A New Stochastic Analysis of Affine Projection Algorithm for Gaussian Inputs and Large Number of Coefficients

Sérgio J. M. de Almeida* and José Carlos M. Bermudez**

*Universidade Católica de Pelotas, Pelotas-RS, Brazil

**Universidade Federal de Santa Catarina, Florianópolis, SC, Brazil

Abstract—**This paper presents a new analytical model for predicting the behavior of Affine Projection (AP) algorithm. The model is derived for autoregressive (AR) Gaussian inputs and for unity step size (fastest convergence). Deterministic recursive equations are determined 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 are presented which show excellent agreement between theory an simulations in steady-state, and fair to good agreement during transient. The model predictions for the transient phase improve as N/P increases. These caracteristics are of special interest in applications such as acoustic echo cancellation.**

I. INTRODUCTION

The least mean squares (LMS) algorithm and its normalized version (NLMS) are among the most employed in adaptive signal processing applications. The performance of these algorithms tends to be insufficient, however, when the signals handled are highly correlated and the required number of adaptive taps is large [1]. One important application with such characteristics is acoustic echo cancellation. A solution for the convergence problems of these algorithms was proposed by Ozeki and Umeda in 1984 [2]; the Affine Projection (AP) algoritm. The AP algorithm applies weight updates in directions that are orthogonal to the last P input vectors. This allows decorrelation of an AR(P) input process, speeding up convergence [3]. Thus, the AP algorithm is a better alternative than LMS or NLMS for applications with highly correlated input signals [4]. The price paid for the best performance is the larger computational complexity. This difficulty becomes less important as more advanced of the semiconductor elements are introduced. Complex algorithms have recently become feasible for applications such as echo cancellation, channel equalization and noise cancellation.

This recent practical applicability has triggered the interest in the analysis of the AP algorithm behavior. Quantitative statistical analysis, however, presents analytical difficulties caused by the complexity of the underdetermined least squares solution embedded in the algorithm. Reference [4] presents a quantitative analysis of the AP algorithm. This analysis is based on a specific model for the input signal originally proposed in [5]. More recenlty, [6] presented a quantitative analysis for autoregressive Gaussian inputs. This analysis follows the work in [3] with the solution of a recursion for the weight error vector variances. The solution uses previous results obtained for the NLMS algorithm with white inputs.

This paper presents a new statistical analysis for the behavior of the AP algorithm for Gaussian autoregressive (AR) inputs. The analytical difficulties are circunvented by assuming a large number of adaptive taps, as compared to the order of the AP algorithm. This condition allows the use of an assumption similar to the independence assumption used to analyze many adaptive algorithms, including the AP algorithm [1], [6]. The analysis assuming long adaptive filters and AR inputs is compatible with applications such as acoustic echo cancellation, where speech signals and long impulse responses are typical. Analytical recursions are derived which predict the behavior of the mean weight vector and the mean square error (MSE). Simulation results show excellent agreement with theoretical predictions in steady-state, and fair to good agreement during the adaptation phase. The new model is conservative and thus suitable for design purposes.

Sérgio J. M. de Almeida is with the 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. E-mails: smelo@eel.ufsc.br, j.bermudez@ieee.org. This work was partially supported by CNPq Grant No. 352084/92-8

II. THE INPUT SIGNAL MODEL

Let $\mathbf{u}(n)$ be 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 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 $\mathbf{z}(n) = [z(n) \ z(n-1) \ \ldots \ z(n-N+1)]^T$ is a vector with samples from a stationary white Gaussian process with variance σ_z^2. The least square estimate of the parameters of $\mathbf{a}$ is

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

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

III. THE AFFINE PROJECTION ALGORITHM

The update equation of the AP algorithm is given by:

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

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

$$e(n) = \mathbf{W}_o^T\mathbf{u}(n) + r(n) - \mathbf{w}^T(n)\mathbf{u}(n) \qquad (4)$$

where $\mathbf{W}_o = [W_{o_0} \ W_{o_1} \ \ldots \ W_{o_{N-1}}]^T$ is the unknown impulse response and $r(n)$ is a white additive observation noise with variance σ_r^2, independent of any other signal. $\mathbf{w}(n) = [w_0(n)\ w_1(n)\ \ldots\ w_{N-1}(n)]^T$ is the adaptive weight vector. The step-size α is assumed unity, which corresponds to maximum convergence speed. The vector $\mathbf{\Phi}(n)$ defines the direction of update, and is given by:

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

Substituting (1) in (5) yiels

$$\mathbf{\Phi}(n) = [I - \mathbf{P}_u(n)]\mathbf{z}(n) = \mathbf{P}_o(n)\mathbf{z}(n) \tag{6}$$

where, $\mathbf{P}_u(n) = \mathbf{U}(n)[\mathbf{U}^T(n)\mathbf{U}(n)]^{-1}\mathbf{U}(n)^T$ is the projection matrix onto the subspace spanned by the columns of $\mathbf{U}(n)$ and $\mathbf{P}_o(n) = I - \mathbf{P}_u(n)$ is the projection matrix onto the orthogonal subspace. The number $(P+1)$ of input vectors used to determine $\mathbf{\Phi}(n)$ defines the order of the AP algorithm.

IV. Analysis

A. Assumptions

In the following analysis, two assumptions are used which are similar to the independence assumption used to analyze many stochastic algorithms [1].

Assumption A1: It is assumed that the statistical dependence between $\mathbf{z}(n)$ and $\mathbf{U}(n)$ can be neglected. This assumption is more realistic for $N \gg P$, and is justified as follows. Eq. (1) shows that $\mathbf{z}(n)$ has an algebraic dependence with the P vectors $\mathbf{u}(n-1)$, …, $\mathbf{u}(n-P)$. Also, $\mathbf{z}(n)$ is of dimension N. Now, $\mathbf{z}(n)$ 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 $\mathbf{z}_\perp(n) = \mathbf{P}_o(n)\mathbf{z}(n)$. Thus, only $\mathbf{z}_u(n)$ is algebraically dependent of $\mathbf{U}(n)$. Moreover, it is reasonable to assume that $\mathbf{z}(n)$ has equal amounts of energy in each direction of the N-dimensional space, since it is drawn from a white process. Thus, only the energy of the N-dimensional $\mathbf{z}(n)$ which is projected onto the P-dimensional subspace defined by $\mathbf{U}(n)$ is responsible for the dependence between $\mathbf{z}(n)$ and $\mathbf{U}(n)$. If $N \gg P$, this dependence can be neglected. This is usually the case in systems with long impulse responses, since P tends to be limited by algorithm complexity considerations.

Assumption A2: Rupp has shown in [3] that $\mathbf{\Phi}(n)$ is a vector whose elements are estimates of the white noise sequence $z(n)$. Based on this property, it is assumed that $\mathbf{\Phi}(n)$ and the weight vector $\mathbf{w}(n)$ are statistically independent. Even though this assumption is not correct for delay line implementations, it has led to very useful results in the analysis of many stochastic adaptive filtering algorithms [1].

B. Statistical Properties of $\mathbf{\Phi}(n)$

Using assumption **A1** and (6), it is easy to show that

$$\mathbf{R}_{\phi\phi} = E\{\mathbf{\Phi}(n)\mathbf{\Phi}^T(n)\} = \sigma_z^2 E\{\mathbf{P}_o(n)\} \tag{7}$$

From (6), each element of $\mathbf{\Phi}(n)$ is determined by the inner product of a line of $\mathbf{P}_o(n)$ and $\mathbf{z}(n)$. Thus each component $\phi(n-i)$ is composed by the sum of N random variables $\{\mathbf{P}_o\}_{ij} z(n-j)$. From assumption **A1** and the fact that $z(n)$ is white, each of these random variables can be assumed independent. Thus, by the Central Limit Theorem, $\mathbf{\Phi}(n)$ has a Gaussian distribution.

C. Mean Weight Behavior

Defining the weight error vector, $\mathbf{v}(n) = \mathbf{w}(n) - \mathbf{W}_o$ and using (4), (3) can be written for $\alpha = 1$ as

$$\begin{aligned}\mathbf{v}(n+1) = \mathbf{v}(n) &- \frac{\mathbf{\Phi}(n)\mathbf{u}^T(n)}{\mathbf{\Phi}^T(n)\mathbf{\Phi}(n)}\mathbf{v}(n) \\ &+ \frac{\mathbf{\Phi}(n)}{\mathbf{\Phi}^T(n)\mathbf{\Phi}(n)} r(n)\end{aligned} \tag{8}$$

Pre-multiplying (8) by $\mathbf{\Phi}^T(n)$, $\mathbf{u}^T(n)$ and $\mathbf{U}^T(n)$ yields [3]

$$\begin{aligned}&\mathbf{\Phi}^T(n)\mathbf{v}(n+1) = \mathbf{\Phi}^T(n)\mathbf{v}(n) - \mathbf{u}^T(n)\mathbf{v}(n) + r(n) \\ &\mathbf{u}(n)\mathbf{v}(n+1) = r(n) \\ &\mathbf{U}^T(n)\mathbf{v}(n+1) = \mathbf{U}^T(n)\mathbf{v}(n)\end{aligned} \tag{9}$$

The last two properties yield $\mathbf{U}^T(n)\mathbf{v}(n+1) = \mathbf{r}(n-1)$, where $\mathbf{r}(n-1) = [r(n-1)\ \ldots\ r(n-P)]^T$.

Using these properties and (5) leads to

$$\begin{aligned}\mathbf{v}(n+1) = \mathbf{v}(n) &- \frac{\mathbf{\Phi}(n)\mathbf{\Phi}^T(n)}{\mathbf{\Phi}^T(n)\mathbf{\Phi}(n)}\mathbf{v}(n) \\ &+ \frac{\mathbf{\Phi}(n)}{\mathbf{\Phi}^T(n)\mathbf{\Phi}(n)} r_a(n)\end{aligned} \tag{10}$$

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

$$r_a(n) = r(n) - \sum_{i=1}^{P} \hat{a}_i(n) r(n-i) \tag{11}$$

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 (10) yields

$$\begin{aligned}E\{\mathbf{v}(n+1)\} &= E\{\mathbf{v}(n)\} \\ &- E\left\{\frac{\mathbf{\Phi}(n)\mathbf{\Phi}^T(n)}{\mathbf{\Phi}^T(n)\mathbf{\Phi}(n)}\right\} E\{\mathbf{v}(n)\}\end{aligned} \tag{12}$$

Each element of the expectation in the r.h.s. of (12) 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. This is equivalent to applying the averaging principle proposed in [7], as $\mathbf{\Phi}^T(n)\mathbf{\Phi}(n)$ tends to be slowly varying when compared to $\phi(n-1)\phi(n-j)$ for large values of N. Hence, the following approximation is used:

$$\begin{aligned}&E\{[\mathbf{\Phi}^T(n)\mathbf{\Phi}(n)]^{-1}\mathbf{\Phi}(n)\mathbf{\Phi}^T(n)\} \\ &\approx E\{[\mathbf{\Phi}^T(n)\mathbf{\Phi}(n)]^{-1}\}\mathbf{R}_{\phi\phi}\end{aligned} \tag{13}$$

where $\mathbf{R}_{\phi\phi}$ is given by (7).

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 N degrees of freedom. Its probability density function is given by [8]

$$f_y(y) = \frac{1}{2^{N/2}\sigma_\phi^N \Gamma(\frac{N}{2})} y^{(N/2)-1} e^{-y/2\sigma_\phi^2} u(y) \tag{14}$$

Determining the expected value in (13) through integration yields

$$E\{[\mathbf{\Phi}^T(n)\mathbf{\Phi}(n)]^{-1}]\} = \frac{\Gamma(N/2-1)}{2\sigma_\phi^2\Gamma(N/2)} \quad (15)$$

Since $\Gamma(n+1) = n\Gamma(n)$, (15) simplifies to:

$$E\{[\mathbf{\Phi}^T(n)\mathbf{\Phi}(n)]^{-1}]\} = \frac{1}{(N-2)\sigma_\phi^2} \quad (16)$$

where σ_ϕ^2 is a diagonal element of $\mathbf{R}_{\phi\phi}$. Using (16) in (12) leads to:

$$E\{\mathbf{v}(n+1)\} = \left[I - \frac{1}{\sigma_\phi^2(N-2)}\mathbf{R}_{\phi\phi}\right]E\{\mathbf{v}(n)\} \quad (17)$$

which is the recursion for the mean weight error vector.

V. Mean Square Error Behavior

Squaring (4) and taking the expected value,

$$\begin{aligned} &E\{e^2(n)\} = \\ &+\left(1 + \mathbf{a}^T\mathbf{a} + \sigma_z^2 tr\left[E\{[\mathbf{U}^T(n)\mathbf{U}(n)]^{-1}\}\right]\right)\sigma_r^2 \\ &+ tr[\mathbf{R}_{\mathbf{\Phi}\mathbf{\Phi}}\mathbf{K}_{vv}(n)] \end{aligned} \quad (18)$$

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

Postmultiplying (10) by its transpose and taking the expected value, yields:

$$\begin{aligned} &E\{\mathbf{v}(n+1)\mathbf{v}^T(n+1)\} = E\{\mathbf{v}(n)\mathbf{v}^T(n)\} \\ &- E\left\{\mathbf{v}(n)\mathbf{v}^T(n)\frac{\mathbf{\Phi}(n)\mathbf{\Phi}^T(n)}{\mathbf{\Phi}^T(n)\mathbf{\Phi}(n)}\right\} \\ &+ E\left\{\mathbf{v}(n)\frac{r_a(n)\mathbf{\Phi}^T(n)}{\mathbf{\Phi}^T(n)\mathbf{\Phi}(n)}\right\} \\ &- E\left\{\frac{\mathbf{\Phi}(n)\mathbf{\Phi}^T(n)}{\mathbf{\Phi}^T(n)\mathbf{\Phi}(n)}\mathbf{v}(n)\mathbf{v}^T(n)\right\} \\ &+ E\left\{\frac{\mathbf{\Phi}(n)\mathbf{\Phi}^T(n)}{\mathbf{\Phi}^T(n)\mathbf{\Phi}(n)}\mathbf{v}(n)\mathbf{v}^T(n)\frac{\mathbf{\Phi}(n)\mathbf{\Phi}^T(n)}{\mathbf{\Phi}^T(n)\mathbf{\Phi}(n)}\right\} \\ &- E\left\{\frac{\mathbf{\Phi}(n)\mathbf{\Phi}^T(n)}{\mathbf{\Phi}^T(n)\mathbf{\Phi}(n)}\mathbf{v}(n)\frac{r_a(n)\mathbf{\Phi}^T(n)}{\mathbf{\Phi}^T(n)\mathbf{\Phi}(n)}\right\} \\ &+ E\left\{\frac{\mathbf{\Phi}(n)r_a(n)}{\mathbf{\Phi}^T(n)\mathbf{\Phi}(n)}\mathbf{v}^T(n)\right\} \\ &- E\left\{\frac{\mathbf{\Phi}(n)r_a(n)}{\mathbf{\Phi}^T(n)\mathbf{\Phi}(n)}\mathbf{v}^T(n)\frac{\mathbf{\Phi}(n)\mathbf{\Phi}^T(n)}{\mathbf{\Phi}^T(n)\mathbf{\Phi}(n)}\right\} \\ &+ E\left\{\frac{\mathbf{\Phi}(n)r_a(n)}{\mathbf{\Phi}^T(n)\mathbf{\Phi}(n)}\frac{r_a(n)\mathbf{\Phi}^T(n)}{\mathbf{\Phi}^T(n)\mathbf{\Phi}(n)}\right\} \end{aligned} \quad (19)$$

The expected values in (19) are evaluated using assumptions **A1** and **A2**, and the same considerations used to evaluate the expected values in (12). Neglecting the statistical dependence of $\mathbf{\Phi}(n)$ and $\mathbf{v}(n)$ and using the independence between $\mathbf{\Phi}(n)$ and $r_a(n)$, the third and seventh terms in the r.h.s. of (19) are equal to zero. The sixth and eighth terms lead to moments of three 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 (19) is approximated by

$$\begin{aligned} &E\left\{\frac{\mathbf{\Phi}(n)\mathbf{\Phi}^T(n)}{\mathbf{\Phi}^T(n)\mathbf{\Phi}(n)}\mathbf{v}(n)\mathbf{v}^T(n)\frac{\mathbf{\Phi}(n)\mathbf{\Phi}^T(n)}{\mathbf{\Phi}^T(n)\mathbf{\Phi}(n)}\right\} \\ &= E\left\{[\mathbf{\Phi}^T(n)\mathbf{\Phi}(n)]^{-2}\right\} \\ &\cdot E\{\mathbf{\Phi}(n)\mathbf{\Phi}^T(n)\mathbf{v}(n)\mathbf{v}^T(n)\mathbf{\Phi}(n)\mathbf{\Phi}^T(n)\} \end{aligned} \quad (20)$$

The first expectation in (20) is evaluated by integration using (14) and yields

$$E\left\{[\mathbf{\Phi}^T(n)\mathbf{\Phi}(n)]^{-2}\right\} = \frac{1}{\sigma_\phi^4(N-2)(N-4)} \quad (21)$$

The second expectation is evaluated using the Gaussian moment factoring theorem [1]:

$$\begin{aligned} &E\{\mathbf{\Phi}(n)\mathbf{\Phi}^T(n)\mathbf{v}(n)\mathbf{v}^T(n)\mathbf{\Phi}(n)\mathbf{\Phi}^T(n)\} = \\ &+2\mathbf{R}_{\phi\phi}\mathbf{K}_{vv}(n)\mathbf{R}_{\phi\phi} + tr[\mathbf{R}_{\phi\phi}\mathbf{K}_{vv}(n)]\mathbf{R}_{\phi\phi} \end{aligned} \quad (22)$$

Using the independence between $\mathbf{\Phi}(n)$ and $r_a(n)$, the last expectation in (19) yields

$$\begin{aligned} &E\left\{\frac{\mathbf{\Phi}(n)}{\mathbf{\Phi}^T(n)\mathbf{\Phi}(n)}r_a^2(n)\frac{\mathbf{\Phi}^T(n)}{\mathbf{\Phi}^T(n)\mathbf{\Phi}(n)}\right\} \\ &= E\{r_a^2(n)\}\cdot E\left\{\frac{\mathbf{\Phi}(n)\mathbf{\Phi}^T(n)}{[\mathbf{\Phi}^T(n)\mathbf{\Phi}(n)]^{-2}}\right\} \end{aligned} \quad (23)$$

From previous results,

$$E\left\{\frac{\mathbf{\Phi}(n)\mathbf{\Phi}^T(n)}{[\mathbf{\Phi}^T(n)\mathbf{\Phi}(n)]^{-2}}\right\} = \frac{1}{\sigma_{\mathbf{\Phi}}^4(N-2)(N-4)}\mathbf{R}_{\phi\phi} \quad (24)$$

Thus,

$$\begin{aligned} &E\{r_a^2(n)\}\cdot E\left\{\frac{\mathbf{\Phi}(n)\mathbf{\Phi}^T(n)}{[\mathbf{\Phi}^T(n)\mathbf{\Phi}(n)]^{-2}}\right\} = \\ &\left(1 + \mathbf{a}^T\mathbf{a} + \sigma_z^2 tr\left[E\{[\mathbf{U}^T(n)\mathbf{U}(n)]^{-1}\}\right]\right) \\ &\cdot\frac{1}{\sigma_\phi^4(N-2)(N-4)}\mathbf{R}_{\phi\phi} \end{aligned} \quad (25)$$

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

Substituting these results in (19) yields a recursion for $\mathbf{K}_{vv}(n)$:

$$\begin{aligned} &\mathbf{K}_{vv}(n) = E\{\mathbf{K}_{vv}(n)\} \\ &- \frac{1}{\sigma_\phi^2(N-2)}[\mathbf{K}_{vv}(n)\mathbf{R}_{\phi\phi} + \mathbf{R}_{\phi\phi}\mathbf{K}_{vv}(n)] \\ &+ \frac{1}{\sigma_\phi^4(N-2)(N-4)}tr[\mathbf{R}_{\phi\phi}\mathbf{K}_{vv}(n)]\mathbf{R}_{\phi\phi} \\ &+\left(1 + \mathbf{a}^T\mathbf{a} + \sigma_z^2 tr\left[E\{[\mathbf{U}^T(n)\mathbf{U}(n)]^{-1}\}\right]\right) \\ &\cdot\frac{1}{\sigma_\phi^4(N-2)(N-4)}\sigma_r^2\mathbf{R}_{\phi\phi} \end{aligned} \quad (26)$$

VI. Simulations

This section presents simulations to verify the accuracy of the analytical models given by equations (17), (18) and (26). Several simulations have been realized using the derived models. The examples presented here are representative of the results obtained. Because of space limitations the examples shown here were selected to illustrate the effect of the ratio N/P on the model's acuracy. In the examples, AR(m) means an autoregressive process of order m, and AP(k) means the AP algorithm of order k (using $k-1$ 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 eigenvalue spread of the input correlation matrix $E\{\mathbf{u}(n)\mathbf{u}^T(n)\}$ is referred to as χ_u. The system to be identified has impulse response represented by the vector $\mathbf{W}_o$ whose elements are given by $W_{o_k} = 0.5\cdot\cos(\pi(k+1)/N)+1$ and normalized so that $\mathbf{W}_o^T.\mathbf{W}_o = 1$.

Matrices $E\{\mathbf{P}_o(n)\}$ and $E\{[\mathbf{U}^T(n)\mathbf{U}(n)]^{-1}\}$ and σ_ϕ^2 were numerically estimated, given the parameters of the input AR process.

Example 1: The parameters for this example are $\sigma_u^2 = 1$; $\sigma_\phi^2 = 0.171$; $\sigma_z^2 = 0.19$, $SNR = 80dB$; $N = 50$ taps; process AR(1) with $a = -0.9$; AP(6) and $\chi_u = 302.4$.

Example 2: The parameters for this example are $\sigma_u^2 = 1$; $\sigma_\phi^2 = 0.185$; $\sigma_z^2 = 0.19$, $SNR = 80dB$; $N = 50$ taps; process AR(1) with $a = -0.9$; AP(2) and $\chi_u = 302.4$.

Example 3: The parameters for this example are $\sigma_u^2 = 1$; $\sigma_\phi^2 = 0.155$; $\sigma_z^2 = 0.19$, $SNR = 80dB$; $N = 50$ taps; process AR(1) with $a = -0.9$; AP(10) and $\chi_u = 302.4$.

Figs. 1, 2 and 3 show the mean weight behavior for some of the weights in examples 1, 2, and 3, respectively. The remaining weights have similar behavior. A very good match between simulation and theoretical predictions can be verified from these plots. Figs. 4, 5 and 6 show the MSE behavior for examples 1, 2 and 3, respectively. Notice that there is excellent match between theory and simulation in all cases for the steady-state behavior. During transient, there is a mismatch between theory and simulation that depends on the ratio N/P. Example 1 (Fig. 4) shows the results for an AP(6) algorithm with $N = 50$ ($N/P = 10$). In Example 2 (Fig. 5), P is reduced to 1 ($N/P = 50$). Notice that the theoretical curve gets closer to the simulation curve. In Example 3 (Fig. 6), the algorithm is changed to AP(10) ($N/P = 5.56$). The theoretical model separates again from the simulation during transient. This behavior has been observed in all the simulations tried. Notice that practical systems for which the AP algorithm is considered usually have large N, such as in acoustic echo cancellation. At the same time, P is not very large because computational complexity and the steady-state misadjustment increase with P [3], [6]. Also, the new model is conservative during transient and very precise in steady-state. Thus, it should be very useful for design purposes.

VII. Summary

This paper has presented a new analytical model for predicting the behavior of AP algorithm. The model was derived for AR Gaussian inputs and for unity step size (fastest convergence). 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 between theory an simulations in steady-state, and fair to good agreement during transient. The model predictions for the transient phase improve as N/P increases.

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.

[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.

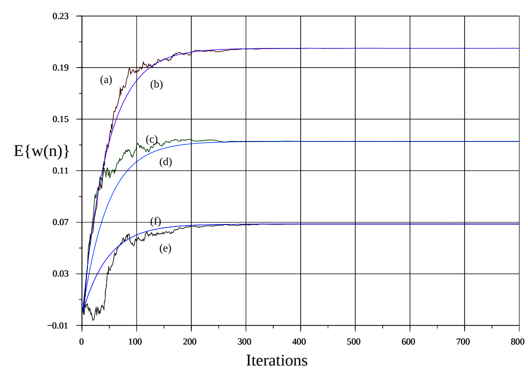

Fig. 1. Mean weight behavior for Example 1. Simulation: (a) w_{50}, (c) w_{25} and (f) w_1. Theory: (b) w_{50}, (d) w_{25} and (e) w_1.

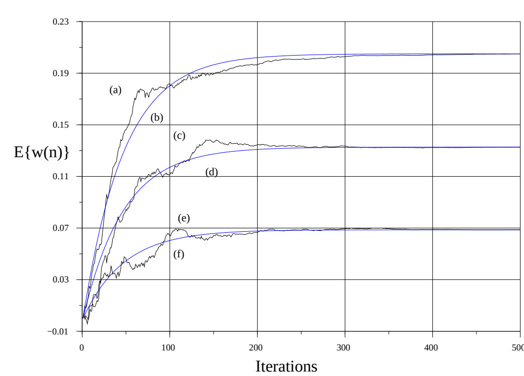

Fig. 2. Mean weight behavior for Example 2. Simulation: (a) w_{50}, (c) w_{25} and (f) w_1. Theory: (b) w_{50}, (d) w_{25} and (e) w_1.

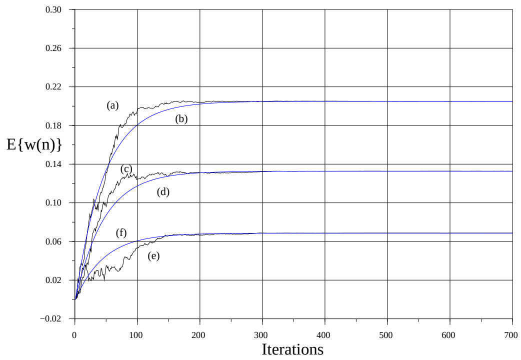

Fig. 3. Mean weight behavior for Example 3. Simulation: (a) w_{50}, (c) w_{25} and (f) w_1. Theory: (b) w_{50}, (d) w_{25} and (e) w_1.

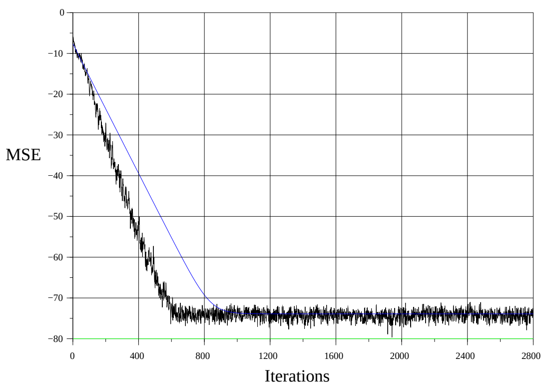

Fig. 4. MSE behavior for Example 1. Simulation: ragged curve. Theory: smooth curve. $N/P = 10$

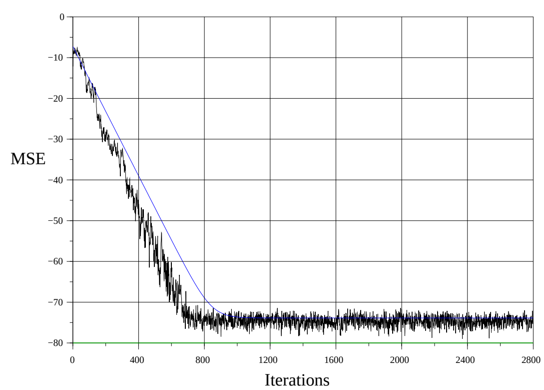

Fig. 5. MSE behavior for Example 2. Simulation: ragged curve. Theory: smooth curve. $N/P = 50$

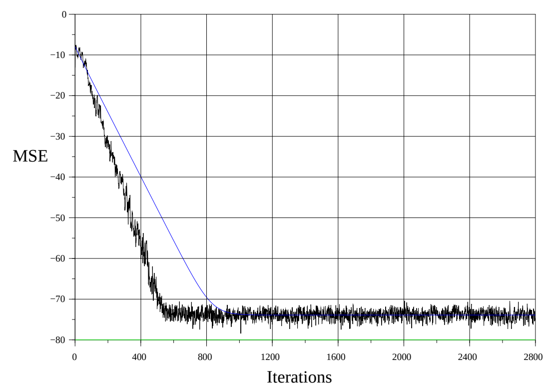

Fig. 6. MSE behavior for Example 3. Simulation: ragged curve. Theory: smooth curve. $N/P = 5.56$