

# HUITIEME COLLOQUE SUR LE TRAITEMENT DU SIGNAL ET SES APPLICATIONS

NICE du 1<sup>er</sup> au 5 JUIN 1981

---

## AUTO-ADAPTATIVITE DES ALGORITHMES STOCHASTIQUES RECURSIFS

Dominique PERRIOT-MATHONNA

THOMSON C.S.F. DSE/SILS 1 rue des Mathurins  
92223 BAGNEUX

---

### RESUME

On présente ici une étude relative à l'auto-adaptativité des algorithmes récursifs stochastiques tels que les algorithmes d'approximation stochastique. Comme Ljung [1] l'a montré et comme on l'a présenté dans [2] l'utilisation d'une suite décroissante de gains d'adaptation permet la convergence avec probabilité 1. Par contre, la vitesse de convergence et donc la capacité d'auto-adaptation décroît vers zéro : on conçoit que, si le vecteur de paramètres poursuivis est soumis à des sauts d'amplitude finie, ce type d'algorithme ne constitue pas une solution efficace dans ce cadre non stationnaire.

Ces différents phénomènes ainsi que les notations utilisées sont rappelés dans une première partie.

La seconde partie présente deux méthodes bien connues permettant de conférer une certaine capacité d'auto-adaptation à ces algorithmes : limitation du pas d'adaptation, pondération des mesures. On montre que malheureusement, il n'y a plus que convergence en probabilité.

Dans la troisième partie, on présente une méthode plus originale consistant à détecter les sauts sur les paramètres poursuivis et, en cas de détection, à réinitialiser la suite des gains d'adaptation : en utilisant des approximations diffusion, on donne une méthode de détermination des nouveaux paramètres de cet algorithme.

Enfin, on présente quelques résultats comparatifs obtenus à partir des méthodes proposées.

### SUMMARY

This paper presents some improvements in recursive stochastic algorithms such as the stochastic approximation algorithms achieved by making them self adaptive.

As proved by Ljung in [1] and presented in [2], the use of a decreasing scalar adaptation gain sequence yields an algorithm which converges with probability one to the desired value ; on the other hand, the convergence rate approaches zero. If the true parameters are perturbed by a sequence of discontinuities, the algorithm becomes uneffective.

This problem is briefly reviewed in the first part of the paper and the notations are introduced.

In the second part, two almost classical methods are presented : step size limitation and measurement weighing. These techniques are asymptotically equivalent and their convergence properties are discussed.

The third part is devoted to a more original method which consists in detecting a possible discontinuity in the estimated parameter and, when a detection occurs, in re-initializing the scalar adaptation gain sequence. New parameters are requested for this algorithm : they are obtained by the use of diffusion approximations.

Finally, some results obtained by these methods are presented.



## AUTO-ADAPTATIVITE DES ALGORITHMES STOCHASTIQUES RECURSIFS

### 1 - POSITION DU PROBLEME

Considérons le système linéaire discret décrit

par :

$$\left. \begin{aligned} x_{n+1} &= (F+G\theta_0 H) x_n + (B+G\theta_0 T) u_n + Ky_n \\ y_n &= \theta_0 H x_n + \theta_0 T u_n + v_n \end{aligned} \right\} (1)$$

Avec :

- $y_n$  : vecteur de sortie du système (dimension p)
- $u_n$  : vecteur d'entrée du système (dimension m)
- $v_n$  : vecteur de bruit sur la sortie ( " p)
- $x_n$  : vecteur d'état du système (dimension  $n_0$ )
- $\theta_0$  : matrice rassemblant les paramètres inconnus du système (dimension pxs).

Les matrices F,G,H,B,K,T sont connues et de dimensions compatibles.

La matrice  $\theta_0$  est estimée récursivement par l'algorithme suivant :

$$\hat{x}_{n+1} = (F+G\theta_n H) \hat{x}_n + (B+G\theta_n T) u_n + Ky_n \quad (2)$$

$$\hat{y}_n = \theta_{n-1} H \hat{x}_n + \theta_{n-1} T u_n \quad (3)$$

$$\theta_n^T = \theta_{n-1}^T + \gamma_n R_n^{-1} (H\hat{x}_n + Tu_n) (y_n - \hat{y}_n)^T \quad (4)$$

$$R_n = R_{n-1} + \gamma_n ((H\hat{x}_n + Tu_n) (H\hat{x}_n + Tu_n)^T - R_{n-1}) \quad (5)$$

Sous certaines hypothèses concernant les propriétés stochastiques de  $\{y_n, u_n\}_{n \in \mathbb{N}}$ , la régularité du terme dit d'actualisation dans (4) et le comportement de la suite scalaire  $\{\gamma_n\}_{n \in \mathbb{N}}$ , Ljung [1] prouve que (4) et (5) génèrent une suite  $\{\theta_n\}_{n \in \mathbb{N}}$  qui converge avec probabilité 1 vers  $\theta_0$  quand n tend vers l'infini, sous une condition bien connue de réelle positivité. Parmi ces conditions, la suite  $\{\gamma_n\}_{n \in \mathbb{N}}$  doit tendre vers 0 et nous supposons sans restriction que :

$$\gamma_n = \frac{1}{n}$$

Supposons maintenant que  $\theta_0$  soit soumise à des sauts d'amplitude finie supposée connue  $\Delta\theta_0$ , la période moyenne entre les sauts étant notée  $T_0$ . Il est alors évident que pour n suffisamment grand, les estimés successifs sont pratiquement constants et cet algorithme adaptatif ne sait pas s'auto-adapter pour poursuivre les variations de  $\theta_0$ .

### 2 - LA LIMITATION DU PAS ET LA PONDERATION DU TERME D'ACTUALISATION

On peut examiner rapidement deux méthodes fréquemment utilisées dans ces types d'algorithmes :

- la limitation du pas consistant à choisir la suite  $\{\gamma_n\}_{n \in \mathbb{N}}$  comme suit :

$$\gamma_n = \begin{cases} \frac{1}{n} & n < N \\ \frac{1}{N} & n \geq N \end{cases} \quad (6)$$

- la pondération du terme d'actualisation consistant à remplacer (5) par :

$$R_n = \frac{1}{n} \sum_{j=0}^{n-1} p^{n-1-j} v_n v_n^T \quad 0 < p \leq 1 \quad (7)$$

en notant :

$$v_n \equiv H\hat{x}_n + Tu_n$$

Notons que dans (5), toutes les mesures ont le même poids (p=1) alors que dans (7), elles ont un poids décroissant avec leur âge.

#### 2.1 - Equivalence asymptotique des deux méthodes

On montre facilement que l'utilisation de (7), au lieu de (5) revient à utiliser dans (4) et (5) une suite  $\{\gamma_n\}_{n \in \mathbb{N}}$  définie par :

$$\gamma_n = \frac{1-p}{1-p^n} \quad 0 < p < 1 \quad (= \frac{1}{n} \text{ pour } p = 1) \quad (8)$$

La comparaison entre (6) et (8) est alors immédiate :

- (6) décroît rapidement et reste saturée à  $1/N$

- (8) décroît plus lentement vers la même valeur  $1/N$  pourvu que :

$$N = \frac{1}{1-p} \quad (9)$$

AUTO-ADAPTATIVITE DES ALGORITHMES STOCHASTIQUES RECURSIFS

Les deux techniques conduisent donc à des comportements transitoires différents sur  $\{\theta_n\}_{n \in \mathbb{N}}$  mais au même régime permanent si  $p$  et  $N$  sont liés par (9) : la suite scalaire de gains ne tend plus vers zéro.

2.2 - Les propriétés de convergence

Du fait de la remarque précédente, on ne peut plus utiliser les résultats de Ljung et on sent bien que les propriétés de convergence avec probabilité 1 ne sont pas conservées; par contre, on peut s'attendre à certaines propriétés de convergence en probabilité avec une borne supérieure de la variance du bruit d'estimation fonction de  $p$  ou de  $N$

L'idée est alors d'utiliser un théorème dû à Khasminskii [3] pour décrire l'écart entre :

- les estimés successifs générés par (4) et (5) où  $\gamma_n$  est remplacé par  $\tau$  taille limite du pas d'adaptation,
- les trajectoires déterministes moyennes solutions des équations différentielles

$$\frac{d}{dt} \theta^T(t) = R^{-1}(t) h(\theta(t)) \quad (10)$$

$$\frac{d}{dt} R(t) = G(\theta(t)) - R(t) \quad (11)$$

avec :

$$h(\theta) = E \left\{ v_n (y_n - \hat{y}_n)^T \right\}$$

$$G(\theta) = E \left\{ v_n v_n^T \right\}$$

$\theta$  étant une valeur fixe quelconque du paramètre.

On introduit la correspondance entre temps discret de (4) et (5) et temps continu de (10) et (11) par :

$$t \sim n\tau$$

Alors sous les conditions convenables (Cf. [3]), le processus linéaire par morceaux

$$y_t^\tau = \frac{1}{\sqrt{\tau}} (\theta_n^T - \theta^T(n\tau))$$

converge (au sens de la convergence faible) quand  $\tau \rightarrow 0$  vers la solution de l'équation différentielle stochastique

$$dy_t = R^{-1} \frac{d}{d\theta} [h(\theta(t))] y_t dt + R^{-1} c(\theta(t)) d\beta_t \quad (12)$$

$$y_0 = 0$$

Avec :

$(\beta_t)_{t \geq 0}$  est le brownien standard

$C(\theta)$  matrice symétrique définie par :

$$C(\theta) C(\theta)^T = \sum_n E \left\{ \left[ v_n (y_n - \hat{y}_n)^T - h(\theta) \right] \left[ v_0 (y_0 - \hat{y}_0)^T - h(\theta) \right]^T \right\}$$

D'autre part, après une période transitoire, on peut espérer que  $\theta_n$  sera suffisamment proche de la valeur vraie  $\theta_0$  pour que l'on puisse remplacer (12) par sa version linéarisée :

$$dy_t = \left[ R^{-1} \frac{d}{d\theta} h(\theta) \right]_{\theta=\theta_0} y_t dt + \left[ R^{-1} C(\theta) \right]_{\theta=\theta_0} d\beta_t$$

$$= F y_t dt + G d\beta_t$$

en supposant que toutes les valeurs propres de  $F$  sont dans le demi-plan gauche.

Dans le temps de l'expérience  $s \sim n$ , le modèle pour le processus linéaire par morceaux

$$x_s^\tau = \theta_n^T - \theta_0^T \quad (13)$$

peut donc être décrit par :

$$dx_s^\tau = \tau F x_s^\tau ds + \tau G d\beta_s \quad (14)$$

On considérera dorénavant que (14) est un bon modèle pour décrire l'évolution de (13) quand  $\tau$  est suffisamment petit. L'intégration de (14) au sens de Itô ou de Stratonovitch produit :

$$x_s^\tau = e^{\tau F(s-s_0)} x_0 + \int_0^s e^{\tau F(s-u)} \tau G d\beta_u$$

qui permet d'établir que :

$$- \lim_{s \rightarrow \infty} E \left\{ x_s^\tau \right\} = 0 \Rightarrow \lim_{s \rightarrow \infty} E \left\{ \theta_s \right\} = \theta_0$$

$$- \lim_{s \rightarrow \infty} E \left\{ x_s^\tau x_s^\tau \right\} = \tau K$$

où  $K$  est une matrice liée à  $F$  et  $G$  indépendante de  $\tau$ .

Ce résultat est bien celui escompté.



#### 4 - LA DETECTION DES SAUTS ET LA REINITIALISATION DE LA SUITE SCALAIRE

Dans le paragraphe précédent, on a évoqué deux méthodes où la taille du pas d'itération n'est jamais nulle ce qui confère à l'algorithme ainsi modifié une certaine capacité d'auto-adaptation au détriment du mode de convergence. On propose ici une méthode qui paraît potentiellement plus puissante en ce sens, qu'entre deux sauts, on garde la propriété de convergence avec probabilité 1.

##### 4.1 - L'idée de base

On se propose donc de détecter les sauts sur le paramètre poursuivi et, en cas de détection, de réinitialiser la suite scalaire  $\{Y_n\}_{n \in \mathbb{N}}$  à une valeur arbitraire comme s'il s'agissait d'un nouveau problème avec de nouvelles conditions initiales. Il s'agit donc d'une utilisation répétée de l'algorithme original avec conservation des propriétés de convergence. Ceci nécessite donc d'utiliser un algorithme de détection en choisissant le signal à tester, le seuil du test et de déterminer conjointement la valeur  $1/N_0$  à laquelle la suite  $\{Y_n\}_{n \in \mathbb{N}}$  sera réinitialisée.

##### 4.2 - Le signal à tester

Ce signal doit contenir l'information sur l'occurrence des sauts.

Il est clair que le terme d'actualisation dans (4)

$$Q_n = v_n (Y_n - \hat{Y}_n)^T$$

est, par construction, le gradient d'un critère quadratique sur l'erreur sortie-sortie du modèle et que  $R_n^{-1}$  procure une correction au second ordre de sorte que :

$$E \{Z_n\} = E \{R_n^{-1} Q_n\} = 0 \iff \theta_n = \theta_0$$

En conséquence, quand  $Z_n$  n'est pas à moyenne nulle, on peut constater une désadaptation de l'algorithme.

##### 4.3 - La méthode de détection

Il s'agit donc de détecter un changement de moyenne sur  $Z_n$  et plusieurs algorithmes sont utilisables dans ce but [4]. On choisit l'algorithme d'Hinkley [5] que l'on peut décrire comme suit : considérons une suite

$$\{Z_n\}_{n \in \mathbb{N}} \text{ telle que :}$$

$$E \{Z_n\} = \begin{cases} \mu_0 & v_n \leq N \\ \mu_1 & v_n > N \end{cases}$$

on évalue :

$$S_n = \sum_{j=0}^n (Z_j - \frac{\mu_0 + \mu_1}{2}) = \sum_{j=0}^n (Z_j - \Delta\mu)$$

$$M_n = \max_{0 < k \leq n} S_k \quad m_n = \min_{0 < k \leq n} S_k$$

$$\Delta_n = M_n - S_n \quad \delta_n = S_n - m_n$$

et on compare pour détecter le changement de moyenne à un seuil  $\Delta$  fixé :

$$\Delta_n \stackrel{?}{\geq} \Delta \quad \delta_n \stackrel{?}{\geq} \Delta$$

Le test sur le maximum permet de détecter une diminution de moyenne et le test sur le minimum ( $\delta_n$ ) une augmentation de moyenne.

##### 4.4 - L'algorithme et ses contraintes

L'algorithme consiste donc à mettre en oeuvre les deux détecteurs d'Hinkley présentés précédemment et, sur l'occurrence d'une détection, à réinitialiser la suite scalaire à la valeur  $1/N_0$ .

Il faut préciser certains paramètres pour la mise en oeuvre :

-  $\Delta\mu$  : cette valeur dépend de la taille des sauts  $\Delta\theta_0$ . Pour  $T_0$  suffisamment grand, l'algorithme sera en régime permanent avant le saut ( $\mu_0 = 0$ ) ; aussi,  $\Delta\mu$  représentera la perturbation minimale que l'on veut détecter,

-  $\Delta, N_0$  : le problème est plus délicat car  $N_0$  régit la vitesse de convergence et le bruit d'estimation, alors que  $\Delta$  agit directement sur le retard à la détection.

Par souci de simplicité, nous nous limiterons dans ce qui suit au cas scalaire.

AUTO-ADAPTATIVITE DES ALGORITHMES STOCHASTIQUES RECURSIFS

4.5 - La détermination de  $\Delta$  et  $N_0$

Ces deux paramètres sont à ajuster conjointement en vue de minimiser le retard à la détection et le nombre de fausses alarmes et sur un temps  $T \gg T_0$ , on peut minimiser :

$$J(\Delta, N_0) = \begin{cases} \text{énergie perdue sur une fausse} \\ \text{alarme} \times \text{nombre moyen de faus-} \\ \text{ses alarmes} \\ + \\ \text{énergie perdue sur un saut} \times \\ \text{nombre moyen de sauts} \end{cases}$$

Ces calculs nécessitent une modélisation du comportement de l'écart  $(\theta_n - \theta_0)$  et du processus de détection : on peut pour cela, utiliser l'approximation-diffusion établie au paragraphe 3 en introduisant la taille du pas

$$\tau = 1 / (t + N_0)$$

auquel cas, la solution de l'équation différentielle stochastique s'écrit :

$$X_s = \left(\frac{N_0}{s + N_0}\right)^{-F} \cdot X_0 + \frac{G}{(s + N_0)^{-F}} \int_0^s \frac{d\beta t}{(t + N_0)^{1+F}}$$

- Energie perdue sur un saut :

On ne s'intéresse qu'à la partie déterministe avec  $X_0 = \Delta\theta_0$  ; aussi

$$E_{\text{saut}} = \int_0^{T_0} X_s^2 ds = \Delta\theta_0^2 \int_0^{T_0} \left(\frac{N_0}{s + N_0}\right)^{-2F} ds$$

- Energie perdue sur une fausse alarme :

On suppose qu'à ce moment on est en régime permanent et seule la partie aléatoire est intéressante ; l'énergie cherchée vaut :

$E_{F.A.}$  = énergie due à l'erreur  $\Delta\theta$  du fait de la fausse alarme

- énergie due à l'erreur  $\Delta\theta$  s'il n'y avait pas eu d'alarme.

Considérons un instant quelconque noté  $T_1$  alors :

$$E_{F.A.} = E \left\{ \int_{T_1}^{T_1+T} X_s^2 ds / \text{réinitialisation à } t = T_1 \right\}$$

$$- E \left\{ \int_{T_1}^{T_1+T} X_s^2 ds / \text{pas de réinitialisation à } t = T_1 \right\}$$

En fait,  $T_1$  peut être interprété comme le temps moyen entre fausses alarmes.

D'autre part, on peut limiter l'intervalle d'intégration à la période  $T_1$  entre deux fausses alarmes auquel cas :

$$E_{F.A.} = E \left\{ \int_0^{T_1} X_s^2 ds - \int_{T_1}^{2T_1} X_s^2 ds \right\} \\ = \int_0^{T_1} \frac{G^2}{(s + N_0)^{-2F}} \left( \int_0^s \frac{dt}{(t + N_0)^{2+2F}} \right) ds - \int_{T_1}^{2T_1} \frac{G^2}{(s + N_0)^{-2F}} \left( \int_0^s \frac{dt}{(t + N_0)^{2+2F}} \right) ds$$

- Temps moyen entre fausses alarmes, retard à la détection

Ces calculs reposent sur la modélisation du signal de test  $Z_n$ . On propose ici une approximation-diffusion qui permet, par les théorèmes d'arrêt, d'accéder aux paramètres recherchés.

Pendant un intervalle de temps suffisamment petit  $\Delta s$ , le pas est supposé de taille constante, l'accroissement du processus est alors :

$$\Delta S_s = S_{s+\Delta s} - S_s = \frac{\theta_{s+\Delta s} - \theta_s}{\gamma_s} - \Delta\mu \Delta s \\ = \frac{X_{s+\Delta s} - X_s}{\gamma_s} - \Delta\mu \Delta s$$

d'où la caractérisation du processus  $\{S_s\}_{s \in \mathbb{R}^+}$  à partir de celle de  $\{X_s\}_{s \in \mathbb{R}^+}$ .

Pour  $s$  suffisamment grand, l'état initial  $X_0$  connu voisin de zéro, alors :

$$dS_t \sim -\Delta\mu dt + G d\beta t$$

- Le temps moyen entre fausses alarmes peut être calculé à partir des théorèmes d'arrêt sur les browniens [6] et ici, on aura :

$$T_1 = \frac{G^2}{2\Delta\mu^2} \left[ e^{\frac{2|\Delta\mu|\Delta}{G^2}} - 1 + \frac{2|\Delta\mu|\Delta}{G^2} \right]$$

- Le retard moyen à la détection se calcule comme un temps moyen de sortie pour la diffusion considérée : la densité de probabilité de ce temps partant de  $S_0$  pour sortie par  $S_0 + \Delta$  vérifie l'équation de Kolmogorov et on peut déduire que :

$$R_D = \frac{\Delta}{|\Delta\mu|}$$



- Le critère à minimiser

On peut alors l'écrire sous la forme suivante:

$$J [A, N_0] = E_{F.A.} \frac{T}{T_1} + (E_{\text{saut}} + R_D) \cdot \frac{T}{T_0}$$

compte tenu des expressions établies précédemment.

La minimisation de ce critère par rapport à  $\Delta$  et  $N_0$  permet d'en déduire les valeurs de ces paramètres.

5 - UN EXEMPLE D'APPLICATION

Cet exemple concerne l'algorithme de filtrage adaptatif de Hampton [7] dont les propriétés ont été étudiées en détail dans [8]. Cet algorithme permet d'estimer la matrice de gain d'un filtre de Kalman quand celui-ci est implémenté sans connaissance des matrices de covariance des bruits de mesure et d'état. Les équations (4) et (5) s'écrivent alors dans le cas scalaire :

$$K_{n+1} = K_n + \frac{1}{n+1} (H\Phi)^{-1} v_{n+1} v_n^T \omega_{n+1}$$

$$\omega_{n+1}^{-1} = \omega_n^{-1} + \frac{1}{n+1} v_n v_n^T$$

où  $H$  et  $\Phi$  sont respectivement les matrices d'observation et d'évolution du système sur lequel agit le filtre,  $v_n$  le résidu en sortie du filtre. L'adaptation du gain repose donc sur un critère de corrélation de ces résidus.

On présente ici trois cas de mise en oeuvre du filtre. Pour ces 3 cas :

$$H = 1 \quad \Phi = 0,99$$

Les matrices  $R$  et  $Q$  sont parfaitement inconnues et on prend  $K_0 = 0$ . Le paramètre optimal  $K_{\text{opt}}$  est soumis à des sauts définis par :

$$\Delta K_{\text{opt}} = \Delta \theta_0 = 0,2 \quad T_0 = 500$$

La planche n° 1 présente le cas d'utilisation standard du filtre : on remarque que l'adaptation est peu sensible : cependant pour  $1000 \leq n < 1500$ , le mauvais fonctionnement du filtre est également dû au peu de sensibilité du critère quand le gain actuel est plus grand que le gain optimal.

La planche n° 2 présente l'évolution des estimés avec une pondération introduite dans le filtre ( $p = 0,996$  ce qui correspond à  $N = 250$ ). On note le bruit résiduel important.

La planche n° 3 présente la mise en oeuvre de l'algorithme décrit au paragraphe 4. Les paramètres sont :

- cas d'augmentation du gain

$$\Delta\mu = 0,15 \quad \Delta = 15 \quad N_0 = 6$$

- cas de diminution du gain

$$\Delta\mu = 0,07 \quad \delta = 20 \quad N_0 = 15$$

On peut noter que les augmentations de moyenne sont bien détectées. Par contre, on a une fausse alarme à  $n = 920$  et non-détection de la diminution de moyenne : enfin dans cette zone, malgré la valeur encore importante de  $\gamma_n$ , l'adaptation est peu sensible.

6 - CONCLUSIONS

Cette étude consacrée à l'auto-adaptativité des algorithmes stochastiques récursifs a permis de dégager deux approches possibles :

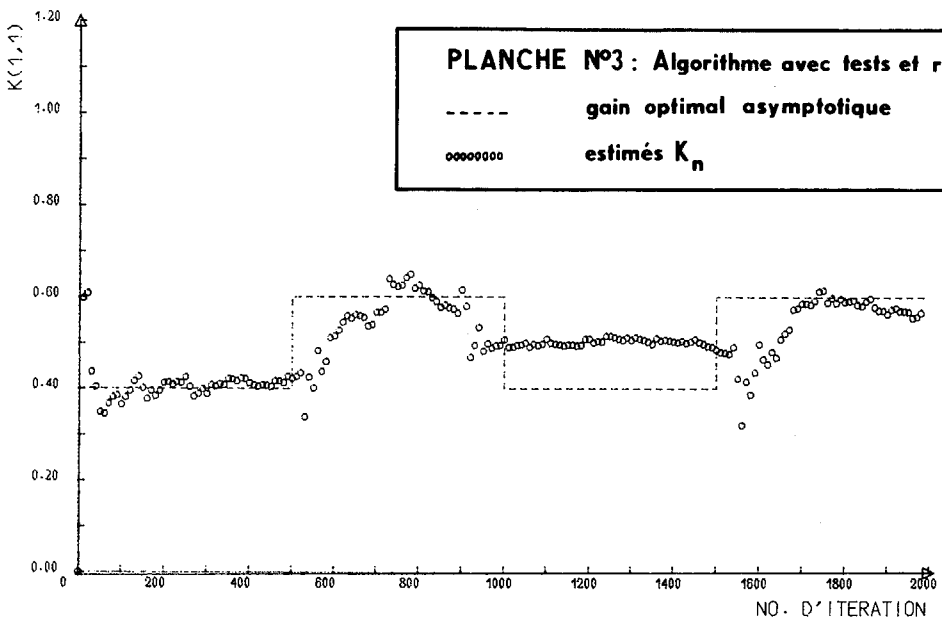
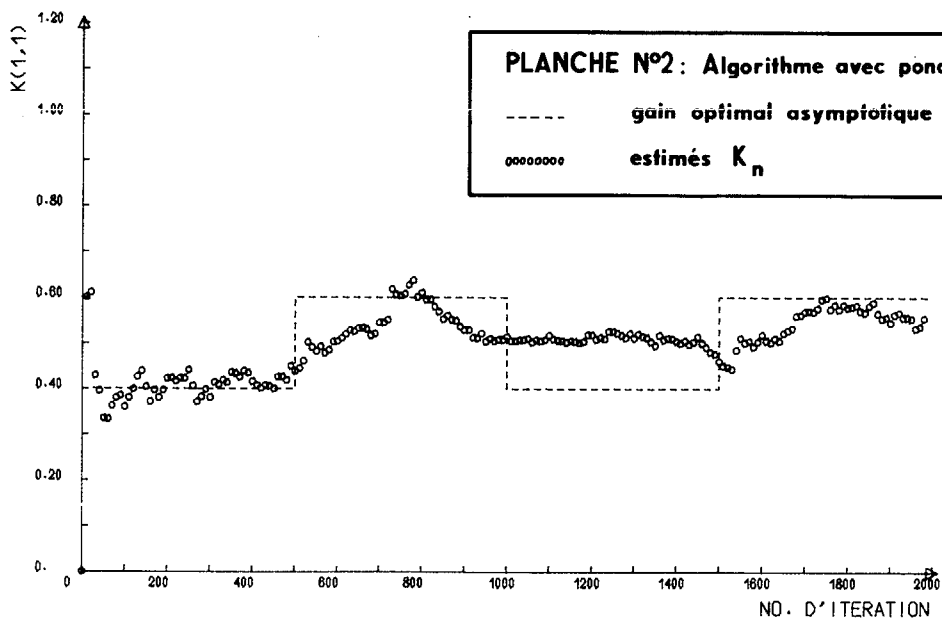
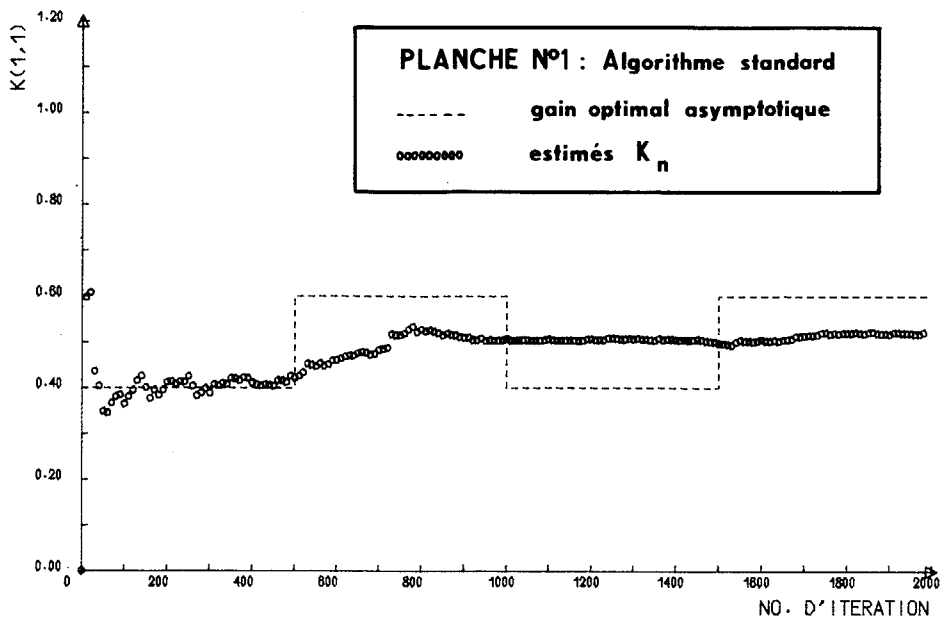
- la pondération des mesures qui ne permet pas de conserver les bonnes propriétés de convergence,
- une méthode basée sur la réinitialisation de la suite scalaire de gains d'adaptation quand un saut sur le paramètre poursuivi est détecté. La seconde méthode nécessite le calcul de paramètres supplémentaires ce qui a été fait en modélisant le signal de test par une approximation-diffusion utilisée pour modéliser le temps moyen entre fausses alarmes et retard à la détection.

Il faut cependant noter que cette approximation peut conduire à des valeurs optimistes du temps moyen entre fausses alarmes en particulier pour des dérives importantes et des seuils très hauts : l'approximation de l'accroissement de la variété mélangeante  $\sum v_k v_{k+1}$  par un brownien devient grossière quand les probabilités de fausses alarmes sont faibles ( $\leq 10^{-4}$ ).

Dans ce cas, l'estimation du temps moyen entre fausses alarmes ne peut se faire que par des approches probabilistes plus fines.

Malgré les approximations utilisées, les exemples proposés montrent l'intérêt de la méthode exposée ici en particulier le fait que les propriétés de convergence avec probabilité 1 sont conservées.

AUTO-ADAPTATIVITE DES ALGORITHMES STOCHASTIQUES RECURSIFS





## AUTO-ADAPTATIVITE DES ALGORITHMES STOCHASTIQUES RECURSIFS

BIBLIOGRAPHIE

- [1] L. LJUNG, "Analysis of recursive stochastic algorithms", IEEE Trans. on Automat Contr. , Vol. AC-22, Août 1977.
- [2] D. PERRIOT-MATHONNA, "On the use of Ljung's results for studying the convergence properties of Hampton's adaptive filter" IEEE Trans. on Automat. Contr., Vol AC-25, Décembre 1980
- [3] R.Z. KHASMINSKII, "On stochastic processes defined by differential equations with a small parameter" Theory of Prob. and Appl., Vol 11, 1966
- [4] M. BASSEVILLE, B. ESPIAU,  
"Détection de contours dans une image numérisée",  
Publ. interne n° 109, IRISA  
Université de Rennes.
- [5] D.V. HINKLEY, "Inference about the change-point from cumulative sum tests", Biometrika, Vol 58, 1971.
- [6] H.M. TAYLOR, "A stopped brownian motion formula", The Annals of Prob., Vol 3, n° 2, 1975
- [7] R.L.T. HAMPTON "Stochastic algorithms for self-adaptive filtering and prediction", NASA O3.OO2.OO6 Semi-annual Report - Part A, N71-34180, 1971
- [8] D. PERRIOT-MATHONNA, "Contribution à l'étude et à la mise en oeuvre d'algorithmes de filtrage de Kalman adaptatif",  
Thèse de Docteur-Ingénieur  
Université de Paris-Sud  
n° 385, 1978