

ANALYSE COMPAREE DU COMPORTEMENT EN CONVERGENCE DES ALGORITHMES NLMS ET APA D'ORDRE 2

Abdelkrim BENAMAR ^{①②}, André GILLOIRE ^①, Ivan BOURMEYSTER ^②

^① CNET LAA/TSS/CMC 2 avenue Pierre Marzin 22307 Lannion Cedex FRANCE

^② ALCATEL MOBILE PHONES 32, avenue Kléber 92707 Colombes Cedex FRANCE

RÉSUMÉ

ABSTRACT

Dans cette contribution, nous présentons une analyse comparée du comportement en convergence des algorithmes NLMS [1] et APA (Affine Projection Algorithm) [2] d'ordre 2. L'utilisation d'un modèle statistique [3] particulier du signal d'entrée nous permet d'exprimer et de comparer analytiquement les désajustements de ces algorithmes. Enfin, l'étude des constantes de temps de convergence modale à l'aide d'un processus d'entrée autorégressif d'ordre 1 met en valeur la différence de comportement transitoire de ces algorithmes suivant la corrélation du processus d'entrée considéré. Ces conclusions sont confirmées par nos résultats expérimentaux.

In this contribution, we derive a second moment weight behavior analysis of the NLMS algorithm [1] and of the Second Order Affine Projection Algorithm (APA) [2]. For that study, we use a specific independent and identically distributed (i.i.d.) model for the input signal vectors [3]. Following a standard development [1], we obtain an expression of the asymptotic mean square error of both algorithms. Introducing a first order Markov model, we also give expressions of the modal time constants linked to the convergence behavior of these algorithms which behave differently depending on the input signal correlation. These expressions exhibit good correspondence with experimental results.

1. INTRODUCTION

Les algorithmes de filtrage adaptatif sont mis en oeuvre dans de nombreuses applications telles que la réduction de bruit, l'annulation d'écho acoustique. Parmi ces algorithmes, le LMS [1] et sa version normalisée, i.e. le NLMS [1] sont de loin les plus utilisés dans la mesure où ils associent une facilité notable de mise en oeuvre et des performances tout à fait satisfaisantes dans de nombreuses applications. Néanmoins, le comportement en convergence de ces algorithmes est très affecté par les propriétés statistiques des signaux d'entrée, et en particulier dans le cas de signaux fortement corrélés comme ceux de parole.

De nombreuses approches ont été mises en oeuvre pour améliorer la convergence du NLMS dans le cas de signaux d'entrée corrélés. Parmi celles-ci, nous pouvons citer celle consistant à préblanchir ces signaux [4]. Une autre approche décrite dans [2] [5] consiste à inclure au sein même de l'algorithme une dépendance explicite par rapport à la corrélation des signaux d'entrée et conduit à l'algorithme APA (Affine Projection Algorithm).

Dans cette contribution, nous présentons une analyse comparée du comportement en convergence des algorithmes NLMS et APA d'ordre 2. Pour ce faire, nous utilisons, en premier lieu, un modèle statistique [3] particulier du signal d'entrée qui nous permet d'exprimer et de comparer analytiquement les désajustements de ces algorithmes. Enfin, par l'intermédiaire de l'étude des constantes de temps de convergence modale à l'aide d'un processus d'entrée autorégressif d'ordre 1 nous mettons en valeur la différence de comportement transitoire de ces algorithmes suivant la corrélation du processus d'entrée considéré.

2. COMPORTEMENT ASYMPTOTIQUE DES ALGORITHMES NLMS ET APA D'ORDRE 2

2.1. Le modèle d'analyse

Ce modèle, introduit dans [3] afin d'analyser le comportement en convergence des algorithmes LMS et NLMS, est caractérisé par les hypothèses suivantes :

(H1) Les vecteurs X_t ($X_t = [x_t, \dots, x_{t-L+1}]^T$) sont constitués du produit de trois variables aléatoires indépendantes entre elles, et constituant chacune une suite indépendante et identiquement distribuée (i.i.d) :

$$X_t = s_t r_t W_t \quad (1)$$

avec $\text{Proba}(s_t = \pm 1) = 1/2$; $r_t \approx \|X_t\|$ (\approx signifie même distribution que), $\text{Proba}(W_t = V_i) = p_i = \frac{\lambda_i}{\text{tr}R}$ et où la matrice de covariance est décomposée sur une base orthonormale de vecteurs propres $\{V_i / i \in [1, L]\}$, i.e. :

$$R = V \Sigma V^T = \sum_{i=1}^L \lambda_i V_i V_i^T \quad (2)$$

(H2) En notant H_{opt} le filtre optimal à identifier, l'observation à l'instant t s'écrit :

$$y_t = (H_{\text{opt}})^T X_t + \varepsilon_t \quad (3)$$

où $\{\varepsilon_t\}$ est supposée i.i.d, de moyenne nulle, de variance σ^2 et indépendante de $\{X_t\}$.



Malgré sa simplicité, ce modèle respecte les statistiques d'ordres 1 et 2 du signal d'entrée et permet de dériver des expressions analytiques du désajustement des algorithmes considérés.

2. 2. Désajustement de l'algorithme APA d'ordre 2

L'algorithme APA d'ordre 2 est caractérisé par les relations [2]:

$$U_t = X_t - \frac{X_t^T X_{t-1}}{X_{t-1}^T X_{t-1}} X_{t-1} \quad (4)$$

$$H_t = H_{t-1} + \mu \left[\frac{y_t - H_{t-1}^T X_t}{U_t^T X_t} \right] U_t \quad (5)$$

En posant $\Delta H_t = H_{opt} - H_t$ (vecteur d'écart), la courbe d'apprentissage est donnée par la relation [1] :

$$\begin{aligned} \xi_t &= E(e_t^2) = E\left(\left[y_t - H_{t-1}^T X_t\right]^2\right) \\ &= \sigma^2 + E\left(\left[\Delta H_{t-1}\right]^T X_t\right)^2 \\ &= \sigma^2 + \text{tr}(R \text{cov}_{t-1}) \end{aligned} \quad (6)$$

avec $\text{cov}_{t-1} = E(\Delta H_{t-1}(\Delta H_{t-1})^T)$, les termes croisés s'annulant du fait du caractère i.i.d de la suite $\{\epsilon_t\}$ et de son indépendance vis à vis de $\{X_t\}$.

Pour expliciter les termes diagonaux de la matrice $R \text{cov}_{t-1}$, nous introduisons les éléments suivants :

$$\tilde{\lambda}_i(t) = (V^T \text{cov}_t V)_{ii} = V_i^T \text{cov}_t V_i \quad (7)$$

La courbe d'apprentissage s'écrit alors :

$$\xi_t = \sigma^2 + \sum_{i=1}^L \lambda_i \tilde{\lambda}_i(t-1) \quad (8)$$

Des équations (3) à (5) nous déduisons une relation de récurrence sur la suite des vecteurs d'écart :

$$\Delta H_t = \left[I - \mu \frac{U_t X_t^T}{U_t^T X_t} \right] \Delta H_{t-1} - \mu \frac{\epsilon_t U_t}{U_t^T X_t} \quad (9)$$

Il vient alors :

$$\begin{aligned} \text{cov}_t &= E \left[\left[I - \mu \frac{U_t X_t^T}{U_t^T X_t} \right] \Delta H_{t-1} (\Delta H_{t-1})^T \left[I - \mu \frac{U_t X_t^T}{U_t^T X_t} \right]^T + \mu^2 E \left[\epsilon_t^2 \frac{U_t U_t^T}{(U_t^T X_t)^2} \right] \right. \\ &\quad \left. - \mu E \left[\left[I - \mu \frac{U_t X_t^T}{U_t^T X_t} \right] \Delta H_{t-1} \frac{\epsilon_t U_t^T}{U_t^T X_t} \right] - \mu E \left[\frac{\epsilon_t U_t^T}{U_t^T X_t} (\Delta H_{t-1})^T \left[I - \mu \frac{U_t X_t^T}{U_t^T X_t} \right] \right] \right] \end{aligned}$$

Du fait du caractère i.i.d de la suite $\{\epsilon_t\}$ et de son indépendance vis à vis de $\{X_t\}$, et si l'on suppose des variations lentes du filtre d'identification H_t (convergence asymptotique ou μ faible), cette expression devient :

$$\text{cov}_t = E \left[\left[I - \mu \frac{U_t X_t^T}{U_t^T X_t} \right] \text{cov}_{t-1} \left[I - \mu \frac{U_t X_t^T}{U_t^T X_t} \right]^T + \mu^2 \sigma^2 E \left[\frac{U_t U_t^T}{(U_t^T X_t)^2} \right] \right] \quad (10)$$

Explicitons maintenant le vecteur U_t à l'aide de l'hypothèse (H1) du modèle d'analyse. Il vient :

$$U_t = s_t r_t \left(W_t - \left[\frac{W_{t-1}^T W_t}{W_{t-1}^T W_{t-1}} \right] W_{t-1} \right) \quad (11)$$

En posant $\text{Proba}(W_t = V_j \text{ et } W_{t-1} = V_k) = p_{jk}$, nous obtenons après simplification :

$$\begin{aligned} \text{cov}_t &= \text{cov}_{t-1} - \mu \sum_{j=1}^L \left(\sum_{k=1}^L p_{jk} \right) V_j V_j^T \text{cov}_{t-1} - \mu \text{cov}_{t-1} \sum_{j=1}^L \left(\sum_{k=1}^L p_{jk} \right) V_j V_j^T \\ &\quad + \mu^2 \sum_{j=1}^L \left(\sum_{k=1}^L p_{jk} \right) V_j V_j^T \text{cov}_{t-1} V_j V_j^T + \mu^2 \sigma^2 E \left(\frac{1}{r_t^2} \right) \sum_{j=1}^L \left(\sum_{k=1}^L p_{jk} \right) V_j V_j^T \end{aligned}$$

Il vient enfin :

$$\begin{aligned} \tilde{\lambda}_i(t) &= V_i^T \text{cov}_t V_i \\ &= \left[\begin{array}{c} 1 - \mu(2 - \mu) \sum_{k=1}^L p_{ik} \\ k \neq i \end{array} \right] \tilde{\lambda}_i(t-1) + \mu^2 \sigma^2 E \left(\frac{1}{r_t^2} \right) \sum_{k=1}^L p_{ik} \end{aligned} \quad (12)$$

Nous définissons alors le désajustement D_{APA2} de l'algorithme APA d'ordre 2 de la façon suivante :

$$D_{APA2} = \frac{\lim_{t \rightarrow +\infty} \xi_t - \sigma^2}{\sigma^2} = \frac{\xi_\infty - \sigma^2}{\sigma^2} \quad (13)$$

En faisant tendre t vers $+\infty$ dans les équations (8) et (12), nous obtenons :

$$D_{APA2} = \frac{\mu}{2 - \mu} (\text{tr}R) E \left(\frac{1}{r_t^2} \right) \quad (14)$$

2. 3. Désajustement de l'algorithme NLMS

L'analyse de l'algorithme NLMS à l'aide du modèle considéré est décrite dans [3]. Il est montré que le désajustement dû au bruit d'observation de cet algorithme est donné par :

$$D_{NLMS} = \frac{\mu}{2 - \mu} (\text{tr}R) E \left(\frac{1}{r_t^2} \right) \quad (15)$$

2. 4. Résultats expérimentaux

Au regard des équations (14) et (15), les expressions analytiques des désajustements des algorithmes NLMS et APA d'ordre 2 sont identiques.

Notons qu'une telle similarité pouvait être aisément supposée pour des signaux faiblement corrélés. En effet, dans le cas d'un bruit blanc, les algorithmes considérés possèdent les mêmes équations en moyenne dans la mesure où le vecteur décorréolé U_t est très peu différent de X_t en moyenne, et ainsi, ces algorithmes ont le même comportement sur ce type de signal (cf. Figure 1). Pour des signaux d'entrée fortement corrélés, cette similarité est confirmée par nos simulations (cf. Figure 2).

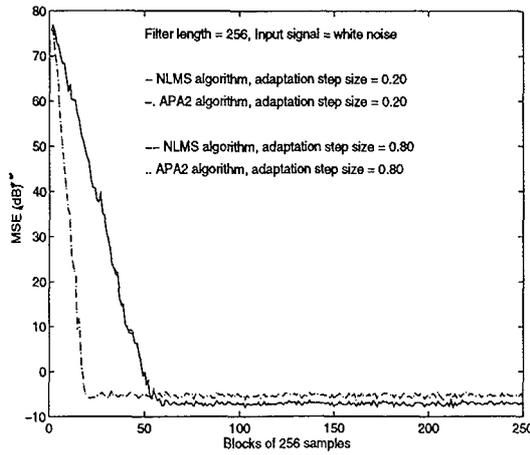


Figure 1 - EQM obtenue sur du bruit blanc

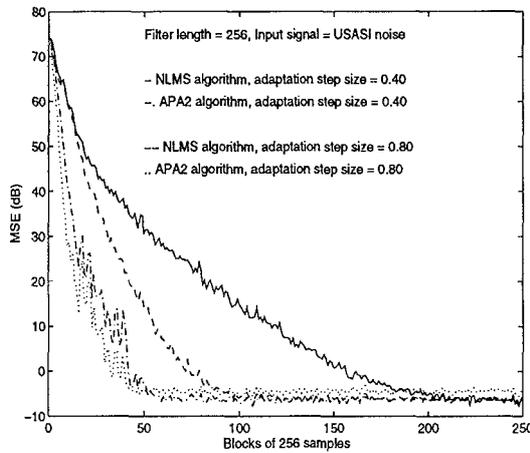


Figure 2 - EQM obtenue sur du bruit USASI

3. INFLUENCE DE LA CORRELATION DES SIGNAUX D'ENTREE SUR LE COMPORTEMENT TRANSITOIRE DES ALGORITHMES NLMS ET APA D'ORDRE 2

3.1. Le modèle d'analyse

Pour étudier l'influence de la corrélation des signaux d'entrée sur le comportement transitoire des algorithmes NLMS et APA d'ordre 2, nous comparons les constantes de temps de convergence modale de ces algorithmes dans le cas où le signal d'entrée est un processus autorégressif d'ordre 1 de paramètre α , i.e. $x_t = \alpha x_{t-1} + b_t$ où b_t est un bruit blanc gaussien.

3.2. Constantes de temps de convergence modale

Les constantes de temps de convergence modale de ces algorithmes sont données pour μ faible par $\tau_i = \frac{1}{\mu \gamma_i}$ où $\{\gamma_i\}$ est l'ensemble des valeurs propres des matrices $F_t = E \left[\frac{X_t X_t^T}{X_t^T X_t} \right]$

pour le NLMS et $G_t = E \left[\frac{U_t X_t^T}{U_t^T X_t} \right]$ pour l'APA d'ordre 2. Ces matrices sont issues de la récurrence régissant l'évolution du vecteur d'écart $\Delta H_t = H_{opt} - H_t$ de ces mêmes algorithmes (équation (9)).

Du fait de la présence de termes de normalisation dans l'expression de ces matrices, le calcul des espérances mathématiques $E \left[\frac{1}{X_t^T X_t} \right]$ et $E \left[\frac{1}{U_t^T X_t} \right]$ s'avère nécessaire pour cette analyse. Pour mener à bien ces calculs, nous procédons, pour une séquence aléatoire $\{z_t / t \geq 0\}$, à l'approximation suivante :

$$E \left[\frac{1}{z_t} \right] \approx \frac{1}{E[z_t]} = \frac{1}{M} \quad (16)$$

Une telle approximation consiste à échanger, pour un processus ergodique, les moyennes arithmétique et géométrique d'une séquence infiniment longue $\{z_t / t \geq 0\}$. Pour un bruit blanc gaussien et $M > 20$, cette approximation introduit une erreur inférieure à 10 %, erreur d'autant plus faible que $M > 20$. Néanmoins, pour de nombreuses situations pratiques cette approximation n'est pas valable. Notons, cependant, que des approximations similaires ont été effectuées par Bottomley et Alexander [6], et, Samson et Reddy [7] pour des dénominateurs variant lentement dans le temps.

Pour notre analyse, nous supposons dorénavant que :

$$E \left[\frac{1}{X_t^T X_t} \right] \approx \frac{1}{E[X_t^T X_t]} \quad (17)$$

$$E \left[\frac{1}{U_t^T X_t} \right] \approx \frac{1}{E[U_t^T X_t]} \quad (18)$$

et, plus généralement, que l'espérance mathématique des quotients intervenant dans l'expression de ces matrices est égale au quotient des espérances mathématiques.

Supposant $L \gg 1$ (filtre long), considérant le processus d'entrée décrit en 3.1 et les approximations précédentes, nous montrons que ces matrices s'écrivent :

$$F_t = [f_{ij}]_{0 \leq i, j \leq L-1} \quad (19)$$

avec

$$\forall (i, j) \in [0, L-1]^2, f_{ij} = \frac{1}{L} \alpha^{|i-j|} \quad (20)$$

$$G_t = [g_{ij}]_{0 \leq i, j \leq L-1}$$

avec

$$\forall (i, j) \in \{(i, j) / 0 \leq j \leq i \leq L-1\},$$

$$g_{ij} = \frac{\alpha^{i-j}}{L(L-1)} [L+j-i+1] - \frac{2}{L(L-1)} \sum_{k=i+1}^{L-1} \alpha^{2k-i-j}$$

$$\forall (i, j) \in \{(i, j) / 0 \leq i < j \leq L-1\},$$

$$g_{ij} = -\frac{2\alpha^{j-i}}{L(L-1)} - \frac{2}{L(L-1)} \sum_{k=j+1}^{L-1} \alpha^{2k-i-j} - \frac{\alpha^{j-i-2}}{L(L-1)} [j-i-1]$$



Pour les signaux faiblement corrélés, i.e. α tendant vers 0, ces matrices deviennent :

$$E[F_t] = [f_{ij}]_{0 \leq i, j \leq L-1} \approx \left[\frac{1}{L} \delta_{ij} \right]_{0 \leq i, j \leq L-1} \quad (21)$$

$$E[G_t] = [g_{ij}]_{0 \leq i, j \leq L-1} \approx \left[\frac{1}{L} \delta_{ij} \right]_{0 \leq i, j \leq L-1} \quad (22)$$

Ces deux matrices possèdent ainsi les mêmes valeurs propres (cf. Figure 3) et, par conséquent, les algorithmes NLMS et APA2 présentent un comportement transitoire identique sur les signaux faiblement corrélés.

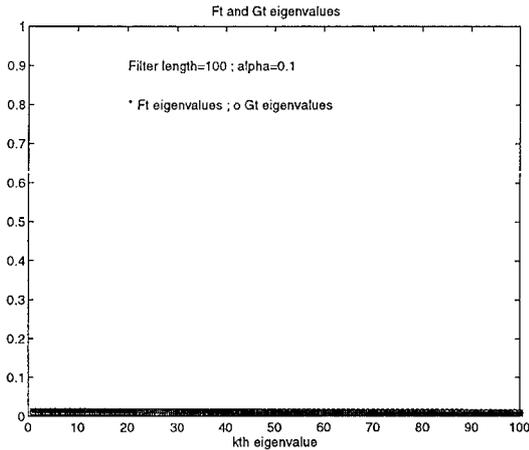


Figure 3 - Valeurs propres pour $\alpha=0.1$

Pour les signaux fortement corrélés, i.e. α tendant vers 1, ces matrices deviennent :

$$E[F_t] = [f_{ij}]_{0 \leq i, j \leq L-1} \approx \left[\frac{1}{L} \right]_{0 \leq i, j \leq L-1} \quad (23)$$

$$E[G_t] = [g_{ij}]_{0 \leq i, j \leq L-1} \quad (24)$$

avec

$$\forall (i, j) \in \{(i, j) / 0 \leq j \leq i \leq L-1\}, g_{ij} \approx \frac{1}{L(L-1)} [i+j-L+1]$$

$$\forall (i, j) \in \{(i, j) / 0 \leq j \leq i \leq L-1\}, g_{ij} \approx \frac{1}{L(L-1)} [i+j-2L+1]$$

Nous notons que les valeurs propres de la matrice $E[F_t]$ sont 1 de multiplicité 1 et 0 de multiplicité $L-1$, ce qui correspond au mode unique du processus autorégressif d'ordre 1. Les valeurs propres de la matrice $E[G_t]$ ne sont quant à elles pas aisément explicables. Nous avons donc recours aux outils numériques pour déterminer le spectre de cette matrice (cf. Figure 4).

Il ressort de ces résultats un comportement très voisin en terme de convergence des algorithmes NLMS et APA d'ordre 2 sur des signaux d'entrée faiblement corrélés. Pour des signaux d'entrée fortement corrélés, nous notons que l'algorithme NLMS privilégie fortement le mode dominant, tandis que l'algorithme APA d'ordre 2 a tendance à pondérer les modes de convergence. Une telle caractéristique contribue à expliquer la vitesse de convergence accrue de ce dernier par rapport à celle du NLMS sur des signaux d'entrée corrélés (cf. Figure 2).

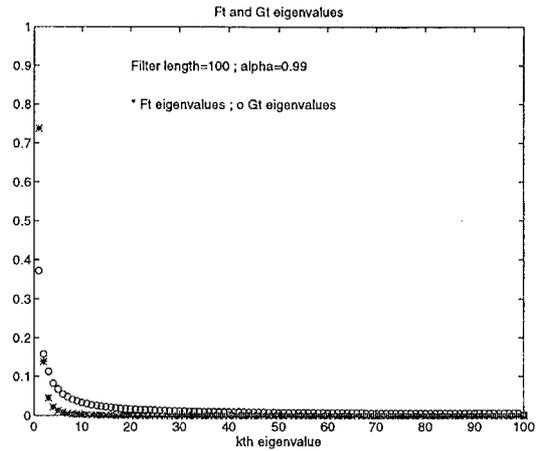


Figure 4 - Valeurs propres pour $\alpha=0.99$

4. CONCLUSION

Dans cette contribution, nous avons présenté une analyse théorique comparée du comportement asymptotique et transitoire des algorithmes NLMS et APA d'ordre 2. Cette analyse a permis de mettre en valeur un comportement asymptotique identique pour ces deux algorithmes d'identification et ce quelque soit la corrélation des signaux d'entrée. Leur comportement transitoire est quant à lui différent pour des signaux fortement corrélés (modèle d'ordre 1) : l'algorithme NLMS privilégie fortement une direction de convergence, celle associée à la plus grande valeur propre de la matrice de covariance du signal d'entrée, tandis que l'algorithme APA d'ordre 2 pondère les modes de convergence. De telles propriétés permettent de considérer les algorithmes de projection affine (APA) comme des solutions intéressantes pour l'identification dans le cas de signaux d'entrée fortement corrélés.

REFERENCES

- [1] S. Haykin, Adaptive Filter Theory, Englewood Cliffs, N.J. Prentice-Hall, Chapter 5, 1991.
- [2] K. Ozeki, T. Umeda, "An Adaptive Filtering Algorithm Using an Orthogonal Projection to an Affine Subspace and its Properties", Electronics and Communications in Japan, vol. 67-A, n°5, pp. 19-27, 1984.
- [3] D.T.M. Slock, "On the Convergence Behavior of the LMS and the Normalized LMS Algorithms", IEEE trans. on S.P., vol. 41, n°9, pp. 2811-2825, September 1993.
- [4] M. Mboup, M. Bonnet, N. Bershad, "Coupled Adaptive Prediction and System Identification: A Statistical Model and Transient Analysis", Proc. IEEE ICASSP'92, San Francisco, pp.IV-1 - IV-4, 1992.
- [5] M. Montazeri, P. Duhamel, "A Set of Algorithms Linking NLMS and Block RLS Algorithms", Proc. IEEE ICASSP'93, Minneapolis, pp. 412-415, 1993.
- [6] G.E. Bottomley, S.T. Alexander, "A novel approach for stabilizing recursive least squares filters," IEEE Trans. on S.P., vol. 39, pp. 1770-1779, August 1991.
- [7] C.G. Samson, V.U. Reddy, "Fixed point error analysis of the normalized ladder algorithm," IEEE Trans. on A.S.S.P., vol. 31, pp. 1177-1191, October 1983.