EXACT CONVERGENCE ANALYSIS OF AFFINE PROJECTION ALGORITHM: THE FINITE ALPHABET INPUTS CASE

H. Besbes[(1)(4)], *Y. Ben Jemaa*[(1)(3)], *M. Jaidane*[(1)(2)]

[(1)] Laboratoire des Systèmes de Télécomunications, ENIT/ESPTT
[(2)] ENIT, Campus Universitaire, Le Belvédère, Tunis, TUNISIE
Phone: (216) 1 874 700, Fax: (216) 1 872 729
[(3)] L.S.S SUPELEC, Plateau de Moulon 91192 GIF sur YVETTE, Paris
Phone: (33) 1 69.85.17.19, Fax: (33) 1 69.41.30.60
[(4)] SUPCOM, Route de Raoued, Km3.5 Cité El Ghazala 2083, Tunis, TUNISIE
Phone: (216) 1 857 000, Fax: (216) 1 856829, Email: hichem.besbes@supcom.rnu.tn

ABSTRACT

The affine projection algorithm (APA) is a very promising algorithm that has good convergence properties when the input signal is correlated. In particular, it's used to perform communications systems: echo cancellation, equalization... However, due to its complexity, there is no available transient and steady state analysis. In this paper, we present an exact analysis approach tailored for digital transmission context. In such context, the input signal remains in a finite alphabet set. With a discrete Markov chain model of the inputs, we can describe accurately the APA's behavior without any unrealistic assumption. In particular we calculate the exact value of critical and optimum step size. Moreover, we provide the exact Mean Square Deviation for all step size and input correlation. The influence of high order statistics can be enhanced.

1. BACKGROUND

The Least-Mean-Square (LMS) algorithm is well known in adaptive filtering, however its convergence speed may be far from acceptable for many applications, when the input sequence is correlated. To overcome this problem, many modifications that aim to decorrelate the input signal are done. In particular, this has lead to the affine projection algorithm (APA) [1], [2]. It has very promising performance in acoustic echo cancellation [3], [4].

The APA algorithm can be also used in adaptive equalization and CDMA applications in order to minimize the training sequence. In such context, the input signal is sampled, quantified and coded, so it remains in a finite alphabet set.

The APA algorithm is given by the following system:

$$\left\{ \begin{array}{rl} H_{k+1} & = H_k + \mu \dfrac{\phi_k}{\phi_k^T \phi_k} e_k \\ e_k & = y_k - H_k^T X_k \end{array} \right. \tag{1}$$

Where $X_k = [x_k, ..., x_{k-L+1}]^T$ is the input observation vector, L is the dimension of the adaptive filter, and μ is a positive step size. The output signal is defined by $y_k = F^T X_k + b_k$, where F is the "impulse response" of the channel and b_k is the observation noise assumed to be zero-mean noise, independent of X_k. The estimates parameters vector is H_k and ϕ_k can be defined as a direction vector since he fixes the direction of the update. It is defined by [1], [2]:

$$\phi_k = X_k - U_k a_k \tag{2}$$

Where $a_k = [U_k^T U_k]^{-1} U_k^T X_k$, and the matrix U_k is the collection of the last m observations of X_k with $m < L$, $U_k = [X_{k-1}, ..., X_{k-m}]$.

It's important to note that the last equation (2) represents a decorrelation operation of the input signal.

The behavior of the algorithm can be described by the evolution of the deviation vector $V_k = H_k - F$. The recursion of V_k is given by:

$$V_{k+1} = V_k - \mu \frac{\phi_k}{\phi_k^T \phi_k + \epsilon} [X_k^T V_k + b_k] \tag{3}$$

Where ϵ is a positive regularization constant which is added to prevent undesired behavior when $\phi_k = 0$.

Analysis of the APA algorithm, in the mean and mean square sense, has been hard to find. Few results are obtained for a modified version of APA, with white noise input sequence [5].

The aim of this paper is to study the convergence and steady state properties of the APA in the digital transmission context. In such context we propose, as in [6], [7] for LMS analysis, to use a discrete Markov chain model of the input data to overcome the complexity of analysis.

In digital transmission contexts, the input signal x_k remains in a finite alphabet set $A = \{a_1, a_2, ..., a_q\}$ such as $M-$ary signal, QAM signal... Consequently, the observation vector X_k remains also in a finite alphabet $\mathcal{A} = \{W_1, W_2, \cdots, W_N\}$ with cardinality $N = q^d$. For example, when $x_k \in \{\pm 1\}$, and the dimension of X_k is $d = 2$, the finite alphabet set is $\mathcal{A} = \left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \end{pmatrix} \begin{pmatrix} -1 \\ -1 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \end{pmatrix} \right\}$.

Since the input sequence is stationary, it can be modeled by a discrete-time Markov chain $\{\theta(k) : k \in \mathbf{Z}^+\}$ with finite state space $\{1, 2, ...N\}$ [6], [7], such that:

$$X_k = W_{\theta(k)} \tag{4}$$

The discrete time Markov chain is characterized by its probability transition matrix $\mathbf{P} = [p_{ij}]$ and its stationary probability vector π_∞.

In the next section we calculate the exact value of critical and optimum step size. Moreover, we provide the exact Mean square Deviation MSD for all step size and input correlation.

2. PERFORMANCE ANALYSIS OF THE AP ALGORITHM

2.1. Proposed approach

In this section we evaluate the APA's performances with order $m = 1$ and versus the step size μ. In this case the vector ϕ_k is given by:

$$\phi_k = f(X_{k-1}, X_k) = X_k - X_{k-1}\frac{X_{k-1}^T X_k}{X_{k-1}^T X_{k-1}} \tag{5}$$

Then,

$$\phi_k^T \phi_k = X_k^T X_k - \frac{X_k^T X_{k-1} X_{k-1}^T X_k}{X_{k-1}^T X_{k-1}} \tag{6}$$

Consequently,

$$\frac{\phi_k}{\phi_k^T\phi_k + \epsilon} = h(X_{k-1}, X_k) = \frac{X_k - X_{k-1}\dfrac{X_{k-1}^T X_k}{X_{k-1}^T X_{k-1}}}{X_k^T X_k - \dfrac{(X_k^T X_{k-1})^2}{X_{k-1}^T X_{k-1}} + \epsilon} \tag{7}$$

Since X_k remains in a finite alphabet set, $\dfrac{\phi_k}{\phi_k^T\phi_k + \epsilon}$ remains also in a finite alphabet. The vector $\dfrac{\phi_k}{\phi_k^T\phi_k + \epsilon}$ can be modeled by:

$$\frac{\phi_k}{\phi_k^T\phi_k + \epsilon} = g_{\theta(k-1),\theta(k)},$$

and the transient matrix of algorithm becomes:

$$I - \mu\frac{\phi_k}{\phi_k^T\phi_k + \epsilon}X_k^T = M_{\theta(k-1),\theta(k)}.$$

So the recursion of the deviation vector can be rewrite as follow:

$$V_{k+1} = M_{\theta(k-1),\theta(k)}V_k - \mu g_{k-1,k} b_k \tag{8}$$

The performance analysis are made through the evolution of $E(V_k V_k^T)$. Since b_k is zero mean independent of the input sequence, the recursion of $E(V_k V_k^T)$ becomes:

$$\begin{aligned} E(V_{k+1}V_{k+1}^T) = \quad & E(M_{\theta(k-1),\theta(k)} V_k V_k^T M_{\theta(k-1),\theta(k)}^T) \\ & +\mu^2\sigma_b^2 E(g_{\theta(k-1),\theta(k)} g_{\theta(k-1),\theta(k)}^T) \end{aligned} \tag{9}$$

The main idea in this paper is, since there is N possibilities of $W_{\theta(k)}$ and N^2 possibilities of $M_{\theta(k-1),\theta(k)}$, we split the matrix $E(V_k V_k^T)$ in N^2 components defined by:

$$q_{i,j}(k) = E(V_k V_k^T 1_{\theta(k-1)=i} 1_{\theta(k)=j}) \tag{10}$$

So we obtain the recursion:

$$\begin{aligned} q_{i,j}(k+1) &= \sum_{l=1}^{N} E(V_{k+1}V_{k+1}^T 1_{\theta(k+1)=j} 1_{\theta(k)=i} 1_{\theta(k-1)=l}) \\ &= \sum_{l=1}^{N} E(M_{l,i} V_k V_k^T 1_{\theta(k+1)=j} 1_{\theta(k)=i} 1_{\theta(k-1)=l} M_{l,i}^T) \\ &+\mu^2\sigma_b^2 \sum_{l=1}^{N} E((g_{l,i} g_{l,i}^T) 1_{\theta(k+1)=j} 1_{\theta(k)=i} 1_{\theta(k-1)=l}) \end{aligned} \tag{11}$$

Since $M_{l,i}$ are constant matrices, the difficulties to analyze the APA are avoided, and one can deduce the recursive formulae between $q_{i,j}(k+1)$ and $q_{j,l}(k)$ without any independence assumption by:

$$\begin{aligned} q_{i,j}(k+1) = \quad & \sum_{l=1}^{N} M_{l,i} q_{l,i}(k) M_{l,i}^T. \\ & .Prob(\theta(k+1) = j/\theta(k) = i, \theta(k-1) = l) \\ & +\mu^2\sigma_b^2 \sum_{l=1}^{N} g_{l,i} g_{l,i}^T \pi_\infty(l) p_{l,i} \end{aligned} \tag{12}$$

Then,

$$\begin{aligned} q_{i,j}(k+1) = \quad & \sum_{l=1}^{N} M_{l,i} q_{l,i}(k) M_{l,i}^T\, p_{i,j}\; sgn(p_{l,i}) \\ & +\mu^2\sigma_b^2 \sum_{l=1}^{N} g_{l,i} g_{l,i}^T \pi_\infty(l) p_{l,i} \end{aligned} \tag{13}$$

Where $sgn(.)$ denotes the sign function.

In order to rewrite the last recursion in linear form, we introduce the useful notations:

$$\tilde{Q}_i(k) = \begin{bmatrix} vec(q_{1,i}(k)) \\ vec(q_{2,i}(k)) \\ \vdots \\ vec(q_{N,i}(k)) \end{bmatrix} \in \mathcal{R}^{NL^2},$$

$$Q_i(k) = \begin{bmatrix} vec(q_{i,1}(k)) \\ vec(q_{i,2}(k)) \\ \vdots \\ vec(q_{i,N}(k)) \end{bmatrix} \in \mathcal{R}^{NL^2}, \quad \tilde{Z}_i = \begin{bmatrix} vec(z_{i,1}) \\ vec(z_{i,2}) \\ \vdots \\ vec(z_{i,N}) \end{bmatrix}.$$

We have then:

$$Q_i(k+1) = \Gamma_i \tilde{Q}_i(k) + \tilde{Z}_i \tag{14}$$

Where

$$\Gamma_i = (K_i^T \otimes I_{L^2}) diag(sgn(p_{li})\, M_{l,i} \otimes M_{l,i})_{l=1..N} \tag{15}$$

and K_i is a matrix $(N \times N)$ which has all the rows equal to the i^{th} row of the transition matrix P.

Finally, we denote:

$$\hat{Q}(k) = \begin{bmatrix} \tilde{Q_1}(k) \\ \tilde{Q_2}(k) \\ \vdots \\ \tilde{Q_N}(k) \end{bmatrix} \quad \hat{Z} = \begin{bmatrix} \tilde{Z}_1 \\ \tilde{Z}_2 \\ \vdots \\ \tilde{Z_N} \end{bmatrix} \quad \underline{Q}(k) = \begin{bmatrix} Q_1(k) \\ Q_2(k) \\ \vdots \\ Q_N(k) \end{bmatrix}$$

We obtain:

$$\underline{Q}(k+1) = diag(\Gamma_i)\hat{Q}(k) + \hat{Z} \qquad (16)$$

We introduce $\widehat{P}$, the permutation matrix that transforms the vector $\hat{Q}(k)$ at $\underline{Q}(k)$:

$$\hat{Q} = \widehat{P}\,\underline{Q}(k) \qquad (17)$$

So, we have:

$$\underline{Q}(k+1) = diag(\Gamma_i)\,\widehat{P}\underline{Q}(k) + \hat{Z} \qquad (18)$$

Finally, we obtain the important equation:

$$\underline{Q}(k+1) = \Gamma\underline{Q}(k) + \hat{Z} \qquad (19)$$

Where

$$\Gamma = diag(\Gamma_i)\,\widehat{P} \qquad (20)$$

The matrix Γ contains all relevant information about the algorithm performances.

It is interesting to note, that through the proposed approach the needed assumptions are a subset of the usual ones. Most importantly, in the finite alphabet case, no independence assumption is needed.

2.2. Exact determination of the algorithm performances

2.2.1. Exact critical and optimal step sizes

In transient state, the critical step size μ^c is defined as the step from which the algorithm diverge. This is found when Γ has an eigenvalue higher than 1. So, we can determine μ^c as follows:

$$\mu^c = arg(\lambda^{\Gamma}_{max}(\mu) = 1) \qquad (21)$$

The optimal step size μ^{opt} is defined as the step that gives the maximal speed convergence. This is found when Γ has all eigenvalues small. So, we can determine μ^{opt} as follows:

$$\mu^{opt} = arg(min(\lambda^{\Gamma}_{max}(\mu))) \qquad (22)$$

The quantity $\lambda^{\Gamma}_{max}(\mu^{opt})$ fixes the speed of convergence.

To illustrate these results, lets as consider a particular identification scheme with $L = 2$ and x_k belongs to the finite alpahbet set $\{+3, +1, -1, -3\}$. For the evaluation of μ_c and μ^{opt}, we consider three cases of input statistics. The first case deals with i.i.d. inputs. The second and the third cases deal with a particular transition matrix wich has the folowing form:

$$\begin{bmatrix} \alpha & 0 & (1-\frac{9\alpha}{7}) & \frac{2\alpha}{7} \\ \frac{2\alpha}{7} & \alpha & 0 & (1-\frac{9\alpha}{7}) \\ (1-\frac{9\alpha}{7}) & \frac{2\alpha}{7} & \alpha & 0 \\ 0 & (1-\frac{9\alpha}{7}) & \frac{2\alpha}{7} & \alpha \end{bmatrix}.$$

The factor α and the transistion matrix fixe all the statistics of the input signal.

The variation of λ^{Γ}_{max} versus the step size is depicted in figure 1. It is interesting to note that in such case, the optimal step size μ^{opt} is equal to 1 for differents inputs statistics. However the critical step size depends on the transition matrix and it is always less than 2.

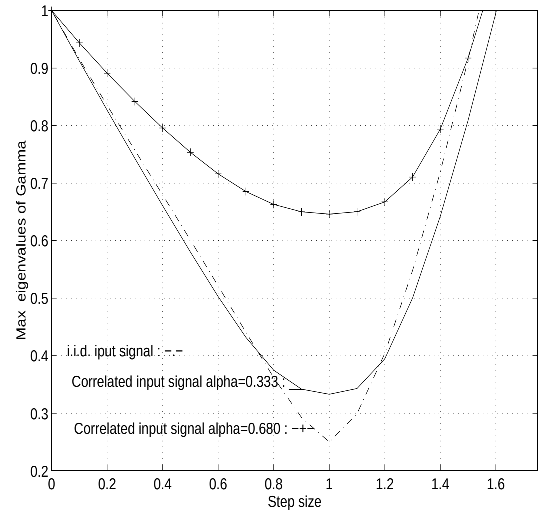

Figure 1: Evolution of λ^{Γ}_{max} versus the step size

2.2.2. Exact MSD

Since the equation (19) is linear, we can deduce easily $\underline{Q}_{\infty}$ by:

$$\underline{Q}_{\infty} = \lim_{k\to\infty} \underline{Q}(k) = (I-\Gamma)^{-1}\hat{Z} \qquad (23)$$

From the last equation we obtain the MSD:

$$MSD = \sum_{i=1}^{N}\sum_{j=1}^{N} tr(q_{ij\,\infty}) \qquad (24)$$

Where $q_{ij\,\infty}$ are deduced from the vector $\underline{Q}_{\infty}$.

In the considered simulation example, the input sequence x_k belongs to the alphabet $\{\pm 1\}$ with a transition matrix:

$$\begin{bmatrix} \frac{3}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{3}{4} \end{bmatrix},$$

and filter lengh is $L = 2$.

Then, the transition matrix for the input signal X_k is:

$$P = \begin{bmatrix} \frac{3}{4} & 0 & 0 & \frac{1}{4} \\ \frac{3}{4} & 0 & 0 & \frac{1}{4} \\ 0 & \frac{1}{4} & \frac{3}{4} & 0 \\ 0 & \frac{1}{4} & \frac{3}{4} & 0 \end{bmatrix}$$

Refering to the equation (24) we can determine the MSD calculated by the new approach.

To determine the MSD, we run a Monte-Carlo simulations over 50 realizations, and estimate the needed results by averaging the MSD after convergence. Figure 2. represents the evolution of the MSD with the step size.

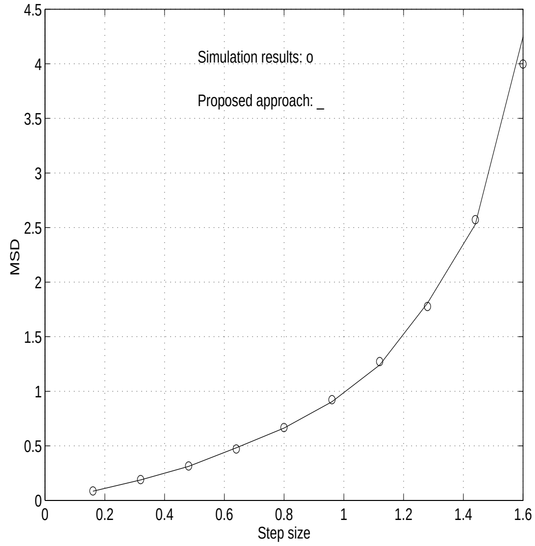

Figure 2: MSD variation with the step size

The main contribution of this finite alphabet tailored analysis is its exactness. This is illustrated by the perfect agreement beween the simulation and theortical results.

3. CONCLUSION

APA's performances are analyzed in the real context of digital transmission where the input signal belongs to a finite alphabet. This exact analysis was easily done without any unrealistic hypothesis. We calculate the exact value of critical and optimum step size. Moreover, we provide the exact Mean Square Deviation for all step size and input correlation.

The finite alphabet approach seems to be a very powerful tool to global optimization of a transmission chain including source coders, adaptive filters, and channel coders...

4. REFERENCES

[1] K. Ozeki and T. Umeda, "An adaptive filtering algorithm using orthogonal projection to an affine subspace and its properties", Electron, Commun. Jpn., vol 67-A, N 5, pp 19-27, 1984.

[2] M. Rupp, "A Family of Adaptive Filter Algorithms with Decorrelating Properties", IEEE Trans. on Signal Processing, Vol 46, N 3, pp 771-775, March 1998.

[3] Q.G. Liu, B. Champagne and K.C. Ho," On the Use of a Modified Fast Affine Projection Algorithm in Subbands for Acoustic Echo Cancellation", Proc. IEEE Digital Signal Processing, pp 354-357, September 1996.

[4] S. Makino, K. Straus, S. Shimauchi, Y. Haneda and A. Nakagawa, "Subband Stereo Echo Cancellar Using the Projection Algorithm with Fast Convergence to the true Echo Path", in Proc ICASSP'97 pp 299-302, April 1997.

[5] M.L.R de Campos, J. A. Apolinário Jr. and P. S. R. Diniz, " Mean-Squared error analysis of binormalized data-reusing LMS algorithm using a discrete angular distribution model for the input signal", in ICASSP'98 pp Seattle, USA, May 1998.

[6] H. Besbes, M. Jaîdane-Saidane and J. Ezzine, "On Exact Convergence Results of Adaptive Filters: the Finite Alphabet Case", submitted to Int. J. Signal Processing of EURASIP, Dec 1997.

[7] H. Besbes, M. Jaîdane-Saidane and J. Ezzine, " Exact analysis of the tracking capability of time-varying channels: the finite alphabet input case." in Proc. of 5^{th} IEEE Int.Conf. on Elect. Circ and Systems ICECS'98, pp 449-452, 7-10 Sept 1998, Lisboa, Portugal.