

# Convergence des filtres adaptatifs dans le contexte alphabet fini

H. Besbes<sup>(1)(2)</sup>, M. Jaïdane-Saïdane<sup>(1)</sup>, J. Ezzine<sup>(1)</sup>

<sup>(1)</sup> L.S.Télécoms, ENIT, Campus Universitaire, Le Belvédère, Tunis, TUNISIE

Tél : (216-1) 874-700, Fax : (216-1) 870-729

<sup>(2)</sup> ESPTT, Route de Raoued, Km3.5 CitÉ El Ghazala 2083, Tunis, TUNISIE

Tél : (216-1) 762-000, Fax : (216-1) 762-819

<sup>(3)</sup> ENIT, Campus Universitaire, BP 37 – 1002 Tunis, TUNISIE

Tél : +216 (1) 874 – 700, Fax : +216 (1) 870 – 729

## RÉSUMÉ

Dans ce papier, nous étudions la convergence en moyenne quadratique des filtres adaptatifs, pilotés par l'algorithme du gradient stochastique (GS). Le contexte où l'entrée appartient à un alphabet fini est considéré ; il correspond au contexte réel de transmission numérique. Nous proposons une approche mathématique adaptée à ce contexte, qui permet une analyse simple et rigoureuse, de convergence et ce, sans passer par des hypothèses simplificatrices classiques de type "indépendance" des vecteurs observations.

## ABSTRACT

This paper gives a tailored approach for convergence analysis of adaptive filters governed by the LMS algorithm. The case of finite alphabet input is considered, it corresponds to the real digital transmission context. A complete and rigorous proof of quadratic mean convergence is given without any constraining independence assumptions on the observation vector.

## 1 Introduction : difficulté de l'analyse de convergence

Du fait de sa faible complexité, l'algorithme du gradient stochastique (GS), est très utilisé dans les applications de transmission numérique du signal : annulation d'écho, codage prédictif, égalisation des canaux... La mise à jour des coefficients  $H_k$ , d'un filtre transverse adaptatif piloté par le GS est donnée par :

$$H_{k+1} = H_k + \mu(y_k - X_k^T H_k)X_k. \quad (1)$$

où  $X_k = (x_k, x_{k-1}, \dots, x_{k-L+1})^T$  est le vecteur de l'observation de l'entrée de longueur  $L$ ,  $\mu > 0$  est le pas d'adaptation,  $y_k = F^T X_k + b_k$  est la sortie bruitée du système (ou canal) de vecteur paramètre  $F$  et  $b_k$  est le bruit d'observation. On supposera les séquences  $\{b_k\}$  et  $\{x_k\}$  stationnaires, indépendantes entre elles et centrées.

Le comportement transitoire du filtre adaptatif est décrit par l'évolution du vecteur déviation défini par  $V_k = H_k - F$  et caractérisant l'écart entre le filtre adaptatif et le filtre optimal. Ce vecteur suit la récurrence suivante :

$$V_{k+1} = (I - \mu X_k X_k^T)V_k + \mu b_k X_k. \quad (2)$$

L'analyse de convergence en moyenne quadratique du GS consiste à étudier le comportement transitoire de  $E(V_k V_k^T)$ . La difficulté essentielle de cette analyse provient du fait que  $X_k$  et  $V_k$  sont dépendants. Ainsi, on ne peut trouver de relation simple entre  $E(V_{k+1} V_{k+1}^T)$  et  $E(V_k V_k^T)$ .

De nombreuses approches ont été proposées pour pallier à cette difficulté :

- Les approches directes consistent à étudier le comportement de la matrice de transition du GS [1, 2].
- Les approches indirectes se basent sur des méthodes d'approximations stochastiques, telles que la méthode de l'ODE [3, 4], et la méthode des perturbations [5].
- Les approches simplificatrices : elles utilisent des modèles particuliers de l'entrée [6].

Les études menées à partir de ces différentes approches conduisent à des analyses mathématiques élaborées selon quatre types d'hypothèses dont seules les trois premières sont réalistes :

- $H1$  Ergodicité de l'entrée.
- $H2$  Excitation persistante de l'algorithme, qui peut se traduire par l'inversibilité de la matrice de covariance de l'entrée.
- $H3$  Bornitude de l'entrée ou de ses moments.

C'est de l'hypothèse ( $H4$ ), relative à la relation entre  $X_k$  et  $V_k$ , que dépend la qualité de l'analyse de convergence. On fait ainsi souvent l'hypothèse irréaliste d'indépendance entre  $X_k$  et  $V_k$ , ou de manière plus réaliste on suppose la  $M$ -indépendance du vecteur observation de l'entrée [1]. Moins l'hypothèse est restrictive, plus les développements mathématiques deviennent relativement complexes.

Dans le contexte de transmission numérique, lorsque l'entrée appartient à un alphabet fini, les analyses de convergence sont déduites des résultats issus de ces différentes approches

classiques. Elles ne profitent pas des caractéristiques de l'entrée.

Dans cet article, une nouvelle approche inspirée de [7] adaptée au contexte "alphabet fini", est proposée pour analyser de manière simple la convergence en moyenne quadratique du GS. Aucune hypothèse simplificatrice de type indépendance (H4) n'est utilisée. Une valeur exacte du pas d'adaptation critique est donnée en fonction des statistiques de l'entrée.

## 2 Contexte alphabet fini et analyse de la convergence

Dans le contexte de transmission numérique, les données  $x_k$  appartiennent à un alphabet fini, par exemple  $\{\pm 1\}$  ou  $\{\pm 1, \pm 3\}$ ; par conséquent, les vecteurs  $X_k$  et les matrices  $(I - \mu X_k X_k^T)$  évoluent eux aussi, respectivement dans des alphabets finis  $\{W_1, W_2, \dots, W_N\}$  et  $\{A_1, A_2, \dots, A_N\}$ . Le vecteur observation peut être donc modélisé par une chaîne de Markov irréductible  $\{\theta(k) : k \in \mathbf{Z}^+\}$  définie sur  $\{1, 2, \dots, N\}$  et de matrice de transition  $\mathbf{P} = \{p_{ij}\}$ , telle que :

$$X_k = W_{\theta(k)}. \quad (3)$$

Compte tenu du caractère "alphabet fini" de l'entrée, une nouvelle approche d'analyse du GS, consiste à étudier la convergence des espérances conditionnelles  $Q_j(k)$  définies par :

$$Q_j(k) = E(V_k V_k^T \mathbf{1}_{(\theta(k)=j)}) \quad (4)$$

où  $\mathbf{1}_{(\cdot)}$  est l'indicateur de Dirac.

D'après les équations (1) et (3), la récurrence entre  $Q_j(k+1)$  et  $Q_i(k)$ , peut être développée de la manière suivante :

$$\begin{aligned} Q_j(k+1) &= \sum_{i=1}^N E(V_{k+1} V_{k+1}^T \mathbf{1}_{(\theta(k+1)=j)} \mathbf{1}_{(\theta(k)=i)}) \\ &= \sum_{i=1}^N E(A_{\theta(k)} V_k V_k^T A_{\theta(k)}^T \mathbf{1}_{(\theta(k+1)=j)} \mathbf{1}_{(\theta(k)=i)}) \\ &\quad + \mu^2 \sum_{i=1}^N E(W_{\theta(k)} W_{\theta(k)}^T b_k^2 \mathbf{1}_{(\theta(k+1)=j)} \mathbf{1}_{(\theta(k)=i)}) \end{aligned}$$

L'élégance de la démonstration provient du fait que  $A_{\theta(k)} \mathbf{1}_{(\theta(k)=i)} = A_i \mathbf{1}_{(\theta(k)=i)}$  et que les matrices  $A_i$  sont constantes et indépendantes de  $V_k$ . On peut alors écrire :

$$Q_j(k+1) = \sum_{i=1}^N E(A_i V_k V_k^T A_i^T \mathbf{1}_{(\theta(k+1)=j)} \mathbf{1}_{(\theta(k)=i)}) + z_j(k) \quad (5)$$

$$\text{avec } z_j(k) = \mu^2 \sum_{i=1}^N E(W_i W_i^T b_k^2 \mathbf{1}_{(\theta(k+1)=j)} \mathbf{1}_{(\theta(k)=i)}).$$

L'analyse de la convergence du GS se fait ainsi, simplement, sans l'hypothèse d'indépendance de type H4.

En utilisant les résultats classiques d'algèbre tensorielle, on trouve alors des relations compactes qui régissent l'évolution de  $E(V_k V_k^T)$  [8] :

$$\tilde{Q}(k+1) = \Gamma \tilde{Q}(k) + \tilde{Z}(k) \quad (6)$$

$$\text{où } \tilde{Q}(k) = \begin{bmatrix} \text{vec}(Q_1(k)) \\ \text{vec}(Q_2(k)) \\ \vdots \\ \text{vec}(Q_N(k)) \end{bmatrix}, \tilde{Z}(k) = \begin{bmatrix} \text{vec}(z_1(k)) \\ \text{vec}(z_2(k)) \\ \vdots \\ \text{vec}(z_N(k)) \end{bmatrix}$$

et  $\Gamma = (\mathbf{P}^T \otimes I_{L^2}) \text{diag}(A_i^T \otimes A_i)$ .

$\otimes$  est le produit de Kronecker matriciel,  $\text{vec}(\cdot)$  est l'opérateur qui transforme une matrice en un vecteur et  $I_{L^2}$  est la matrice identité de dimension  $L^2$ . Pour un ensemble de matrices carrées  $D_i, i = 1 \dots N$  de dimension  $l$ , on note  $\text{diag}(D_i)$  la matrice carrée de dimension  $Nl$  formée par les  $D_i$  sur la diagonale et zéro ailleurs. Notons que la matrice  $\Gamma$  est de dimension  $N.L^2$  et que cette dimension devient rapidement élevée lorsque la taille de l'alphabet augmente et l'ordre  $L$  du filtre adaptatif augmente.

Du fait que la matrice  $\Gamma$  est constante et ne dépend que du pas d'adaptation  $\mu$  et des caractéristiques statistiques de l'entrée, la convergence en moyenne quadratique de l'algorithme n'est liée qu'aux valeurs propres de cette matrice.

La proposition suivante résume les résultats trouvés, concernant la convergence du GS.

### Proposition

Dans le contexte "alphabet fini" de l'entrée, et sous les deux hypothèses suivantes :

- H'1 La chaîne de Markov  $\{\theta(k)\}$  est irréductible (c'est l'hypothèse H1 de l'ergodicité de l'entrée).
- H'2 La famille  $\{W_1, \dots, W_N\}$  engendre  $\mathfrak{R}^L$  : (c'est la condition H2 d'excitation persistante du GS).

Il existe un réel  $\mu_c$  supérieur à  $\mu_c^{\max}$  défini par :

$$\mu_c^{\max} = \frac{2}{\max_{i=1 \dots N} (W_i^T W_i)} \quad (7)$$

tel que pour  $\mu < \mu_c$ , la valeur propre maximale de  $\Gamma$  est inférieure à 1, et le GS converge en moyenne quadratique.

## 3 Détermination du pas d'adaptation critique et simulations

Du point de vue pratique, la connaissance du pas critique est cruciale. Les approches classiques basées sur des hypothèses réalistes caractérisant le lien entre  $X_k$  et  $V_k$  ne fournissent pas de borne précise du pas critique.

Notons que souvent on admet pour pas critique  $\frac{2}{\lambda_{\max}}$  où  $\lambda_{\max}$  est la valeur propre maximale de la matrice de corrélation  $E\{X_k X_k^T\}$ . Un tel résultat est pourtant relatif à la convergence en moyenne de  $V_k$  et suppose de plus l'indépendance des vecteurs observations.

En utilisant les statistiques d'ordre 4 de l'entrée et encore sous l'hypothèse d'indépendance, on montre de manière plus précise [1] qu'il existe une borne

$$\mu_c^{\text{ind}} = 2 \cdot \frac{\lambda_{\min}}{\nu_{\max}} \quad (8)$$

où  $\lambda_{min}$  est la valeur propre minimale de  $E\{X_k X_k^T\}$  et  $v_{max}$  est la valeur propre maximale de  $E\{|X_k|^2 X_k X_k^T\}$ , telle que, si  $\mu < \mu_c^{ind}$ , il y a convergence en moyenne quadratique du GS.

La détermination du pas d'adaptation critique  $\mu_c$  par la nouvelle approche, passe par l'estimation de la matrice de transition  $\mathbf{P}$  puis par la recherche de la valeur propre maximale de  $\Gamma$ . Cette matrice dépend à la fois du pas d'adaptation  $\mu$  et des caractéristiques statistiques de l'entrée (eq 6).

Pour valider les résultats obtenus, on a considéré une entrée corrélée appartenant à l'alphabet  $\{\pm 1, \pm 3\}$  et deux exemples d'identification avec des filtre de longueur  $L = 1$  et  $L = 3$ .

1<sup>er</sup> cas : Dans le cas d'identification  $L = 1$ ,  $F = 10$  la valeur de  $\mu_c^{ind} = 10/41$  est indépendante de la corrélation de l'entrée, tandis que le pas d'adaptation critique théorique décroît lorsque la corrélation de l'entrée augmente (fig 1).

Par exemple pour une corrélation de 0.98, le pas critique  $\mu_c = 0.2233 < \mu_c^{ind}$ . On a relevé, dans ce cas, l'évolution de  $E(|V_k|^2)$ , par des simulations de Monte Carlo faites sur 350 réalisations, pour deux valeurs  $\mu_1$  et  $\mu_2$  du pas d'adaptation,  $\mu_c < \mu_1 = 0.24 < \mu_c^{ind}$  et  $\mu_2 = 0.223 < \mu_c$  (fig 2). On constate pour  $\mu_1$ , la divergence exponentielle de  $E(|V_k|^2)$  d'où l'inexactitude de l'approche classique.

2<sup>ème</sup> cas : Dans le cas d'identification  $L = 3$  avec  $F = (1, 0.6, 0.5)^T$ . Notons que  $N = 64$  et que la dimension de la matrice  $\Gamma$  est déjà de  $NL^2 = 576$ . Pour chaque valeur de la corrélation, on a déterminé par simulations les valeurs  $\mu_c^{simu}$  à  $10^{-3}$  près qui provoquent la divergence de  $E\{V_k^T V_k\}$ .

Les résultats de simulation  $\mu_c^{simu}$  sont comparés avec les valeurs théoriques  $\mu_c$ , trouvées par l'approche proposée et les valeurs  $\mu_c^{ind}$  utilisant les hypothèses classiques d'indépendance.

La figure 3 montre l'exactitude de l'approche proposée puisque  $\mu_c$  et  $\mu_c^{simu}$  sont confondus.

Remarquons que lorsque la corrélation de l'entrée est faible, les valeurs de  $\mu_c^{ind}$  et  $\mu_c$  sont effectivement voisines.

Il est intéressant de noter, par ailleurs, que pour les fortes corrélations de l'entrée,  $\mu_c^{ind}$  estimé avec l'hypothèse d'indépendance tend vers 0 pour  $L = 3$ , un tel résultat n'étant d'aucun intérêt du point de vue pratique. Par contre, dans une telle situation,  $\mu_c$  estimé sans cette hypothèse d'indépendance tend vers  $\mu_c^{max}$  (eq 7) qui est différent de 0 et qui de surcroît est une borne qui ne nécessite aucune connaissance sur les statistiques de l'entrée mais simplement une connaissance de l'alphabet transmis.

## 4 Conclusion

Nous avons proposé dans cet article une approche mathématique adaptée à l'analyse des performances des filtres adaptatifs dans un contexte d'"alphabet fini" de l'entrée, contexte rencontré en transmission numérique du signal. Cette approche permet une analyse simple, rigoureuse et complète de la convergence en moyenne quadratique de l'algorithme GS. Aucune hypothèse simplificatrice, telle que l'indépendance ou

la mémoire finie des observations successives n'est utilisée. Une borne précise du pas d'adaptation critique est donnée.

## Références

- [1] O. Macchi, "LMS Adaptive Processing with Applications in Transmission", Wiley, 1994.
- [2] V.Solo, "The Limiting Behavior of the LMS", IEEE trans. ASSP, Vol 37, N 12, pp 1909-1922, December 1989.
- [3] M.A Kourizin, "On Almost-Sure Bounds for the LMS Algorithm", IEEE trans. Information Theory, Vol 40, N 2, pp 372-383, March 1994.
- [4] A. Benveniste, "Design of Adaptive Algorithms for the Tracking of Time-Varying Systems", Int. J. Adaptive Contr. Signal Processing, pp 3-29, 1987.
- [5] A. Perrier, B.Deylon, E.Moulines, "Performance Analysis of Stochastic Gradient Identification Algorithms Using a Perturbation Expansion Method, Part I : Theoretical results", à paraître .
- [6] Dirk T.M. Slock, "On the Convergence Behavior of the LMS and Normalized LMS Algorithms", IEEE trans on ASSP-41, Vol 9, pp 2811-2825, September 1993.
- [7] O.L.V. Costa and M.D. Fragoso, "Stability Results for Discrete-Time Linear Systems with Markovian Jumping Parameters", Journal of Mathematical Analysis and Applications, Vol 179, N 1, pp 154-178, October 1993.
- [8] H. Besbes, M.Jaidane, J.Ezzine "Convergence des filtres adaptatifs en transmission numérique : Contexte de l'alphabet fini", 16<sup>èmes</sup> Journées Tunisiennes d'Electrotechnique et d'Automatique, Nabeul Nov 1996.

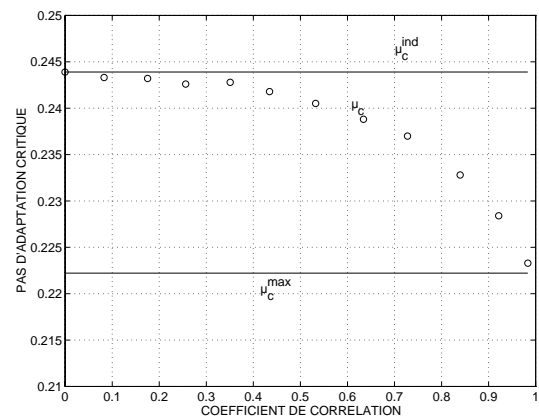


FIG. 1 — Evolution du pas d'adaptation critique en fonction de la corrélation : cas  $L = 1$

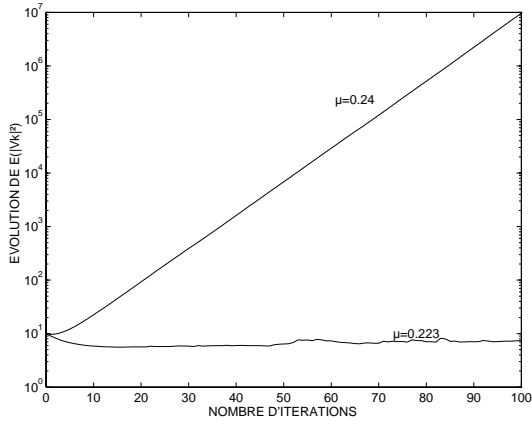


FIG. 2 — Evolution de  $E(|V_k|^2)$  : cas d'une corrélation de 0.98,  $L = 1$

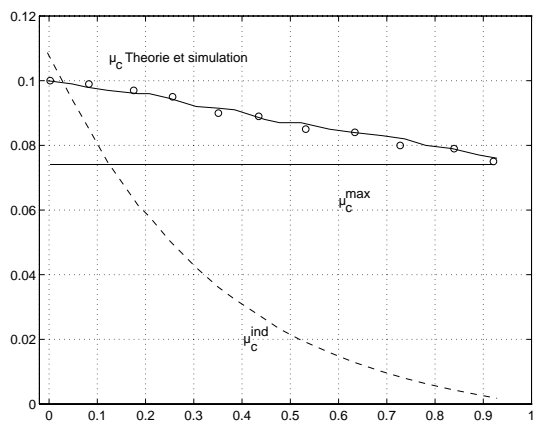


FIG. 3 — Evolution du pas d'adaptation critique en fonction de la corrélation : cas  $L = 3$