\mu}(\delta_t^{(1)}, \delta_t^{(2)})^T
\end{aligned}$$

alors

$$\delta_t = \sqrt{\mu}\delta_t^{(1)} + \frac{\gamma}{\sqrt{\mu}}\delta_t^{(2)} \tag{9}$$

et l'équation (3) equivaut à (8) avec

$$P_t = \begin{pmatrix} L_t\varphi_t^T & 0 \\ 0 & L_t\varphi_t^T \end{pmatrix}, \quad \zeta_t = \begin{pmatrix} L_t e_t \\ w_t \end{pmatrix} \tag{10}$$

Donc, quand μ tend vers zéro, δ est donné par l'équation (9) où le vecteur $(\delta^{(1)}, \delta^{(2)})$ is approximé (indépendemment de γ) par un vecteur Gaussien de matrice de covariance d'ordre 1. Notons que le choix optimal du gain est $\mu \simeq \gamma$. Les cas $\mu = o(\gamma)$ ou $\gamma = o(\mu)$ ne sont pas dénués d'intéret car, en pratique, le paramètre γ est assez difficile à estimer. Introduisons la notation suivante

Definition 1: *Soit*

$$\delta_t = \sqrt{\mu}\delta_t^{(1)} + \frac{\gamma}{\sqrt{\mu}}\delta_t^{(2)}$$

où $\delta_t^{(1)}$ et $\delta_t^{(2)}$ sont deux processus aléatoires dépendant du paramètre μ seulement. Si $(\delta_t^{(1)}, \delta_t^{(2)})$ converge en loi vers la variable Gaussienne $N(0,V)$, avec

$$V = \begin{pmatrix} V_1 & 0 \\ 0 & V_2 \end{pmatrix}$$

alors, on notera $\delta_t \asymp N(0, \mu V_1 + \frac{\gamma^2}{\mu}V_2)$.

3 Etude des algorithmes

Afin de simplifier la presentation on supposera que e_t et w_t sont des accroissements de martingale.

On note $\mathcal{F}'_t = \sigma\{\ldots, \varphi_{t+1}, e_t, w_t\}$ et $\mathcal{G}'$ la tribu invariante de la filtration (φ_t, e_t, w_t). On considérera les hypothèses:

[B1] (φ_t, e_t, w_t) est une suite strictement stationnaire ergodique; $E(e_t|\mathcal{F}'_{t-1}) = 0$, $E(w_t|\mathcal{F}'_{t-1}) = 0$, $Ee_t^4 < \infty$, $E(e^2|\mathcal{F}'_{t-1}) = \sigma_e^2$, $E(w_t w_t^T|\mathcal{F}'_{t-1}) = R_w$, et $E(w_t e_t|\mathcal{F}'_{t-1}) = 0$.

[B2] $E|\varphi_t|^4 < \infty$ et $E\varphi_t\varphi_t^T > 0$.

[B3] $\Gamma = \Gamma^T > 0$.

3.1 Algorithme de Widrow (LMS)

On considère l'algorithme [14]

$$\hat{\theta}_t = \hat{\theta}_{t-1} + \mu\Gamma\varphi_t(y_t - \varphi_t^T\hat{\theta}_{t-1}) \tag{11}$$

ainsi qu'une version modifiée

$$\hat{\theta}_t = \hat{\theta}_{t-1} + \mu\Gamma\frac{\varphi_t}{1 + \mu\varphi_t^T\Gamma\varphi_t}(y_t - \varphi_t^T\hat{\theta}_{t-1}) \tag{12}$$

Posons $D = E\varphi_1\varphi_1^T$ et définissons les matrices V_i à l'aide des équations de Liapounov

$$\Gamma D V_1 + V_1 D\Gamma = \sigma_e^2\Gamma D\Gamma \tag{13}$$

$$\Gamma D V_2 + V_2 D\Gamma = R_w \tag{14}$$

éorème 2 *Sous les hypothèses B1-B3, l'algorithme (11) (12) (pour ce dernier on exige aussi $E|\varphi_t|^8 < \infty$) satisit*

$$\hat\theta_t - \theta_t \asymp \mathcal{N}(0, \mu V_1 + V_2\gamma^2/\mu)$$

emonstration: Posons

$$\delta_t = \Gamma^{-1/2}(\hat\theta_t - \theta_t),\ \bar\varphi_t = \Gamma^{1/2}\varphi_t,\ \bar w_t = \Gamma^{-1/2}w_t$$

$$\bar P_t^0 = \bar\varphi_t\bar\varphi_t^T \quad \zeta_t^0 = \begin{pmatrix} \bar\varphi_t e_t \\ \bar w_t \end{pmatrix}$$

$$\bar P_t^1 = \frac{\bar\varphi_t\bar\varphi_t^T}{1+\mu|\bar\varphi_t|^2} \quad \zeta_t^1 = \begin{pmatrix} \frac{\bar\varphi_t e_t}{1+\mu|\bar\varphi_t|^2} \\ \bar w_t \end{pmatrix}$$

vec ces notations, il vient

$$\delta_t = \sqrt{\mu}\delta_t^{(1)} + \delta_t^{(2)}\gamma/\sqrt{\mu}$$

$\Delta_t = \sqrt{\mu}(\delta_t^{(1)}, \delta_t^{(2)})$ satisfait

$$\Delta_t = (I - \mu\bar P_t)\Delta_{t-1} + \mu\zeta_t \tag{15}$$

$$\bar P_t = \begin{pmatrix} \bar P_t^i & 0 \\ 0 & \bar P_t^i \end{pmatrix},\ \zeta_t = \zeta_t^i.$$

n choisit i=0 ou i=1 pour l'algorithme (11) ou (12) respecvement. Les processus P_t^0, ζ_t^0 vérifient A1-A5 du théorème et donc le théorème est immediat pour l'algorithme (11). Notons que P_t^1 se decompose en $P_t^1 = P_t^0 + P_t'$ où

$$P_t' = -\mu\frac{|\bar\varphi_t|^2\bar\varphi_t\bar\varphi_t^T}{1+\mu|\bar\varphi_t|^2}$$

n vérifie sans peine que l'hypothèse A5 est vérifiée. De plus,

$$E|\zeta_t^1 - \zeta_t^0| \leq \mu E(|\bar\varphi_t|^3)\sigma_e$$

e qui implique A4. On peut donc appliquer le théorème 1. ∎

Pour terminer l'étude asymptotique de l'algorithme de Wirow, développons sa version optimale. Soit $\mu = \gamma$. Alors

$$\gamma^{-1/2}(\hat\theta_t - \theta_t) \xrightarrow{D} \mathcal{N}(0, V_3)$$

ù $V_3 = V_1 + V_2$ est la solution de l'équation:

$$\Gamma D V_3 + V_3 D\Gamma = \sigma_e^2\Gamma D\Gamma + R_w \tag{16}$$

)n a alors le corollaire suivant:

orollaire 1 *Considérons une matrice de gain Γ^* satisfaisant*

$$\Gamma^* D\Gamma^* = \frac{R_w}{\sigma_e^2}.$$

lors pour tout choix de Γ satisfaisant B3

$$V(\Gamma) \geq V(\Gamma^*) = \sigma_e^2\Gamma^*.$$

La démonstration de ce résultat se trouve dans [11]. Noons que $V(\Gamma^*)$ vaut

$$V(\Gamma^*) = \sigma_e D^{-1/2}(D^{1/2}R_w D^{1/2})^{1/2}D^{-1/2} \tag{17}$$

i les bruits e_t et w_t ont une distribution Gaussienne, cette aleur est la plus petite possible qu'un algorithme d'estimaion puisse atteindre (cf [11]).

3.2 Algorithme LMS normalisé

On considère l'algorithme

$$\hat\theta_t = \hat\theta_{t-1} + \mu\Gamma\frac{\varphi_t}{1+|\varphi_t|^2}(y_t - \varphi_t^T\hat\theta_{t-1}). \tag{18}$$

Il fut introduit dans la littérature (cf [9]) comme une modification du LMS habituel. La stabilité dans L_p de cette méthode a été prouvée dans [9] et des estimées de la vitesse de convergence y sont données.

On note $D = E\varphi_t\varphi_t^T(1+|\varphi_t|^2)^{-1}$ et $G = E\varphi_t\varphi_t^T(1+|\varphi_t|^2)^{-2}$. Remarquons que si B2 est satisfaite alors $D > 0$ et $G > 0$. Définissons les matrices V_i par les équations de Liapounov

$$\Gamma D V_1 + V_1 D\Gamma = \sigma_e^2\Gamma G\Gamma$$
$$\Gamma D V_2 + V_2 D\Gamma = R_w$$

Théorème 3 *Sous B1-B3, l'algorithme (18) satisfait*

$$\hat\theta_t - \theta_t \asymp \mathcal{N}(0, \mu V_1 + V_2\gamma^2/\mu)$$

La démonstration est analogue à celle du théorème 2.

3.3 Algorithme avec facteur d'oubli

$\hat\theta_t$ satisfait ici l'équation

$$\begin{aligned} \hat\theta_t &= \hat\theta_{t-1} - \mu R_t^{-1}\varphi_t(y_t - \hat\theta_{t-1}^T\varphi_t) \\ R_t &= (1-\mu)R_{t-1} + \mu\varphi_t\varphi_t^T,\ R_0 = \varrho I \end{aligned} \tag{19}$$

avec $\varrho > 0$. Posons $D = E\varphi_t\varphi_t^T$, $V_1 = \sigma_e^2 D^{-1}/2$ et $V_2 = R_w/2$

Théorème 4 *Sous B1 et B2, l'algorithme (19)*

$$\hat\theta_t - \theta_t \asymp \mathcal{N}(0, \mu V_1 + V_2\gamma^2/\mu)$$

Des conditions suffisantes pour la stabilité dans L_p ainsi que des estimées de la vitesse de convergence se trouvent dans [3] et [13].
On aura besoin du lemme suivant (admis):

Lemme 1 *Soit (φ_t) un processus stationnaire ergodique à valeurs dans $\mathbf{R}^N$, $E\varphi_t\varphi_t^T = D$ and $E|\varphi_t|^4 < \infty$. Définissons R_t par*

$$R_t = (1-\mu)R_{t-1} + \mu\varphi_t\varphi_t^T,\ R_0 = \varrho I > 0$$

Alors $\sup_{0<\mu<1}\sup_t E|R_t - D|^2 < \infty$, *et* $\lim_{\mu\to 0}\lim_{t\to\infty} E|R_t - D|^2 = 0$.

Démonstration du théorème 4: On ne considérera que le cas $\mu = \gamma$. Le processus $x_t = R_t(\hat\theta_t - \theta_t)$ satisfait l'équation (cf. [2])

$$x_t = (1-\gamma)x_{t-1} + \gamma\varphi_t e_t + \gamma R_t w_t.$$

On peut alors faire la décomposition

$$x_t = \bar x_t + \bar\delta_t$$

où

$$\begin{aligned} \bar\delta_t &= (1-\gamma)\bar\delta_{t-1} + \gamma(R_t - D)w_t,\ \bar\delta_0 = 0 \\ \bar x_t &= (1-\gamma)\bar x_{t-1} + \gamma\varphi_t e_t + \gamma D w_t,\ \bar x_0 = x_0. \end{aligned}$$

On déduit du lemme 1 que les conditions du théorème 1 sont satisfaites par le processus $(\bar{x}_t)$; par conséquent

$$\gamma^{-1/2}\bar{x}_t \xrightarrow{D} \mathcal{N}(0, \bar{V})$$

avec $\bar{V} = \sigma_e^2 D + DR_w D$. Puisque $(R_t - D)w_t$ est un accroissement de martingale, on a

$$E|\bar{\delta}_t|^2 \leq (1-\gamma)^2 E|\bar{\delta}_{t-1}|^2 + \gamma^2 |R_w| E|R_t - D|^2 = o(\gamma)$$

Donc $\gamma^{-1/2}\bar{\delta}_t \xrightarrow{P} 0$. Le lemme 1 permet de conclure que pour tout $0 < \varepsilon \leq 1/2$ il existe $\mu(\varepsilon)$ tel que $P(|R_t - D| > \varepsilon) < \varepsilon$ dès que $\mu \leq \mu(\varepsilon)$. Et donc $P(|R_t^{-1} - D^{-1}| > 2\lambda_{min}^{-1}(D)\varepsilon) < \varepsilon$. Ceci entraine que $\gamma^{-1/2}(R_t^{-1}x_t - D^{-1}x_t) \xrightarrow{P} 0$ et $\gamma^{-1/2}(\hat{\theta}_t - \theta_t) \sim D^{-1}x_t$. ∎

Notons que si $\mu = \gamma$, la matrice de covariance de l'erreur normalisée $V = (\sigma_e^2 D^{-1} + R_w)/2$ diffère de la valeur optimale (17). Cette méthode n'est donc pas optimale.

3.4 LMS stabilisé

$\hat{\theta}_t$ satisfait l'équation

$$\hat{\theta}_t = \hat{\theta}_{t-1} + \mu\Gamma_t\varphi_t(y_t - \varphi_t^T\hat{\theta}_{t-1}) \qquad (20)$$

$$\Gamma_t = (\mu\rho^{-1}R_t + \Gamma^{-1})^{-1}, \ \Gamma = \Gamma^T > 0 \qquad (21)$$

$$R_t = (1-\rho)R_{t-1} + \rho\varphi_t\varphi_t^T, \quad R_0 = 0. \qquad (22)$$

Ici $0 < \rho < 1$ et $\mu > 0$ sont les parametres de l'algorithme. Quand $\mu \to 0$ les équations (20) - (22) peuvent être vues comme une version régularisée du LMS (11) avec le gain constant Γ et également, comme une version régularisée de l'algorithme avec facteur d'oubli. L'avantage de cette méthode est la stabilité dans L_p pour toutes valeurs de μ et ρ sous certaines conditions d'excitation (cf. [4]).

Théorème 5 *Sous B1-B3, l'algorithme (20) - (22) satisfait*

$$\hat{\theta}_t - \theta_t \asymp \mathcal{N}(0, \mu V_1 + V_2\gamma^2/\mu)$$

où V_1 et V_2 sont définies par (13) et (14).

Demonstration: Posons $P_t^0 = \Gamma$; alors $\Gamma_t = P_t^0 + P_t'$ où

$$P_t' = -\Gamma(\Gamma + \mu^{-1}\rho R_t^{-1})^{-1}\Gamma,$$

et

$$|P_t'| \leq |\Gamma|^2 |(\Gamma + \mu^{-1}\rho R_t^{-1})^{-1}| \leq |\Gamma|^2\mu\rho^{-1}|R_t|.$$

Introduisons la version stationnaire de la suite (R_t):

$$S_0 = \rho\sum_{j=-\infty}^{0}(1-\rho)^{-j}\varphi_j\varphi_j^T$$

$$S_t = (1-\rho)S_{t-1} + \rho\varphi_t\varphi_t^T$$

Remarquons que

$$S_t - R_t = (1-\rho)(S_{t-1} - R_{t-1})$$

et $R_0 = 0$ impliquent $R_t \leq S_t$; donc le choix

$$p_t' = |\Gamma|^2\mu\rho^{-1}|S_t|$$

satisfait [A5]. Le démonstration se termine comme pour le théorème 2. ∎

Puisque l'expression de la covariance asymptotique de l'erreur est la même que dans le théorème 2, le corollaire 1 reste vrai pour l'algorithme (20) - (22).

References

[1] A. BENVENISTE, M. METIVIER and P. PRIOURET *Adaptive Algorithms and Stochastic Approximations* Springer-Verlag, Berlin, 1990.

[2] S. BITTANTI and M. CAMPI Adaptive RLS algorithms under stochastique excitation - L^2 convergence analysis. *IEEE Trans. Automatic Control*, AC-36, N. 8, 963-967, 1991.

[3] L.GUO, L. LJUNG and P. PRIOURET Performance Analysis of Forgetting factor RLS, *Proceedings of CDC'31*, Tucson, Arizona, December 1992.

[4] A. JUDITSKY and P. PRIOURET A robust algorithm for random parameter tracking. Submitted for publication.

[5] H. KUSHNER, H. HUANG, Asymptotic properties of stochastique approximations with constant coefficients, *SIAM J. Control, Optim*, 19, 87-105, 1981.

[6] L.LJUNG, S.GUNNARSON *Adaptation and Tracking in System Identification-A Survey*, Automatica, Vol.26, No. 1, pp. 7-21,1990.

[7] L.LJUNG *Optimal and ad-hoc adaptation mechanisms*, Linkoping University Internal Repport No.LITH-ISY-I-1181.

[8] L.LJUNG *An Exact Formula for the Tracking Ability of Adaptive Algorithms* Linkoping University Internal Repport No.LITH-ISY-I-1189.

[9] L.LJUNG, P.PRIOURET *A Result on the Mean Square Error Obtained Using General Tracking Algorithm*, Int.J. of Adaptive Control and Signal Processing,Vol. 5, pp.231-250, 1992.

[10] L.LJUNG, P.PRIOURET *Remarks on the Mean Square Tracking Error*, Int.J. of Adaptive Control and Signal Processing,Vol. 6,1992.

[11] A. JUDITSKY, A. NAZIN, Optimal and robust estimation of slowly drifting parameters of linear regression, *Avtomatika i Telemekhanika*, N 6, 66-75, 1991 (in Russian). To be translated as *Automation & Remote Control.*

[12] O. MACCHI and E. EWEDA Second order convergence analysis of stochastique adaptive linear filtering *IEEE Trans. Automatic Control*, AC-28, 76-85, 1983.

[13] M. NIEDŹMIEWCKI and L. GUO Nonasymptotic results for finite-memory WLS filters.

[14] B. WIDROW, J. Mc COOL, M.G. LARIMORE, C.R. JOHNSON Stationary and nonstationary learning characteristics of the LMS adaptive filter, *Proc. of the IEEE*, 64, N. 8, 1151-1161, 1976.