Statistical Analysis of a Pseudo Affine Projection Adaptive Algorithm

Sérgio J. M. de Almeida*, José Carlos M. Bermudez** and Neil J. Bershad***

*Catholic University of Pelotas, Pelotas-RS, Brazil
**Federal University of Santa Catarina, Florianópolis, SC, Brazil
***University of California, Irvine, USA
e-mails: smelo@eel.ufsc.br, j.bermudez@ieee.org, bershad@ece.uci.edu.

Abstract— **This paper presents a statistical analysis of a Pseudo Affine Projection (PAP) algorithm, obtained by introducing a step size control parameter in the weight update equation of the Affine Projection (AP) Algorithm. An analytical model is derived for predicting the algorithm stochastic behavior for any step size $\alpha \leq 1$. Deterministic recursive equations are derived for the mean weight and for the mean square error behaviors for a large number of adaptive taps N, as compared to the order P of the algorithm. Simulations are presented which show excellent agreement between theory and simulations in steady-state and during transient. The learning behavior of the PAP algorithm is of special interest in applications where a tradeoff between convergence speed and steady-state misadjustment is necessary.**

I. Introduction

The Affine Projection (AP) algorithm proposed by Ozeki and Umeda in 1984 [2] applies weight updates in directions that are orthogonal to the last P input vectors. This decorrelates the input signal and speeds up convergence [3], making the algorithm attractive for applications with highly correlated input signals [4]. The price to be paid for the improved performance is an increased computational complexity and a slightly higher noise floor, when compared to algorithms such as the normalized least mean squares (NLMS). The complexity cost has been decreasing with recent advances in the semiconductor industry. The higher noise floor, however, is an intrinsic property of the algorithm. One way to reduce the steady-state misadjustment is to introduce a step size control. Such a non-unity step size allows a tradeoff between steady-state misadjustment and convergence speed. Such reduction of the maximum speed step size has also been tipically used for the NLMS algorithm in practical echo cancellers to prevent undesirable effects of the system's non-idealities of the real world implementation [12]. Thus, it is of interest to study the behavior of the AP algorithm when its weight update equation is modified to incorporate a step size control parameter α.

Reference [4] has presented a quantitative analysis of the AP algorithm. The analysis is based upon an independent input signal model originally proposed in [5]. However, the independent signal model cannot handle the pre-whitening properties of the AP algorithm. Reference [6] presented a quantitative analysis for autoregressive (AR) Gaussian inputs and unit step size. This analysis follows the work in [3] (which was also restricted to unit step size) for obtaining the solution of a recursion for the weight error vector variances. The solution uses previous results for the NLMS algorithm with white inputs. Reference [10] presented a new statistical analysis for the behavior of the AP algorithm for Gaussian AR inputs and unit step size. Analytical difficulties are avoided for the case of a large number of adaptive taps compared to the AP algorithm order. More recently, the results in [10] have been improved through the use of new statistical assumptions and approximations, as well as the consideration of vector properties of certain variables [11]. The results in [11] are again for unit step size.

This paper extends the work in [11] to the statistical analysis of the AP algorithm when the update equation is modified by the introduction of a step size control parameter $\alpha \leq 1$. The resulting algorithm is no longer the AP algorithm, and is thus named here the Pseudo Affine Projection (PAP) algorithm. The use of a non-unit step size requires new considerations for the analysis, since the original structure of the AP algorithm has been modified. The analysis assumes an AR input and a large number of adaptive taps, as compared to the algorithm's order. Analytical recursions are derived which predict the behavior of the mean weight vector and the mean square error (MSE). Monte Carlo simulation results show excellent agreement with theoretical predictions during the adaptation phase and in steady-state.

II. The Input Signal Model

The adaptive system attempts to estimate a desired signal $d(n)$ that can be modeled by

$$d(n) = \mathbf{w^{o^T}}\mathbf{u}(n) + r(n)$$

where $\mathbf{w^o} = [w_1^o \; w_2^o \; \ldots \; w_{N-1}^o]^T$ is the vector of the model parameters and $r(n)$ is a white noise with variance σ_r^2, which accounts for measurement noise and modeling errors.

The input signal $u(n)$ is assumed to be a stationary AR process, modeling input signals for many practical applications. Let $\mathbf{u}(n)$ be a vector of N samples of an AR process of order P. Thus,

$$\mathbf{u}(n) = \sum_{i=1}^{P} a_i \mathbf{u}(n-i) + \mathbf{z}(n) = \mathbf{U}(n)\mathbf{a} + \mathbf{z}(n) \qquad (1)$$

where the matrix $\mathbf{U}(n) = [\mathbf{u}(n-1) \ldots \mathbf{u}(n-P)]$ is a collection of P past input vectors $\mathbf{u}(n-k) = [u(n-k) \; \ldots \; u(n-k-N+1)]^T$ and $\mathbf{z}(n) = [z(n) \; \ldots \; z(n-N+1)]^T$ is a vector with

Sérgio J. M. de Almeida and is with Escola de Engenharia e Arquitetura, Universidade Católica de Pelotas, Pelotas-RS, Brazil, José Carlos M. Bermudez is with Laboratório de Pesquisas em Processamento Digital de Sinais, Departamento de Engenharia Elétrica, Universidade Federal de Santa Catarina, Florianópolis-SC, Brazil, 88040-900, Phone: +55 48 331 7627 Fax:+55 48 9280. Neil J. Bershad is with the Department of Eletrical and Computer Engineering, University of California, Irvine, CA 92697, 949-824-6709, FAX-949-824-3203.
This work was partially supported by CNPq Grant No. 352084/92-8

samples from a stationary white Gaussian process with variance σ_z^2.

The least squares estimate of the parameter vector $\mathbf{a}$ is given by:

$$\hat{\mathbf{a}}(n) = [\mathbf{U}^T(n)\mathbf{U}(n)]^{-1}\mathbf{U}^T(n)\mathbf{u}(n) \quad (2)$$

where $\mathbf{U}^T(n)\mathbf{U}(n)$ is assumed of rank P.

III. The Weight Update Equation

The weight update equation of the PAP algorithm, obtained from the AP algorithm recursion [3] by the inclusion of a step size parameter α is given by:

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \alpha \frac{\mathbf{\Phi}(n)}{\mathbf{\Phi}^T(n)\mathbf{\Phi}(n)} e(n) \quad (3)$$

where the error signal $e(n)$ is given by

$$e(n) = \mathbf{w}^{o^T}\mathbf{u}(n) + r(n) - \mathbf{w}^T(n)\mathbf{u}(n) \quad (4)$$

and $\mathbf{w}(n) = [w_0(n)\ w_1(n)\ \ldots\ w_{N-1}(n)]^T$ is the adaptive weight vector.

Note that the original AP algorithm [2] with non-unit step size would lead to a vector error signal, instead of the scalar error in (3). Recursion (3), however, coincides with the AP update equation when $\alpha = 1$.

The vector $\mathbf{\Phi}(n)$ defines the direction of update, and is given by:

$$\mathbf{\Phi}(n) = \mathbf{u}(n) - \mathbf{U}(n)\hat{\mathbf{a}}(n) \quad (5)$$

The order of the AP algorithm is defined by the number $(P+1)$ of input vectors used to determine $\mathbf{\Phi}(n)$.

IV. Vector and Statistical Properties of $\mathbf{\Phi}$

The following analysis invokes assumptions similar to the independence assumption used to analyze many stochastic algorithms [1].

Assumption A1: The statistical dependence between $\mathbf{z}(n)$ and $\mathbf{U}(n)$ can be neglected. This assumption is justified as follow and is more realistic for $N \gg P$.

Equation (1) shows an algebraic dependence between $\mathbf{z}(n)$ and vectors $\mathbf{u}(n-1), \ldots, \mathbf{u}(n-P)$. Also, $\mathbf{z}(n)$ is of dimension N. Consider $\mathbf{P}_u(n) = \mathbf{U}(n)[\mathbf{U}^T(n)\mathbf{U}(n)]^{-1}\mathbf{U}^T(n)$, the projection matrix onto the subspace spanned by the columns of $\mathbf{U}(n)$, and $\mathbf{P}_o(n) = \mathbf{I} - \mathbf{P}_u(n)$, the projection matrix onto the orthogonal complement subspace. Then, $\mathbf{z}(n)$ can be decomposed as $\mathbf{z}_u(n) + \mathbf{z}_\perp(n)$, where $\mathbf{z}_u(n) = \mathbf{P}_u(n)\mathbf{z}(n)$ and $\mathbf{z}_\perp(n) = \mathbf{P}_o(n)\mathbf{z}(n)$. Only $\mathbf{z}_u(n)$ is algebraically dependent upon $\mathbf{U}(n)$. Moreover, since $z(n)$ is white, the average energy of $\mathbf{z}(n)$ is equally distributed among its N dimensions. Thus, only the energy in $\mathbf{z}_u(n)$ creates a dependence between $\mathbf{z}(n)$ and $\mathbf{U}(n)$. This dependence can be neglected if $N \gg P$.

Assumption A2: $\mathbf{\Phi}(n)$ and the weight vector $\mathbf{w}(n)$ are statistically independent.

This assumption is similar to the independence assumption applied to delay line adaptive filters with white inputs since $\mathbf{\Phi}(n)$ is a vector whose elements are estimates of the white noise sequence $z(n)$ [3].

Substituting (1) in (5) yields

$$\mathbf{\Phi}(n) = \mathbf{I} - \mathbf{P}_u(n)]\mathbf{z}(n) = \mathbf{P}_o(n)\mathbf{z}(n) = \mathbf{z}_\perp(n) \quad (6)$$

Eq. (6) shows that $\mathbf{\Phi}(n)$ is orthogonal to the columns of $\mathbf{U}(n)$.

The structure and the properties of the correlation matrix $\mathbf{R}_{\phi\phi}$ require consideration of the vector and statistical properties of $\mathbf{\Phi}(n)$.

First, $\mathbf{z}_\perp(n)$ is a vector with power only in $(N-P)$ dimensions of the N-dimensional space. The vector $\mathbf{z}_u(n)$ contributes the power in the remaining P dimensions. Consider a given iteration (a fixed value for n). In general, the dimensions excited by $\mathbf{z}_\perp(n)$ are different for each sample function of the adaptive process because of the randomness of $u(n)$. On average, this is equivalent to all dimensions excited at each run (for any given n) with $(N-P)/N$ of the power in $\mathbf{z}(n)$. This reasoning is detailed in the following calculations.

From (6), the correlation matrix of $\mathbf{\Phi}(n)$ can be written as:

$$\mathbf{R}_{\phi\phi} = E\{\mathbf{\Phi}(n)\mathbf{\Phi}^T(n)\} = E\{\mathbf{z}_\perp(n)\mathbf{z}_\perp^T(n)\} \quad (7)$$

Using $\mathbf{z}(n) = \mathbf{z}_\perp(n) + \mathbf{z}_u(n)$ and noting that $E\{\mathbf{z}_\perp(n)\mathbf{z}_u^T(n)\} = 0$ and $E\{\mathbf{z}_u(n)\mathbf{z}_\perp^T(n)\} = 0$, since for each run $\mathbf{z}_\perp(n)$ and $\mathbf{z}_u(n)$ always have powers in different directions, it is easy to show that

$$\mathbf{R}_{\phi\phi} = E\{\mathbf{z}(n)\mathbf{z}^T(n)\} - E\{\mathbf{z}_u(n)\mathbf{z}_u^T(n)\} \quad (8)$$

An expression for $\mathbf{R}_{\phi\phi}$ is now derived based on a equal distribution of the average power in each dimension. The total power contributed by each term on the r.h.s. of (8) is given by

$$tr[E\{\mathbf{z}(n)\mathbf{z}^T(n)\}] = N \cdot \sigma_z^2 \quad (9)$$

and

$$tr[E\{\mathbf{z}_u(n)\mathbf{z}_u^T(n)\}] = P \cdot \sigma_z^2 \quad (10)$$

Distributing the power equally in all dimensions results in

$$\mathbf{R}_{\phi\phi} = E\{\Phi(n)\Phi^T(n)\} = \sigma_\phi^2 \cdot I = \left(\frac{N-P}{N}\right) \cdot \sigma_z^2 \cdot I \quad (11)$$

Assumption A3: $\mathbf{\Phi}(n)$ is a zero mean Gaussian random vector.

Eq. (6) shows that each component $\phi(n-i)$ of $\mathbf{\Phi}(n)$ is determined by $\sum_{j=1}^{N} P_{o_{ij}} z(n-j-1)$. From assumption A1 and $z(n)$ white, the random variables in this sum are independent. Thus, by the Central Limit Theorem, the distribution of $\mathbf{\Phi}(n)$ tends to a Gaussian for large N.

V. Mean Weight Behavior

Defining the weight error vector, $\mathbf{v}(n) = \mathbf{w}^o - \mathbf{w}(n)$ and using (3), (4) leads to

$$\begin{aligned}\mathbf{v}(n+1) = \mathbf{v}(n) &- \alpha \frac{\mathbf{\Phi}(n)\mathbf{u}^T(n)}{\mathbf{\Phi}^T(n)\mathbf{\Phi}(n)}\mathbf{v}(n) \\ &- \alpha \frac{\mathbf{\Phi}(n)}{\mathbf{\Phi}^T(n)\mathbf{\Phi}(n)} r(n)\end{aligned} \quad (12)$$

Defining the the instantaneous error $e_a(n) = \mathbf{u}^T(n)\mathbf{v}(n)$, such that $e(n) = e_a(n) + r(n)$, (12) can be written as

$$\mathbf{v}(n+1) = \mathbf{v}(n) - \alpha \frac{\mathbf{\Phi}(n)}{\mathbf{\Phi}^T(n)\mathbf{\Phi}(n)} e_a(n) - \alpha \frac{\mathbf{\Phi}(n)}{\mathbf{\Phi}^T(n)\mathbf{\Phi}(n)} r(n) \quad (13)$$

Pre-multiplying (12) by $\mathbf{u}^T(n)$ and $\mathbf{U}^T(n)$, using (5) as $\mathbf{u}^T(n) = \mathbf{\Phi}^T(n) + \hat{\mathbf{a}}^T(n)\mathbf{U}^T(n)$, and noting from (6) that $\mathbf{U}^T(n)\mathbf{\Phi}(n) = 0$ yields [3]

$$\mathbf{u}^T(n)\mathbf{v}(n+1) = (1-\alpha)\mathbf{u}^T(n)\mathbf{v}(n) - \alpha r(n) \tag{14}$$

$$\mathbf{U}^T(n)\mathbf{v}(n+1) = \mathbf{U}^T(n)\mathbf{v}(n) \tag{15}$$

Eqs. (14) and (15) yield $\mathbf{U}^T(n)\mathbf{v}(n+1) = (1-\alpha)\mathbf{U}^T(n)\mathbf{v}(n) - \alpha\mathbf{r}(n-1)$, where $\mathbf{r}(n-1) = [r(n-1) \ \ldots \ r(n-P)]^T$. Using these properties, $e_a(n)$ results in

$$\begin{aligned} e_a(n) &= \mathbf{u}^T(n)\mathbf{v}(n) = \mathbf{\Phi}^T(n)\mathbf{v}(n) + \hat{\mathbf{a}}^T(n)\mathbf{U}^T(n)\mathbf{v}(n) \\ &= \mathbf{\Phi}^T(n)\mathbf{v}(n) + \hat{\mathbf{a}}^T(n)[(1-\alpha)\mathbf{U}^T(n)\mathbf{v}(n) \\ &\quad - \alpha\mathbf{r}(n-1)] \end{aligned} \tag{16}$$

Defining $\epsilon(n) = \mathbf{U}^T(n)\mathbf{v}(n)$ and assuming that $\hat{\mathbf{a}}(n) \approx \mathbf{a}$ [6], (16) can be written as

$$e_a(n) = \mathbf{\Phi}^T(n)\mathbf{v}(n) + (1-\alpha)\mathbf{a}^T\epsilon(n) - \alpha\mathbf{a}^T\mathbf{r}(n-1) \tag{17}$$

Now,

$$\mathbf{a}^T\epsilon(n) = \sum_{i=1}^{P} a_i \sum_{k=0}^{N-1} u(n-k-i)v_k(n) \tag{18}$$

can be approximated by

$$\begin{aligned} \mathbf{a}^T\epsilon(n) &\approx \sum_{i=1}^{P} a_i \sum_{k=0}^{N-1} u(n-k)v_k(n) \\ &= \Big(\sum_{i=1}^{P} a_i\Big)e_a(n) \end{aligned} \tag{19}$$

if it is assumed that $u(n-\ell) \approx u(n-\ell-1) \approx \cdots \approx u(n-\ell-P)$, i.e., that the sequence $\{u(n)\}$ presents small variations in P consecutive samples. This assumption is reasonable for highly correlated input signals and small P. Both conditions are usually satisfied in applications where the AP algorithm is employed.

Using (19), (17) can be rewritten as

$$e_a(n) = \frac{1}{1-(1-\alpha)\sum_{i=1}^{P} a_i}\left[\mathbf{\Phi}^T(n)\mathbf{v}(n) - \alpha\sum_{i=1}^{P} a_i r(n-i)\right] \tag{20}$$

Substituting (20) into (13), results in

$$\begin{aligned} \mathbf{v}(n+1) = \mathbf{v}(n) &- \left[\frac{\alpha}{1-(1-\alpha)\sum_{i=1}^{P} a_i}\right]\frac{\mathbf{\Phi}(n)\mathbf{\Phi}^T(n)}{\mathbf{\Phi}^T(n)\mathbf{\Phi}(n)}\mathbf{v}(n) \\ &+ \alpha\frac{\mathbf{\Phi}(n)}{\mathbf{\Phi}^T(n)\mathbf{\Phi}(n)}r_a(n) \end{aligned} \tag{21}$$

where $r_a(n)$ is the filtered noise sequence [3]

$$r_a(n) = r(n) - \alpha\frac{\sum_{i=1}^{P} a_i r(n-i)}{1-(1-\alpha)\sum_{i=1}^{P} a_i} \tag{22}$$

Under assumption **A2** and noting that $E\{\mathbf{\Phi}(n)r_a(n)\} = 0$ because $r(n)$ is zero mean and independent of any other signal, the expected value of (21) yields

$$\begin{aligned} E\{\mathbf{v}(n+1)\} = &\left(1 - \left[\frac{\alpha}{1-(1-\alpha)\sum_{i=1}^{P} a_i}\right]\right. \\ &\left.\times E\left\{\frac{\mathbf{\Phi}(n)\mathbf{\Phi}^T(n)}{\mathbf{\Phi}^T(n)\mathbf{\Phi}(n)}\right\}\right)E\{\mathbf{v}(n)\} \end{aligned} \tag{23}$$

Each element of the expectation in the r.h.s. of (23) has a numerator given by $\phi(n-i)\phi(n-j)$ and a denominator given by $\sum_{k=0}^{N-1}\phi^2(n-k)$. Since the components of $\mathbf{\Phi}(n)$ in the numerator affect only two out of N terms in the denominator, numerator and denominator can be assumed weakly correlated for large values of N. For ergodic inputs, this approximation is equivalent to apply the averaging principle [7], as $\mathbf{\Phi}^T(n)\mathbf{\Phi}(n)$ tends to be slowly varying when compared to $\phi(n-i)\phi(n-j)$ for large values of N. Hence, the following approximation is used:

$$\begin{aligned} &E\{[\mathbf{\Phi}^T(n)\mathbf{\Phi}(n)]^{-1}\mathbf{\Phi}(n)\mathbf{\Phi}^T(n)\} \\ &\approx E\{[\mathbf{\Phi}^T(n)\mathbf{\Phi}(n)]^{-1}\}\mathbf{R}_{\phi\phi} \end{aligned} \tag{24}$$

where $\mathbf{R}_{\phi\phi}$ is given by (11).

The expected value of $E\{[\mathbf{\Phi}^T(n)\mathbf{\Phi}(n)]^{-1}\}$ is determined using the assumption that $\mathbf{\Phi}(n)$ is Gaussian distributed and neglecting the statistical dependence between its components (recall that they are estimates of a white sequence). Thus, $y = \mathbf{\Phi}^T(n)\mathbf{\Phi}(n)$ has a chi-square distribution with $G = N - P$ degrees of freedom. The value of G arises from the statistical properties of $\mathbf{\Phi}(n)$ determined in the previous section. Thus, [8]

$$f_y(y) = \frac{1}{2^{G/2}\sigma_\phi^G\Gamma(\frac{G}{2})}y^{(G/2)-1}e^{-y/2\sigma_\phi^2}u(y) \tag{25}$$

Determining the expected value in (24) through integration yields

$$E\{[\mathbf{\Phi}^T(n)\mathbf{\Phi}(n)]^{-1}]\} = \frac{1}{\sigma_\phi^2(G-2)} \tag{26}$$

where $\sigma_\phi^2 = \sigma_z^2(N-P)/N$. Using (26) in (23) leads to:

$$\begin{aligned} E\{\mathbf{v}(n+1)\} = &\left(1 - \frac{\alpha}{\sigma_\phi^2(G-2)[1-(1-\alpha)\sum_{i=1}^{P} a_i]}\right)\mathbf{R}_{\phi\phi} \\ &\times E\{\mathbf{v}(n)\} \end{aligned} \tag{27}$$

which is the recursion for the mean weight error vector.

VI. Mean Square Error Behavior

This section presents the main steps in the determination of the mean square error behavior. Details of some evaluations are supressed due to space limitations.

Squaring (4) and taking the expected value yields

$$\begin{aligned} E\{e^2(n)\} = \sigma_r^2\Bigg[1 &+ \frac{\alpha^2}{[1-(1-\alpha)\sum_{i=1}^{P} a_i]^2} \\ &\times\left(\mathbf{a}^T\mathbf{a} + \sigma_z^2 tr[E\{[\mathbf{U}^T(n)\mathbf{U}(n)]^{-1}\}]\right)\Bigg] \\ &+ tr[\mathbf{R}_{\Phi\Phi}\mathbf{K}(n)] \end{aligned} \tag{28}$$

where $\mathbf{K}(n) = E\{\mathbf{v}(n+1)\mathbf{v}^T(n+1)\}$ is the weight-error correlation matrix.

Postmultiplying (21) by its transpose and taking the expected value, yields:

$$
\begin{aligned}
&E\{\mathbf{v}(n+1)\mathbf{v}^T(n+1)\} = E\{\mathbf{v}(n)\mathbf{v}^T(n)\} \\
&- \frac{\alpha}{1-(1-\alpha)\sum_{i=1}^{P} a_i} E\left\{ \mathbf{v}(n)\mathbf{v}^T(n) \frac{\mathbf{\Phi}(n)\mathbf{\Phi}^T(n)}{\mathbf{\Phi}^T(n)\mathbf{\Phi}(n)} \right\} \\
&- \alpha E\left\{ \mathbf{v}(n) \frac{r_a(n)\mathbf{\Phi}^T(n)}{\mathbf{\Phi}^T(n)\mathbf{\Phi}(n)} \right\} \\
&- \frac{\alpha}{1-(1-\alpha)\sum_{i=1}^{P} a_i} E\left\{ \frac{\mathbf{\Phi}(n)\mathbf{\Phi}^T(n)}{\mathbf{\Phi}^T(n)\mathbf{\Phi}(n)} \mathbf{v}(n)\mathbf{v}^T(n) \right\} \\
&+ \frac{\alpha^2}{[1-(1-\alpha)\sum_{i=1}^{P} a_i]^2} \\
&\times E\left\{ \frac{\mathbf{\Phi}(n)\mathbf{\Phi}^T(n)}{\mathbf{\Phi}^T(n)\mathbf{\Phi}(n)} \mathbf{v}(n)\mathbf{v}^T(n) \frac{\mathbf{\Phi}(n)\mathbf{\Phi}^T(n)}{\mathbf{\Phi}^T(n)\mathbf{\Phi}(n)} \right\} \\
&+ \frac{\alpha^2}{1-(1-\alpha)\sum_{i=1}^{P} a_i} E\left\{ \frac{\mathbf{\Phi}(n)\mathbf{\Phi}^T(n)}{\mathbf{\Phi}^T(n)\mathbf{\Phi}(n)} \mathbf{v}(n) \frac{r_a(n)\mathbf{\Phi}^T(n)}{\mathbf{\Phi}^T(n)\mathbf{\Phi}(n)} \right\} \\
&- \alpha E\left\{ \frac{\mathbf{\Phi}(n) r_a(n)}{\mathbf{\Phi}^T(n)\mathbf{\Phi}(n)} \mathbf{v}^T(n) \right\} \\
&+ \frac{\alpha^2}{1-(1-\alpha)\sum_{i=1}^{P} a_i} E\left\{ \frac{\mathbf{\Phi}(n) r_a(n)}{\mathbf{\Phi}^T(n)\mathbf{\Phi}(n)} \mathbf{v}^T(n) \frac{\mathbf{\Phi}(n)\mathbf{\Phi}^T(n)}{\mathbf{\Phi}^T(n)\mathbf{\Phi}(n)} \right\} \\
&+ \alpha^2 E\left\{ \frac{\mathbf{\Phi}(n) r_a(n)}{\mathbf{\Phi}^T(n)\mathbf{\Phi}(n)} \frac{r_a(n)\mathbf{\Phi}^T(n)}{\mathbf{\Phi}^T(n)\mathbf{\Phi}(n)} \right\}
\end{aligned}
\tag{29}
$$

The expected values in (29) are evaluated using assumptions **A1** and **A2**, and the same considerations used to evaluate the expected values in (23). Neglecting the statistical dependence of $\mathbf{\Phi}(n)$ and $\mathbf{v}(n)$ and assuming $\mathbf{\Phi}(n)$ and $\mathbf{v}(n) r_a(n)$ uncorrelated, the third and seventh terms in the r.h.s. of (29) are equal to zero. The sixth and eighth terms lead to third order moments of zero-mean Gaussian variates (components of $\mathbf{\Phi}(n)$). Using the properties of cumulants of order three for Gaussian variables [9], it is easy to show that these terms are also equal to zero.

The fifth term in (29) is approximated by

$$
\begin{aligned}
&E\left\{ \frac{\mathbf{\Phi}(n)\mathbf{\Phi}^T(n)}{\mathbf{\Phi}^T(n)\mathbf{\Phi}(n)} \mathbf{v}(n)\mathbf{v}^T(n) \frac{\mathbf{\Phi}(n)\mathbf{\Phi}^T(n)}{\mathbf{\Phi}^T(n)\mathbf{\Phi}(n)} \right\} \\
&= E\left\{ \frac{1}{[\mathbf{\Phi}^T(n)\mathbf{\Phi}(n)]^2} \right\} \\
&\quad \times E\{\mathbf{\Phi}(n)\mathbf{\Phi}^T(n)\mathbf{v}(n)\mathbf{v}^T(n)\mathbf{\Phi}(n)\mathbf{\Phi}^T(n)\}
\end{aligned}
\tag{30}
$$

The first expectation in (30) is evaluated by integration using (25), where $y = \mathbf{\Phi}^T(n)\mathbf{\Phi}(n)$. Thus,

$$
\frac{1}{E\{[\mathbf{\Phi}^T(n)\mathbf{\Phi}(n)]^2\}} = \frac{1}{\sigma_\phi^4 (G^2+2G)} \tag{31}
$$

To evaluate the second expectation, note that its (i,j)-th element can be written as:

$$
tr[\mathbf{\Phi}(n)\mathbf{\Phi}^T(n)\mathbf{v}(n)\mathbf{v}^T(n)] \cdot \phi_i \phi_j = \sum_{l=0}^{N-1} \sum_{k=0}^{N-1} \phi_l \phi_k v_k v_l \phi_i \phi_j \tag{32}
$$

Assuming, as done before, that $\phi_i \phi_j$ is weakly correlated with the trace in (32), which is a function of all ϕ_k, and noting that $E\{\phi_\ell \phi_k\} = 0$ for $\ell \neq k$, (32) simplifies to

$$
\begin{aligned}
&E\{\mathbf{\Phi}(n)\mathbf{\Phi}^T(n)\mathbf{v}(n)\mathbf{v}^T(n)\mathbf{\Phi}(n)\mathbf{\Phi}^T(n)\} = \\
&E\left\{ \sum_{k=0}^{N-1} \left[\phi_k^2 \cdot E\{v_k^2\} \right] \right\} \cdot E\{\phi_i \phi_j\}
\end{aligned}
\tag{33}
$$

$E\{v_k^2\}$ is then determined by expanding $\sigma_{v_k}^2 = E\{(v_k - E\{v_k\})^2\} = E\{v_k^2\} - E^2\{v_k\}$. Then, assuming that all v_k have equal variances $\sigma_{v_k}^2 = tr[\mathbf{C}(n)]/N$ where $\mathbf{C}(n) = \mathbf{K}(n) - E\{\mathbf{v}(n)\}E\{\mathbf{v}^T(n)\}$ is the covariance matrix of $\mathbf{v}(n)$, yields:

$$
E\{v_k^2\} = \frac{1}{N}\left[tr[\mathbf{K}(n)] - E\{\mathbf{v}^T(n)\}E\{\mathbf{v}(n)\} \right] + E^2\{v_k\}, \quad \forall k \tag{34}
$$

Inserting (34) into (33) and noting that $\mathbf{\Phi}(n)$ is zero-mean, white and independent of $\mathbf{v}(n)$, yields:

$$
\begin{aligned}
&E\left\{ \sum_{k=0}^{N-1} \left[\phi_k^2 \cdot E\{v_k^2\} \right] \right\} \cdot E\{\phi_i \phi_j\} = \\
&\sigma_\phi^2 \cdot \left[\frac{G}{N} \cdot tr[\mathbf{K}(n)] + \left(1 - \frac{G}{N}\right) \cdot E\{\mathbf{v}^T(n)\} \cdot E\{\mathbf{v}(n)\} \right] \mathbf{R}_{\phi\phi}
\end{aligned}
\tag{35}
$$

The last expectation in (29) is determined using the uncorrelatedness of $\mathbf{\Phi}(n)$ and $r_a(n)$:

$$
\begin{aligned}
&E\{r_a^2(n)\} \cdot E\left\{ \frac{\mathbf{\Phi}(n)\mathbf{\Phi}^T(n)}{[\mathbf{\Phi}^T(n)\mathbf{\Phi}(n)]^2} \right\} = \\
&\frac{\sigma_r^2}{\sigma_\phi^4 (G-2)(G-4)} \left[1 + \frac{\alpha^2}{[1-(1-\alpha)\sum_{i=1}^{P} a_i]^2} \right. \\
&\left. \times \left(\mathbf{a}^T\mathbf{a} + \sigma_z^2 tr[E\{[\mathbf{U}^T(n)\mathbf{U}(n)]^{-1}\}] \right) \right] \mathbf{R}_{\phi\phi}
\end{aligned}
\tag{36}
$$

Finally, the second and fourth terms are approximated by $\mathbf{K}(n)\mathbf{R}_{\phi\phi}$ and $\mathbf{R}_{\phi\phi}\mathbf{K}(n)$, respectively.

Puting all the above results together in (29) yields a recursion for $K(n)$:

$$
\begin{aligned}
&\mathbf{K}(n+1) = \mathbf{K}(n) \\
&- \frac{\alpha}{\sigma_\phi^2 [1-(1-\alpha)\sum_{i=1}^{P} a_i](G-2)} [\mathbf{K}(n)\mathbf{R}_{\phi\phi} + \mathbf{R}_{\phi\phi}\mathbf{K}(n)] \\
&+ \frac{\alpha^2}{\sigma_\phi^2 [1-(1-\alpha)\sum_{i=1}^{P} a_i]^2 (G^2+2G)} \\
&\times \left[\frac{G}{N} tr[\mathbf{K}(n)] + \left(1 - \frac{G}{N}\right) E\{\mathbf{v}^T(n)\}E\{\mathbf{v}(n)\} \right] \mathbf{R}_{\phi\phi} \\
&+ \frac{\alpha^2 \sigma_r^2}{\sigma_\phi^4 (G-2)(G-4)} \left[1 + \frac{\alpha^2}{[1-(1-\alpha)\sum_{i=1}^{P} a_i]^2} \right. \\
&\left. \times \left(\mathbf{a}^T\mathbf{a} + \sigma_z^2 tr[E\{[\mathbf{U}^T(n)\mathbf{U}(n)]^{-1}\}] \right) \right] \mathbf{R}_{\phi\phi}
\end{aligned}
\tag{37}
$$

The matrix $E\{[\mathbf{U}^T(n)\mathbf{U}(n)]^{-1}\}$ must be numerically estimated from the input signal $u(n)$. For adaptive filters of sufficient order, however, it is reasonable to assume $\hat{\mathbf{a}}(n) \approx \mathbf{a}$, and

the term $\sigma_z^2 \text{ tr}\left[E\{[\mathbf{U}^T(n)\mathbf{U}(n)]^{-1}\}\right]$ can be neglected in both (28) and (37), when compared with $\mathbf{a}^T\mathbf{a}$. The resulting model does not require any matrix estimation and is as accurate as the complete model for most practical purposes. Also, notice that the new model coincides with the model derived in [11] when $\alpha = 1$.

VII. Simulations

This section presents simulations to verify the accuracy of the analytical models given by equations (27), (28) and (37). Several simulations have been realized using the derived models. The examples presented here are representative of the results obtained. In all cases, the term $\sigma_z^2 \text{ tr}\left[E\{[\mathbf{U}^T(n)\mathbf{U}(n)]^{-1}\}\right]$ has been neglected in both (28) and (37). In the examples, AR(P) means an autoregressive process of order P, and AP($P+1$) means the AP algorithm of order $P+1$ (using P input vectors in $\mathbf{U}(n)$). The signal-to-noise ratio of the adaptive system is defined as $SNR = 10\log(\sigma_u^2/\sigma_r^2)$dB. The system to be identified has impulse response represented by the vector $\mathbf{w}^o$ and normalized so that $\mathbf{w}^{o^T}\mathbf{w}^o = 1$.

Figs. 1 and 2 show the mean weight behavior for some of the weights. The remaining weights have similar behavior. A very good match between simulation and theoretical predictions can be verified from these plots.

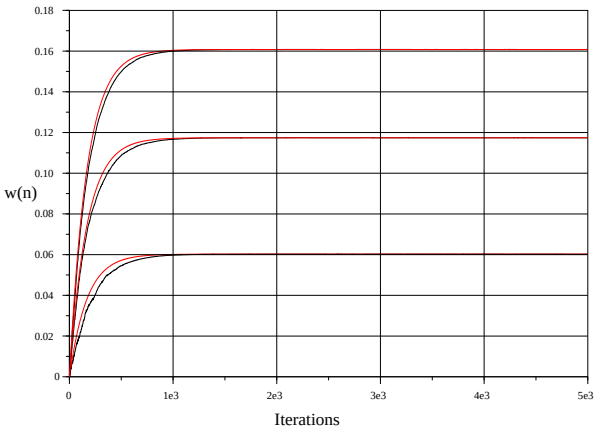

Fig. 1. Mean weight behavior for some of the weights. Monte Carlo simulations (400 runs) (black); the analytical model proposed (red). $\alpha = 0.5$, $N = 64$ and $P = 4$.

Figs. 3 and 4 show that the steady-state error increases with the step size α. Comparing Figs. 3 and 5, whose simulations differ only by the value of P, the steady-state error does not show noticeable improvement by increasing P from 4 to 8. Theresults in Figs. 6 and 7 highlight the effect of α on the tradeoff between convergence speed and steady-state misadjustment. Notice that there is excellent agreement between simulations and analytical predictions in all cases, both during the adaptation phase and in steady-state.

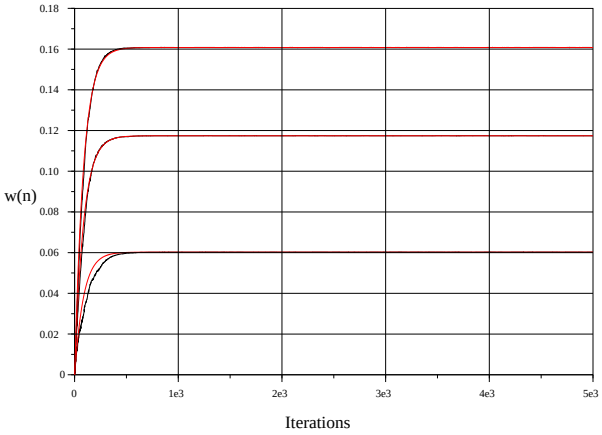

Fig. 2. Mean weight behavior for some of the weights. Monte Carlo simulations (400 runs) (black); the analytical model proposed (red). $\alpha = 0.8$, $N = 64$ and $P = 4$.

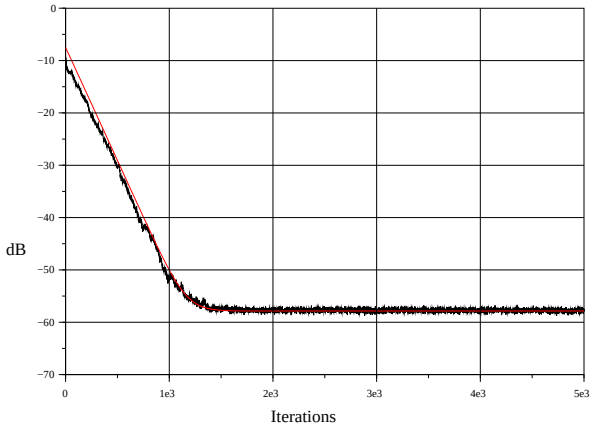

Fig. 3. MSE: Comparisons between (black) Monte Carlo simulations (400 runs); (red) the analytical model proposed. $\alpha = 0.5$, $N = 64$ and $P = 4$.

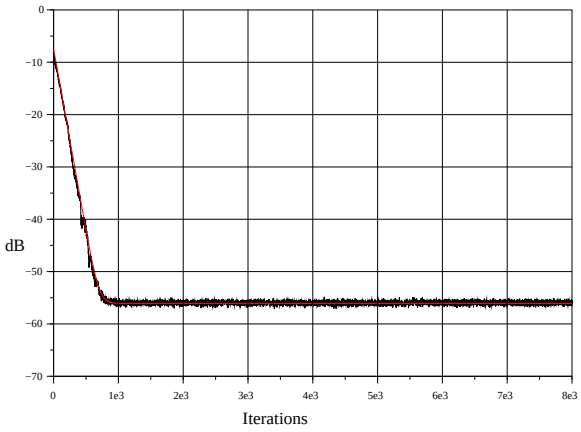

Fig. 4. MSE: Comparisons between (black) Monte Carlo simulations (400 runs); (red) the analytical model proposed. $\alpha = 0.8$, $N = 64$ and $P = 4$.

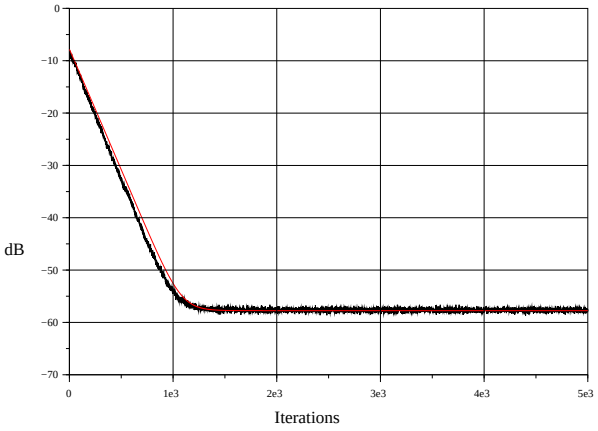

Fig. 5. MSE: Comparisons between (black) Monte Carlo simulations (400 runs); (red) the analytical model proposed. $\alpha = 0.5$, $N = 64$ and $P = 8$.

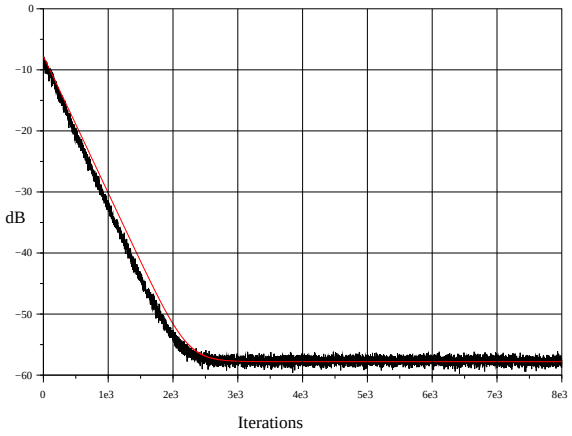

Fig. 6. MSE: Comparisons between (black) Monte Carlo simulations (200 runs); (red) the analytical model proposed. $\alpha = 0.5$, $N = 128$ and $P = 16$.

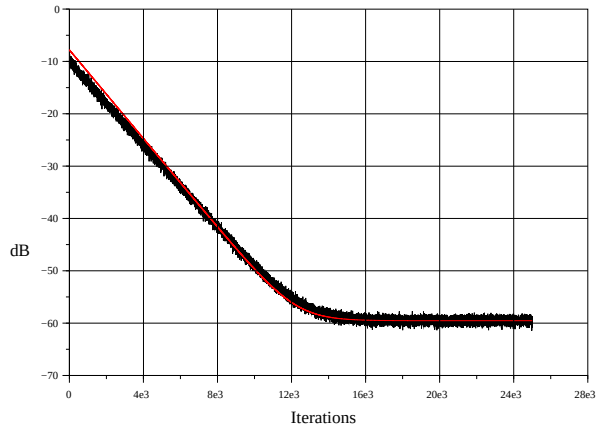

Fig. 7. MSE: Comparisons between (black) Monte Carlo simulations (200 runs); (red) the analytical model proposed. $\alpha = 0.1$, $N = 128$ and $P = 16$.

VIII. Summary

This paper has presented a analytical model for predicting the behavior of AP algorithm for step size $\alpha \leq 1$. Deterministic recursive equations were derived for the mean weight and the mean square error behaviors for a large number of adaptive taps N, as compared to the order P of the algorithm. Simulation results have shown excellent agreement with theorical predictions during the adaptation phase and in steady-state.

References

[1] S. Haykin, *Adaptive Filter Theory* second edition, Prentice-Hall, 1991.

[2] K. Ozeki, T. Umeda,"An Adaptive Filtering Algorithm Using Orthogonal Projection to an Affine Subspace and Its Properties,"*Eletronics and Communications in Japan*, vol.67-A, No.5, pp. 19-27, February 1984.

[3] M. Rupp,"A Family of Filter Algorithms with Decorrelating Properties," *IEEE Trans. on Signal Processing*, vol.46 , No.3, pp.771-775, March 1998.

[4] S. G. Sankaran, A. A. (Louis) Beex,"Convergence Behavior of Affine Projection Algorithms,"*IEEE Trans. on Signal Processing*, vol.48, No.4, pp.1086-1096, April 2000.

[5] D. T. M. Slock, "On the Convergence Behavior of the LMS and the Normalized LMS Algorithms," *IEEE Trans. on Signal Processing*, vol.41 , No. 9, pp. 2811-2825, September 1993.

[6] N.J. Bershad, D. Linebarger, S. McLaughlin,"A Stochastic Analysis of the Affine Projection Algorithm for Gaussian Autoregressive Inputs,"*Acoustics, Speech, and Signal Processings*. 2001 *IEEE International Conference on*, vol.6, pp.3837-3840, 2001.

[7] C. Samson, V.U. Reddy, "Fixed point error analysis of the Normalized Ladder Algorithms. *IEEE Trans. on Acoustic, Speech and Signal Processing*, vol.31, No.5, pp.1177-1191, October 1983.

[8] A. Papoulis, *Probability, Random Variables, and Stochastic Processs*, first edition, 1965.

[9] C.L. Nikias, A.P. Petropulu, *Higher-Order Spectra Analysis*-A Nonlinear Signal Processing Framework, Prentice-Hall, 1993.

[10] S. J. M. Almeida, J. C. M. Bermudez, "A New Stochastic Analysis of Affine Projection Algorithm for Gaussian Inputs and Large Number of Coefficients," *International Telecommunications Symposium-ITS2002*, September 2002. Natal, Brazil.

[11] S. J. M. Almeida, J. C. M. Bermudez, N. J. Bershad and M. H. Costa, "A Stochastic Model for the Convergence Behavior of the Affine Projection Algorithm for Gaussian Inputs," *International Conference on Acoustics, Speech, and Signal Processing - ICASSP -2003*, April 2003. Hong Kong, China.

[12] D. L. Duttweiler, "Proportionate Normalized Least-Mean-Squares Adaptation in Echo Cancelers," *IEEE Trans. on Acoustic, Speech and Signal Processing*, vol.8, No.5, pp.508-518, September 2000.