A STOCHASTIC MODEL FOR THE CONVERGENCE BEHAVIOR OF THE AFFINE PROJECTION ALGORITHM FOR GAUSSIAN INPUTS

Sérgio J. M. de Almeida[1] , *José C. M. Bermudez*[2], *Neil J. Bershad*[3] *and Márcio H. Costa*[1]

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

ABSTRACT

This paper presents an analytical model for predicting the stochastic 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 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 better agreement between theory and simulations in transient and steady-state than previous models described in the literature. The learning behavior of the AP algorithm is of great interest in applications such as acoustic echo cancellation.

1. INTRODUCTION

The least mean squares (LMS) and its normalized version (NLMS) are among the most often used algorithms in adaptive signal processing applications. However, their convergence speeds are insufficient when the input signals are highly correlated and the number of adaptive taps is large [1]. Acoustic echo cancellation is one important application with such 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 weights in directions that are orthogonal to the last P input vectors. This whitens an AR(P) input and speeds convergence [3]. Thus, AP is a better choice than LMS or NLMS for applications with highly correlated input signals [4]. Higher computational complexity is the price of the faster convergence. This cost decreases as more advanced semiconductor elements are introduced. Complex algorithms have recently become feasible for applications such as echo cancellation, channel equalization and noise cancellation.

This feasability has created interest in the stochastic analysis of the AP algorithm behavior. However, quantitative statistical analysis is extremely difficult because of the underdetermined least squares solution embedded in the algorithm. Reference [4] presents a quantitative analysis of the AP algorithm which is based on an independent input signal model [5]. However, this input model limits the ability of the analysis to model the pre-whitening properties of the AP algorithm. Reference [6] presented a quantitative analysis for autoregressive Gaussian inputs. This analysis follows the work in [3] obtaining the solution of a recursion for the weight error vector variances. The solution uses previous results for the NLMS algorithm with white inputs. More recently [9] presented a new statistical analysis for the behavior of the AP algorithm for Gaussian AR inputs. Analytical difficulties are avoided for the case of a large number of adaptive taps compared to the AP algorithm order. This case allows an assumption similar to the independence assumption which has been used to successfully analyze many adaptive algorithms, including the AP algorithm [1, 6].

This paper follows the work in [9]. Results introduced in [6] are used to determine statistical properties of the decorrelated input signal. New assumptions are used to evaluate the mean weight and the mean square error behavior. Monte Carlo simulation results show excellent agreement with theoretical predictions during the adaptation phase and in steady-state.

2. THE INPUT SIGNAL MODEL

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

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

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,

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

where the matrix $U(n) = [\mathbf{u}(n-1)\ldots\mathbf{u}(n-P)]$ 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 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:

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

where $U^T(n)U(n)$ is assumed of rank P.

Sérgio J. M. de Almeida and Márcio H. Costa are 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

3. THE AFFINE PROJECTION ALGORITHM

The weight update equation of the AP algorithm with step size unity (maximum convergence speed) can be written as [3]:

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

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

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

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

4. 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 $U(n)$ 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 $P_u(n) = U(n)[U^T(n)U(n)]^{-1}U^T(n)$, the projection matrix onto the subspace spanned by the columns of $U(n)$, and $P_o(n) = I - P_u(n)$, 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 $\mathbf{z}_u(n) = P_u(n)\mathbf{z}(n)$ and $\mathbf{z}_\perp(n) = P_o(n)\mathbf{z}(n)$. Only $\mathbf{z}_u(n)$ is algebraically dependent upon $U(n)$. 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 $U(n)$. 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

$$\mathbf{\Phi}(n) = [I - P_u(n)]\mathbf{z}(n) = P_o(n)\mathbf{z}(n) = \mathbf{z}_\perp(n) \qquad (6)$$

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

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:

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

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

$$R_{\phi\phi} = E\{\mathbf{z}(n)\mathbf{z}^T(n)\} - E\{\mathbf{z}_u(n)\mathbf{z}_u^T(n)\} \qquad (8)$$

An expression for $R_{\phi\phi}$ 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

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

and

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

Distributing the power equally in all dimensions results in

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

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.

5. MEAN WEIGHT BEHAVIOR

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

$$\mathbf{v}(n+1) = \mathbf{v}(n) - \frac{\mathbf{\Phi}(n)\mathbf{u}^T(n)}{\mathbf{\Phi}^T(n)\mathbf{\Phi}(n)}\mathbf{v}(n) + \frac{\mathbf{\Phi}(n)}{\mathbf{\Phi}^T(n)\mathbf{\Phi}(n)}r(n) \qquad (12)$$

Pre-multiplying (12) by $\mathbf{\Phi}^T(n)$, $\mathbf{u}^T(n)$ and $U^T(n)$, and using the properties derived in [3] yields

$$\mathbf{v}(n+1) = \mathbf{v}(n) - \frac{\mathbf{\Phi}(n)\mathbf{\Phi}^T(n)}{\mathbf{\Phi}^T(n)\mathbf{\Phi}(n)}\mathbf{v}(n) + \frac{\mathbf{\Phi}(n)}{\mathbf{\Phi}^T(n)\mathbf{\Phi}(n)}r_a(n) \qquad (13)$$

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

$$r_a(n) = r(n) - \sum_{i=1}^{P} \hat{a}_i(n)r(n-i) \qquad (14)$$

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 (13) yields

$$E\{\mathbf{v}(n+1)\} = E\{\mathbf{v}(n)\} - E\left\{\frac{\mathbf{\Phi}(n)\mathbf{\Phi}^T(n)}{\mathbf{\Phi}^T(n)\mathbf{\Phi}(n)}\right\}E\{\mathbf{v}(n)\} \qquad (15)$$

Each element of the expectation in the r.h.s. of (15) 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 apply the averaging principle proposed in [7], as $\phi(n-1)\phi(n-j)$ tends to be slowly varying when compared to $\mathbf{\Phi}^T(n)\mathbf{\Phi}(n)$ for large values of N. Hence, the following approximation is used:

$$E\{[\mathbf{\Phi}^T(n)\mathbf{\Phi}(n)]^{-1}\mathbf{\Phi}(n)\mathbf{\Phi}^T(n)\} \approx E\{[\mathbf{\Phi}^T(n)\mathbf{\Phi}(n)]^{-1}\}R_{\phi\phi} \tag{16}$$

where $R_{\phi\phi}$ 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 (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]

$$f_y(y) = \frac{1}{2^{G/2}\sigma_\phi^G\Gamma(\frac{G}{2})}y^{(G/2)-1}e^{-y/2\sigma_\phi^2}u(y) \tag{17}$$

and direct integration leads to

$$E\{[\mathbf{\Phi}^T(n)\mathbf{\Phi}(n)]^{-1}]\} = \frac{1}{\sigma_\phi^2(G-2)} \tag{18}$$

where $\sigma_\phi^2 = (N-P)/N\ \sigma_z^2$. Using (18) in (15) leads to:

$$E\{\mathbf{v}(n+1)\} = \left[I - \frac{1}{\sigma_\phi^2(G-2)}R_{\phi\phi}\right]E\{\mathbf{v}(n)\} \tag{19}$$

which is the recursion for the mean weight error vector.

6. MEAN SQUARE ERROR BEHAVIOR

Squaring (4) and taking the expected value leads, after some algebraic manipulations, to

$$\begin{aligned} E\{e^2(n)\} = &\left(1 + \mathbf{a}^T\mathbf{a} + \sigma_z^2 tr\left[E\{[U^T(n)U(n)]^{-1}\}\right]\right)\sigma_r^2 \\ &+ tr[R_{\phi\phi}K(n)] \end{aligned} \tag{20}$$

where $K(n) = E\{\mathbf{v}(n)\mathbf{v}^T(n)\}$ is the weight-error correlation matrix. The first term of (20) is a function of the input statistics. The second term needs to be determined.

Postmultiplying (13) by its transpose, taking the expected value, using assumptions A1 and A2, and using the same considerations to determine (19) leads to the recursive expression:

$$\begin{aligned} K(n+1) = &K(n) - K(n)E\left\{\frac{\mathbf{\Phi}(n)\mathbf{\Phi}^T(n)}{\mathbf{\Phi}^T(n)\mathbf{\Phi}(n)}\right\} \\ &- E\left\{\frac{\mathbf{\Phi}(n)\mathbf{\Phi}^T(n)}{\mathbf{\Phi}^T(n)\mathbf{\Phi}(n)}\right\}K(n) \\ &+ E\left\{\frac{\mathbf{\Phi}(n)\mathbf{\Phi}^T(n)}{\mathbf{\Phi}^T(n)\mathbf{\Phi}(n)}\mathbf{v}(n)\mathbf{v}^T(n)\frac{\mathbf{\Phi}(n)\mathbf{\Phi}^T(n)}{\mathbf{\Phi}^T(n)\mathbf{\Phi}(n)}\right\} \\ &+ E\left\{\frac{\mathbf{\Phi}(n)r_a(n)}{\mathbf{\Phi}^T(n)\mathbf{\Phi}(n)}\frac{r_a(n)\mathbf{\Phi}^T(n)}{\mathbf{\Phi}^T(n)\mathbf{\Phi}(n)}\right\} \end{aligned} \tag{21}$$

The first two expectations in (21) 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

$$\begin{aligned} &E\left\{\frac{\mathbf{\Phi}(n)\mathbf{\Phi}^T(n)}{\mathbf{\Phi}^T(n)\mathbf{\Phi}(n)}\mathbf{v}(n)\mathbf{v}^T(n)\frac{\mathbf{\Phi}(n)\mathbf{\Phi}^T(n)}{\mathbf{\Phi}^T(n)\mathbf{\Phi}(n)}\right\} = \\ &\frac{1}{E\{[\mathbf{\Phi}^T(n)\mathbf{\Phi}(n)]^2\}}\cdot E\{\mathbf{\Phi}(n)\mathbf{\Phi}^T(n)\mathbf{v}(n)\mathbf{v}^T(n)\mathbf{\Phi}(n)\mathbf{\Phi}^T(n)\} \end{aligned} \tag{22}$$

The equation (22) can be evaluated using some further approximations which cannot be presented here for reasons of space. Using these approximations in (21) yields a recursion for $K(n)$.

$$\begin{aligned} &K(n+1) = K(n) - \frac{1}{\sigma_\phi^2(G-2)}[K(n)R_{\phi\phi} + R_{\phi\phi}K(n)] \\ &+\left[\frac{G}{N}\cdot tr[K(n)] + \left(1 - \frac{G}{N}\right)\cdot E\{\mathbf{v}^T(n)\}\cdot E\{\mathbf{v}(n)\}\right] \\ &\cdot\frac{R_{\phi\phi}}{\sigma_\phi^2(G^2+2G)} + \left(1 + \mathbf{a}^T\mathbf{a} + \sigma_z^2 tr\left[E\{[U^T(n)U(n)]^{-1}\}\right]\right) \\ &\cdot\frac{\sigma_r^2 R_{\phi\phi}}{\sigma_\phi^4(G-2)(G-4)} \end{aligned} \tag{23}$$

7. SIMULATIONS

This section presents simulations to verify the accuracy of the analytical models given by equations (19), (20) and (23). The results are compared to results for previous models described in the literature [6], [9]. In all cases, the matrix $E\{[U^T(n)U(n)]^{-1}\}$ has been numerically estimated using the input process.

The following parameters have been used in the examples: $\sigma_u^2 = 1$ and $\sigma_r^2 = 10^{-6}$; input process $AR(1)$ with $a = -0.9$; $\alpha = 1$. Orders of the algorithm: $AP(9)$ and $AP(17)$. Orders of the adaptive filter: $N = 64$ and $N = 128$.

Figs. 1 – 4 show the MSE behavior for the examples. Note that there is excellent agreement between simulations (150 runs) and the analytical predictions of the new model, both during transient and in steady-state. Also shown are the results predicted by the models derived in [6] and [9]. The new model yields better predictions of algorithm behavior in all cases.

8. CONCLUSIONS

This paper has presented an analytical model for predicting the stochastic behavior of the AP algorithm 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.

9. REFERENCES

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

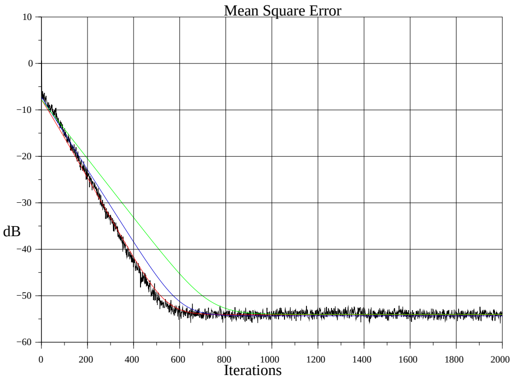

Fig. 1. MSE:Comparisons between (black) Monte Carlo simulations; (red) the new analytical model; (blue) model described in [6]; (green) model described in [9]. $AP(9)$; $N = 64$.

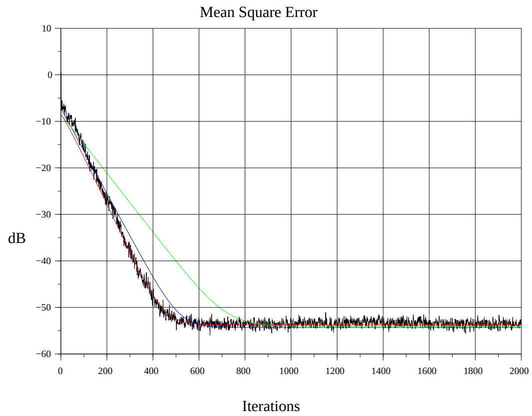

Fig. 2. MSE:Comparisons between (black) Monte Carlo simulations; (red) the new analytical model; (blue) model described in [6]; (green) model described in [9]. $AP(17)$; $N = 64$.

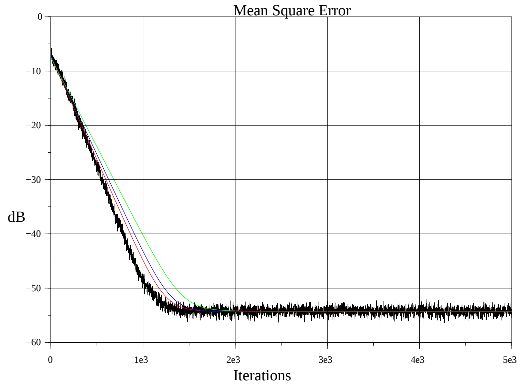

Fig. 3. MSE:Comparisons between (black) Monte Carlo simulations; (red) the new analytical model; (blue) model described in [6]; (green) model described in [9]. $AP(9)$; $N = 128$.

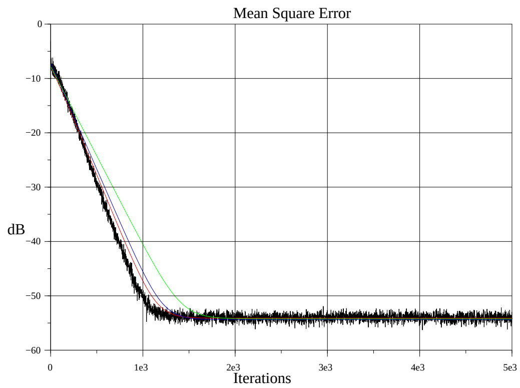

Fig. 4. MSE:Comparisons between (black) Monte Carlo simulations; (red) the new analytical model; (blue) model described in [6]; (green) model described in [9]. $AP(17)$; $N = 128$.

[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] 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.