

# A Stochastic Model for the Affine Projection Algorithm Operating in a Nonstationary Environment

Sérgio J. M. de Almeida  
Escola de Engenharia e Arquitetura  
Universidade Católica de Pelotas  
Pelotas-RS, Brazil  
E-mail:smelo@eel.ufsc.br

José C. M. Bermudez  
Dept. of Electrical Engineering  
Federal University of Santa Catarina  
Florianópolis-SC, Brazil  
E-mail:j.bermudez@ieee.org

Neil J. Bershad  
Dept. of Electrical Engineering and Computer Science  
University of California, Irvine, CA 92697  
E-mail:bershad@ece.uci.edu

**Abstract**—This paper presents an analytical model for predicting the stochastic behavior of the Affine Projection (AP) algorithm operating in a nonstationary environment. The model is derived for autoregressive (AR) Gaussian inputs and for unity step size (fastest convergence). Deterministic recursive equations are presented for the mean weight and mean square error for a large number of adaptive taps  $N$  as compared to the algorithm order  $P$ . The model predictions show excellent agreement with Monte Carlo simulations in transient and steady-state. The learning behavior of the AP algorithm in nonstationary environments is of great interest in applications such as acoustic echo cancellation.

## I. INTRODUCTION

Adaptive filtering is used in a large number of engineering applications. The least mean squares (LMS) adaptive algorithm and its normalized version (NLMS) are among the most often used adaptive filtering algorithms. However, their convergence rates are significantly reduced for non-white (highly correlated) inputs [1]. Acoustic echo cancellation is one important application with such input signal characteristics. The Affine Projection (AP) algorithm was proposed by Ozeki and Umeda in 1984 [2] as a solution to this problem. The AP algorithm updates the adaptive filter weights in directions that are orthogonal to the last  $P$  input vectors. It has been shown that the AP algorithm converges much faster than LMS or NLMS for correlated inputs.

Analyses of the AP algorithm in stationary environments for different input models have been presented in [3], [4] and by several other authors. However, very few results are available for the important case of nonstationary environments. In [5], tracking properties of the NLMS-OCF algorithm (a generalization of the AP algorithm) have been derived based on an independent input signal model and for a random walk non-stationarity model. The tracking model derived in [5] has been shown to agree with simulations for white inputs and for reasonably large input vector delays. Results for the AP algorithm (unit input vector delay) with highly correlated input signals were not presented.

This paper presents a new statistical analysis of the AP algorithm for nonstationary environments and autoregressive input signals with arbitrary zero-mean distribution. Analytical recursive models are derived for the mean weight and mean square error behaviors for nonstationarities modeled by a random walk model. Monte Carlo simulations show excellent agreement between theory and algorithm behavior for different degrees of nonstationarity and for highly correlated input signals.

## II. THE NONSTATIONARY DATA MODEL

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

$$d(n) = \mathbf{w}^{\circ T}(n)\mathbf{u}(n) + r(n) \quad (1)$$

where  $\mathbf{w}^{\circ}(n) = [w_0^{\circ}(n) \ w_1^{\circ}(n) \ \dots \ w_{N-1}^{\circ}(n)]^T$  denotes the time-varying optimum tap-weight vector and  $r(n)$  is a white noise with variance  $\sigma_r^2$ , which accounts for measurement noise and modeling errors.

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

$$\mathbf{u}(n) = \sum_{i=1}^P a_i \mathbf{u}(n-i) + \mathbf{z}(n) = \mathbf{U}(n)\mathbf{a} + \mathbf{z}(n) \quad (2)$$

where matrix  $\mathbf{U}(n) = [\mathbf{u}(n-1) \ \dots \ \mathbf{u}(n-P)]$  is a collection of  $P$  past input vectors  $\mathbf{u}(n-k) = [u(n-k) \ \dots \ u(n-k-N+1)]^T$  and  $\mathbf{z}(n) = [z(n) \ \dots \ z(n-N+1)]^T$  is a vector with samples from a stationary white independent Gaussian process with variance  $\sigma_z^2$ .

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

$$\hat{\mathbf{a}}(n) = [\mathbf{U}^T(n)\mathbf{U}(n)]^{-1}\mathbf{U}^T(n)\mathbf{u}(n) \quad (3)$$

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

The time variations of the optimum tap-weight vector  $\mathbf{w}^{\circ}(n)$  are assumed to follow the random walk model

$$\mathbf{w}^{\circ}(n+1) = \mathbf{w}^{\circ}(n) + \mathbf{q}(n), \quad (4)$$

where  $\mathbf{q}(n)$  is an independent zero-mean white noise vector process with variance  $\sigma_q^2$ . This simplified model is frequently used with good results in analysis of adaptive filters in nonstationary environments [5], [6].

## III. THE AFFINE PROJECTION ALGORITHM

The weight update equation of the AP algorithm with step size (maximum convergence speed) unity can be written as [2], [7]:

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \frac{\Phi(n)}{\Phi^T(n)\Phi(n)}e(n) \quad (5)$$

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

$$e(n) = \mathbf{w}^{\circ T}(n)\mathbf{u}(n) + r(n) - \mathbf{w}^T(n)\mathbf{u}(n) \quad (6)$$

where  $\mathbf{w}(n) = [w_0(n) \ w_1(n) \ \dots \ w_{N-1}(n)]^T$  is the adaptive weight vector. The vector  $\Phi(n)$  defines the direction of update, and is given by:

$$\Phi(n) = \mathbf{u}(n) - \mathbf{U}(n)\hat{\mathbf{a}}(n) \quad (7)$$

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

#### IV. VECTOR AND STATISTICAL PROPERTIES OF $\Phi$

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 follows and is more realistic for  $N \gg P$ .

Equation (2) shows an algebraic dependence between  $\mathbf{z}(n)$  and vectors  $\mathbf{u}(n-1), \dots, \mathbf{u}(n-P)$ . 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  $\mathbf{z}(n)$  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:**  $\Phi(n)$  and the weight vector  $\mathbf{w}(n)$  are statistically independent.

Assumption A2 is similar to the independence assumption applied to delay line adaptive filters with white inputs since  $\Phi(n)$  is a vector of estimates of the white noise sequence  $z(n)$  [7].

Substituting (2) in (7) yields

$$\Phi(n) = [\mathbf{I} - \mathbf{P}_u(n)]\mathbf{z}(n) = \mathbf{P}_o(n)\mathbf{z}(n) = \mathbf{z}_\perp(n) \quad (8)$$

Eq. (8) shows that  $\Phi(n)$  is orthogonal to the columns of  $\mathbf{U}(n)$ .

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

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 (8), the correlation matrix of  $\Phi(n)$  can be written as:

$$\mathbf{R}_{\phi\phi} = E\{\Phi(n)\Phi^T(n)\} = E\{\mathbf{z}_\perp(n)\mathbf{z}_\perp^T(n)\} \quad (9)$$

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

$$\mathbf{R}_{\phi\phi} = E\{\mathbf{z}(n)\mathbf{z}^T(n)\} - E\{\mathbf{z}_u(n)\mathbf{z}_u^T(n)\} \quad (10)$$

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 (10) is given by

$$\text{tr}[E\{\mathbf{z}(n)\mathbf{z}^T(n)\}] = N \cdot \sigma_z^2 \quad (11)$$

and

$$\text{tr}[E\{\mathbf{z}_u(n)\mathbf{z}_u^T(n)\}] = P \cdot \sigma_z^2 \quad (12)$$

Distributing the power equally in all dimensions results in

$$\mathbf{R}_{\phi\phi} = E\{\Phi(n)\Phi^T(n)\} = \sigma_\phi^2 \cdot \mathbf{I} = \left(\frac{N-P}{N}\right) \cdot \sigma_z^2 \cdot \mathbf{I} \quad (13)$$

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

Eq. (8) shows that each component  $\phi(n-i)$  of  $\Phi(n)$  is determined by  $\sum_{j=1}^N P_{oij} z(n-j+1)$ . From assumption A1 and  $z(n)$  white, the

random variables in this sum are independent. Thus, by the Central Limit Theorem, the distribution of  $\Phi(n)$  tends to a Gaussian for large  $N$ .

#### V. MEAN WEIGHT BEHAVIOR

Defining the weight error vector,  $\mathbf{v}(n) = \mathbf{w}(n) - \mathbf{w}^o(n)$  and using (4) and (6), (5) can be written as

$$\mathbf{v}(n+1) = \mathbf{v}(n) - \frac{\Phi(n)\mathbf{u}^T(n)}{\Phi^T(n)\Phi(n)}\mathbf{v}(n) + \frac{\Phi(n)}{\Phi^T(n)\Phi(n)}r(n) - \mathbf{q}(n) \quad (14)$$

Pre-multiplying (14) by  $\mathbf{u}^T(n)$  and  $\mathbf{U}^T(n)$ , and using the properties derived in [7] yields

$$\begin{aligned} \mathbf{v}(n+1) &= \mathbf{v}(n) - \frac{\Phi(n)\Phi^T(n)}{\Phi^T(n)\Phi(n)}\mathbf{v}(n) + \frac{\Phi(n)}{\Phi^T(n)\Phi(n)}r_a(n) \\ &\quad + \frac{\Phi(n)}{\Phi^T(n)\Phi(n)}\hat{\mathbf{a}}^T(n)\mathbf{g}(n) - \mathbf{q}(n) \end{aligned} \quad (15)$$

where  $r_a(n)$  is the filtered noise sequence [7]

$$r_a(n) = r(n) - \sum_{i=1}^P \hat{a}_i(n)r(n-i) \quad (16)$$

and  $\mathbf{g}(n)$  is the  $(P \times 1)$  vector

$$\mathbf{g}(n) = \text{diag}\{\mathbf{U}^T(n)\mathbf{Q}(n)\mathbf{L}\} \quad (17)$$

where  $\mathbf{L}$  is a  $(P \times P)$  upper triangular matrix with all nonzero elements equal to one and  $\mathbf{Q}(n)$  is the  $(N \times P)$  matrix:

$$\mathbf{Q}(n) = [\mathbf{q}(n-1) \quad \mathbf{q}(n-2) \dots \mathbf{q}(n-P)] \quad (18)$$

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

Under assumption A2 and noting that  $r(n)$  and  $q(n)$  being zero-mean and independent of any other signal makes the three last expected values in the r.h.s. of (19) equal to zero, yields

$$E\{\mathbf{v}(n+1)\} = E\{\mathbf{v}(n)\} - E\left\{\frac{\Phi(n)\Phi^T(n)}{\Phi^T(n)\Phi(n)}\right\}E\{\mathbf{v}(n)\} \quad (20)$$

Each element of the expectation in the r.h.s. of (20) has a numerator given by  $\phi(n-i)\phi(n-j)$  and a denominator equal to  $\sum_{k=0}^{N-1} \phi^2(n-k)$ . Since the components of  $\Phi(n)$  affect only two out of  $N$  terms in the denominator, numerator and denominator can be assumed weakly correlated for  $N$  large. For ergodic signals, this is equivalent to applying the averaging principle [8], as  $\phi(n-1)\phi(n-j)$  tends to be slowly varying when compared to  $\Phi^T(n)\Phi(n)$  for large  $N$ . Hence, the following approximation is used:

$$E\{[\Phi^T(n)\Phi(n)]^{-1}\Phi(n)\Phi^T(n)\} \approx E\{[\Phi^T(n)\Phi(n)]^{-1}\}\mathbf{R}_{\phi\phi} \quad (21)$$

where  $\mathbf{R}_{\phi\phi}$  is given by (13).

The expected value of  $E\{[\Phi^T(n)\Phi(n)]^{-1}\}$  is determined using the assumption that  $\Phi(n)$  is Gaussian distributed and neglecting the statistical dependence between its components (estimates of a white sequence). Thus,  $\mathbf{y} = \Phi^T(n)\Phi(n)$  has a chi-square distribution with  $G = N - P$  degrees of freedom. The value of  $G$  arises from the statistical properties of  $\Phi(n)$  determined in the previous section. Thus, [4]

$$E\{[\Phi^T(n)\Phi(n)]^{-1}\} = \frac{1}{\sigma_\phi^2(G-2)} \quad (22)$$

where  $\sigma_\phi^2 = (N-P)/N \sigma_z^2$ . Using (22) in (20) leads to:

$$E\{\mathbf{v}(n+1)\} = \left[ I - \frac{1}{\sigma_\phi^2(G-2)} \mathbf{R}_{\phi\phi} \right] E\{\mathbf{v}(n)\} \quad (23)$$

which is the recursion for the mean weight error vector.

## VI. MEAN SQUARE ERROR BEHAVIOR

Squaring (6) and taking the expected value leads, after some algebraic manipulations, to

$$E\{e^2(n)\} = \left( 1 + \mathbf{a}^T \mathbf{a} + \sigma_z^2 \text{tr}[E\{[\mathbf{U}^T(n)\mathbf{U}(n)]^{-1}\}] \right) \sigma_r^2 + \text{tr}[\mathbf{R}_{\phi\phi} \mathbf{K}(n)] \quad (24)$$

where  $\mathbf{K}(n) = E\{\mathbf{v}(n)\mathbf{v}^T(n)\}$  is the weight-error correlation matrix. In determining (24), it was assumed that the algorithm has sufficient order (greater or equal to  $P$ ). Thus,  $\hat{\mathbf{a}}(n) \approx \mathbf{a}$  was used.

The first term of (24) is a function of the input statistics. The second term needs to be determined. Postmultiplying (15) by its transpose, taking the expected value, using assumptions A1 and A2, and using the same considerations to determine (23) and (24) leads to the recursive expression:

$$\begin{aligned} \mathbf{K}(n+1) = & \mathbf{K}(n) - \mathbf{K}(n) E \left\{ \frac{\Phi(n)\Phi^T(n)}{\Phi^T(n)\Phi(n)} \right\} \\ & - E \left\{ \frac{\Phi(n)\Phi^T(n)}{\Phi^T(n)\Phi(n)} \right\} \mathbf{K}(n) \\ & + E \left\{ \frac{\Phi(n)\Phi^T(n)}{\Phi^T(n)\Phi(n)} \mathbf{v}(n)\mathbf{v}^T(n) \frac{\Phi(n)\Phi^T(n)}{\Phi^T(n)\Phi(n)} \right\} \\ & + E \left\{ \frac{\Phi(n)}{\Phi^T(n)\Phi(n)} r_a(n) r_a(n) \frac{\Phi^T(n)}{\Phi^T(n)\Phi(n)} \right\} \\ & + E \left\{ \frac{\Phi(n)}{\Phi^T(n)\Phi(n)} \mathbf{a}^T \mathbf{g}(n) \mathbf{g}^T(n) \mathbf{a} \frac{\Phi^T(n)}{\Phi^T(n)\Phi(n)} \right\} \\ & + E\{\mathbf{q}(n)\mathbf{q}^T(n)\} \end{aligned} \quad (25)$$

The first two expectations in (25) have already been determined. Since the distribution of  $\mathbf{v}(n)$  is unknown, the evaluation of the third expectation requires further approximations. Extensive simulations have shown that an adequate approximation is the ratio of the expected values. Thus, the following approximation is used

$$\begin{aligned} E \left\{ \frac{\Phi(n)\Phi^T(n)}{\Phi^T(n)\Phi(n)} \mathbf{v}(n)\mathbf{v}^T(n) \frac{\Phi(n)\Phi^T(n)}{\Phi^T(n)\Phi(n)} \right\} \approx \\ \frac{1}{E\{[\Phi^T(n)\Phi(n)]^2\}} \cdot E\{\Phi(n)\Phi^T(n)\mathbf{v}(n)\mathbf{v}^T(n)\Phi(n)\Phi^T(n)\} \end{aligned} \quad (26)$$

Eq. (26) can be evaluated using some further approximations which cannot be presented here for reasons of space. To evaluate the fifth expectations, the statistical dependence between  $\Phi(n)$  and  $\mathbf{U}(n)$  is neglected, besides the approximations already explained. Results these approximations in (25) yields a recursion for  $\mathbf{K}(n)$ .

$$\begin{aligned} \mathbf{K}(n+1) = & \mathbf{K}(n) - \frac{1}{\sigma_\phi^2(G-2)} [\mathbf{K}(n)\mathbf{R}_{\phi\phi} + \mathbf{R}_{\phi\phi}\mathbf{K}(n)] \\ & + \left[ \frac{G}{N} \cdot \text{tr}[\mathbf{K}(n)] + \left( 1 - \frac{G}{N} \right) \cdot E\{\mathbf{v}^T(n)\} \cdot E\{\mathbf{v}(n)\} \right] \end{aligned}$$

$$\begin{aligned} & \times \frac{\mathbf{R}_{\phi\phi}}{\sigma_\phi^2(G^2+2G)} + \left( 1 + \mathbf{a}^T \mathbf{a} + \sigma_z^2 \text{tr}[E\{[\mathbf{U}^T(n)\mathbf{U}(n)]^{-1}\}] \right) \\ & \times \frac{\sigma_r^2 \mathbf{R}_{\phi\phi}}{\sigma_\phi^4(G-2)(G-4)} + \sigma_q^2 \left( 1 + \frac{1}{\sigma_\phi^4} \frac{\mathbf{a}^T \Psi(n) \mathbf{a}}{(G-2)(G-4)} \right) \mathbf{R}_{\phi\phi} \end{aligned} \quad (27)$$

where  $\Psi(n)$  is the  $(P \times P)$  matrix given by:

$$\Psi(n) = E \left\{ \mathbf{U}^T(n) \mathbf{A} \mathbf{U}(n) \right\} \quad (28)$$

with

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & 2 & 2 & \dots & 2 \\ 1 & 2 & 3 & \dots & 3 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 2 & 3 & \dots & N \end{bmatrix} \quad (29)$$

## VII. STEADY-STATE BEHAVIOR – TRACKING

Using  $\mathbf{R}_{\phi\phi} = \sigma_\phi^2 \mathbf{I}$  and noting from (23) that  $\lim_{n \rightarrow \infty} E\{\mathbf{v}(n)\} = 0$ , (27) in steady-state becomes

$$\begin{aligned} \mathbf{K}(n+1) = & \left( 1 - \frac{2}{G-2} \right) \mathbf{K}(n) + \frac{1}{N(G+2)} \text{tr}[\mathbf{K}(n)] \mathbf{I} \\ & + \left( 1 + \mathbf{a}^T \mathbf{a} + \sigma_z^2 \text{tr}[E\{[\mathbf{U}^T(n)\mathbf{U}(n)]^{-1}\}] \right) \frac{\sigma_r^2}{\sigma_\phi^2(G-2)(G-4)} \mathbf{I} \\ & + \sigma_q^2 \left( \sigma_\phi^2 + \frac{1}{\sigma_\phi^2} \frac{\mathbf{a}^T \Psi(n) \mathbf{a}}{(G-2)(G-4)} \right) \mathbf{I} \end{aligned} \quad (30)$$

which shows that  $\lim_{n \rightarrow \infty} \mathbf{K}(n)$  is a diagonal matrix. Taking the trace of (30) and solving for  $\text{tr}[\mathbf{K}(\infty)] = \lim_{n \rightarrow \infty} \text{tr}[\mathbf{K}(n)]$  yields

$$\begin{aligned} \text{tr}[\mathbf{K}(\infty)] = & \frac{N(G^2-4)}{G+6} \left\{ \left( 1 + \mathbf{a}^T \mathbf{a} + \sigma_z^2 \text{tr}[E\{[\mathbf{U}^T(n)\mathbf{U}(n)]^{-1}\}] \right) \right. \\ & \times \frac{\sigma_r^2}{\sigma_\phi^2(G-2)(G-4)} + \left. \left( \sigma_\phi^2 + \frac{1}{\sigma_\phi^2} \frac{\mathbf{a}^T \Psi(n) \mathbf{a}}{(G-2)(G-4)} \right) \sigma_q^2 \right\} \end{aligned} \quad (31)$$

Using (31) in (24) leads to

$$\begin{aligned} \lim_{n \rightarrow \infty} E\{e^2(n)\} = & \left( 1 + \mathbf{a}^T \mathbf{a} + \sigma_z^2 \text{tr}[E\{[\mathbf{U}^T(n)\mathbf{U}(n)]^{-1}\}] \right) \\ & \times \left( 1 + \frac{N(G+2)}{(G-4)(G+6)} \right) \sigma_r^2 \\ & + \left( \sigma_\phi^4 + \mathbf{a}^T \Psi(n) \mathbf{a} \right) \frac{N(G+2)}{(G-4)(G+6)} \sigma_q^2 \end{aligned} \quad (32)$$

In (32) containing  $\sigma_r^2$  is the fluctuation error and the term containing  $\sigma_q^2$  is the lag error, the latter responsible for the tracking performance of the algorithm. Thus, the lag misadjustment of the AP algorithm is given by

$$\begin{aligned} \mathcal{M}_{lag-AP} = & \frac{\sigma_\phi^4 + \mathbf{a}^T \Psi(n) \mathbf{a}}{1 + \mathbf{a}^T \mathbf{a} + \sigma_z^2 \text{tr}[E\{[\mathbf{U}^T(n)\mathbf{U}(n)]^{-1}\}]} \\ & \times \frac{N(G+2)}{(G-4)(G+6)} \frac{\sigma_q^2}{\sigma_r^2} \end{aligned} \quad (33)$$

## VIII. SIMULATIONS

This section presents simulations to verify the accuracy of the analytical models given by equations (23), (24) and (27). In all cases, the matrices  $E\{[\mathbf{U}^T(n)\mathbf{U}(n)]^{-1}\}$  and  $\Psi(n)$  have been numerically estimated using the input process.

The following parameters have been used in the examples:  $\sigma_o^2 = 1$  and  $\sigma_r^2 = 10^{-6}$ , input process AR(1) with  $a_1 = -0.9$ , order of the algorithm - AP(9) ( $P = 8$ ), length of the adaptive filter  $N = 64, 128$  and  $256$ . The variance  $\sigma_q^2$  is selected to obtain a desired *degree of nonstationary* of  $S = \sqrt{N}\sigma_u\sigma_q/\sigma_r$ . Practical cases occur for  $S < 2$  [6].

Figs. 1 - 3 show the MSE behavior for the examples. In all cases, there is excellent agreement between simulations (a) (100 runs) and the analytical predictions (b) of the proposed model, both during transient and in steady-state.

Fig. 4 compares the lag misadjustments of the AP and LMS algorithms for a range of  $S$ . The LMS lag misadjustment [9], given by  $\mathcal{M}_{lag-LMS} = N\sigma_q^2/4\mu\sigma_r^2$  has been evaluated for two values of the step size  $\mu$ : (i)  $\mu_{min\_misad} = \frac{\sigma_q}{2\sigma_r} \sqrt{\frac{N}{tr[\mathbf{R}_{uu}]}}$ , which leads to the minimum overall misadjustment; (ii)  $\mu_{max\_vel} = \frac{1}{tr[\mathbf{R}_{uu}]}$ , which leads to the maximum convergence rate. Fig. 4 shows that the AP algorithm has a lag misadjustment comparable (but smaller) to LMS with  $\mu = \mu_{min\_misad}$ .

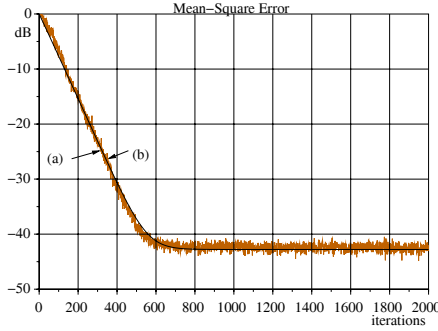


Fig. 1. MSE:  $S = 2, \sigma_u^2 = 5.433, \sigma_q^2 = 1.15 \times 10^{-8}, N = 64$  and  $P = 2$

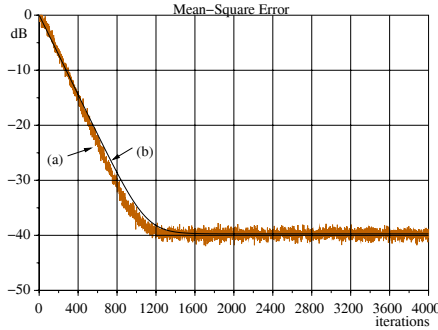


Fig. 2. MSE:  $S = 2, \sigma_u^2 = 5.35, \sigma_q^2 = 5.85 \times 10^{-9}, N = 128$  and  $P = 2$

## IX. CONCLUSIONS

This paper has presented an analytical model for predicting the stochastic behavior of the AP algorithm operating in a nonstationary environment 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 for a large number of adaptive taps  $N$  compared to the algorithm order  $P$ . The new theory yields excellent agreement with Monte Carlo simulations in both transient and steady-state phases of adaptation.

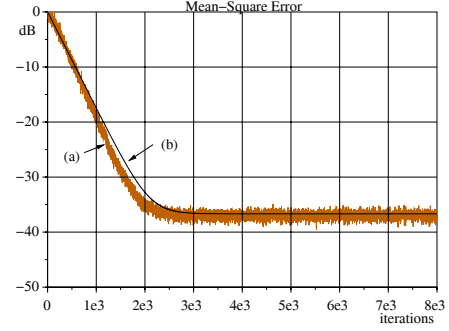


Fig. 3. MSE:  $S = 2, \sigma_u^2 = 5.30, \sigma_q^2 = 2.95 \times 10^{-9}, N = 256$  and  $P = 2$

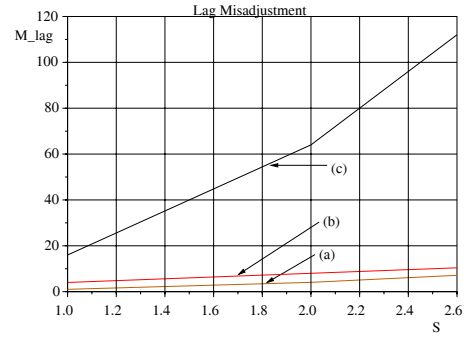


Fig. 4. Lag Misadjustments: (a) AP. (b) LMS with  $\mu = \mu_{min\_misad}$ . (c) LMS with  $\mu = \mu_{max\_vel}$

## REFERENCES

- [1] S. Haykin, *Adaptive Filter Theory* fourth edition, Prentice-Hall, 2002.
- [2] K. Ozeki, T. Umeda, "An adaptive filtering algorithm using orthogonal projection to an affine subspace and its properties," *Electronics and Communications in Japan*, vol.67-A, No.5, pp. 19-27, February 1984.
- [3] 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, 2000.
- [4] 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," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. VI, pp. 313-316, 2003.
- [5] S. G. Sankaran and A. A. Beex, "Tracking analysis results for NLMS and APA," in Proc. 2002 ICASSP, vol. 2, pp. 1105-1108, 2002.
- [6] S. Marcos and O. Macchi, "Tracking capability of the least mean square algorithm: Application to an asynchronous echo canceller," *IEEE Trans. Acoust. Speech, Signal Processing*, vol. ASSP-35, pp. 1570-1578, November 1987.
- [7] M. Rupp, "A family of filter algorithms with decorrelating properties," *IEEE Trans. on Signal Processing*, vol.46, No.3, pp.771-775, March 1998.
- [8] 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.
- [9] D. G. Manolakis, V. K. Ingle and S. M. Kogon, *Statistical and Adaptive Signal Processing: Spectral estimation, Signal Modeling, Adaptive Filtering and Array Processing*, McGraw-Hill, 2000.