OPTIMUM VARIABLE EXPLICIT REGULARIZED AFFINE PROJECTION ALGORITHM

Hernán Rey[1], *Leonardo Rey Vega*[1], *Sara Tressens*[1], *Jacob Benesty*[2]

[1]Facultad de Ingeniería, Universidad de Buenos Aires,
Paseo Colón 850, Buenos Aires, Argentina (e-mail: hrey,lrey,stres@fi.uba.ar).
[2]INRS-EMT, Université du Québec,
800 de la Gauchetiere Ouest, Montréal, Québec, Canada (e-mail: benesty@emt.inrs.ca).

ABSTRACT

A variable regularized Affine Projection Algorithm (VR-APA) is introduced, which does not require the classical step size. Its use is supported from different points of view. First, it has the property of being H$^\infty$ optimal, providing robust behavior against perturbations and model uncertainties. Second, the time varying regularization parameter is obtained by maximizing the speed of convergence of the algorithm. At each time step, it needs knowledge of the power of the estimation error vector, which can be estimated by averaging *observable* quantities. Although we first derive it for a linear time invariant (LTI) system, we show that the same expression holds if we consider a time varying system following a first order Markov model. Simulation results are presented to test the performance of the proposed algorithm and to compare it with other schemes under different situations.

1. INTRODUCTION

Adaptive filtering appear very frequently as a solution in engineering problems [1]. Adaptive filtering schemes have not only the ability of solving problems with less computational cost but can also deal with time variations of the system (nonstationary environments).

In this work, we focus on the *Affine Projection Algorithm* (APA) [2]. We propose an APA with a time varying regularization, which is supported from different points of view. First, the regularization parameter can control the changes along the direction of update without an upper stability bound (for any positive value), so the classic step size μ is no longer needed [3]. Second, we recently showed in [3] that a regularized APA is H$^\infty$ optimal providing robust behavior against perturbations and model uncertainties. This approach includes the traditional use of the explicit regularization to provide numerical stability to those algorithms that have to deal with ill conditioned matrix inversions. However, the regularization can also help to reduce other instability sources, e.g. measurement noise.

In APA, some of the proposed methods for step-size control require extra processing for their implementation (like pre-whitening or delay coefficients estimation) [4], while others have many parameters [5], which have not been linked to any expression that could turn them into parameters of design (e.g., showing the steady state estimation error as a function of these parameters). Although in [3], we proposed an optimal choice for the regularization factor to achieve maximum speed of convergence, it requires the knowledge or computation of the system error power, $E\left[||\tilde{\mathbf{w}}_i||^2\right]$. To skip this issue, we introduce here a variant that depends only on the power of the estimation error vector, i.e. $E\left[||\mathbf{e}_i||^2\right]$. We estimate it by time averaging the observable quantity $||\mathbf{e}_i||^2$, allowing us to derive a new *Variable Regularized APA* (VR-APA). The simple expression derived here gives information about the relationship between the regularization factor and the convergence behavior of the algorithm. In addition, we prove that the same optimal regularization choice holds if we consider a time varying system using a first order Markov model.

In section 2 we introduce the APA recursion and propose to use a variable regularized version of it. The optimal choice for the regularization sequence is analyzed in section 3. Finally, simulation results are presented in section 4. Boldface symbols are used for vectors (in lower case letters) and matrices (in upper case letters).

2. THE APA FAMILY

Let $\mathbf{w}_i = \left(w_i^0, w_i^1, \ldots, w_i^{M-1}\right)^T$ be an unknown $M \times 1$ linear finite-impulse response system. The $M \times 1$ input vector at time i, $\mathbf{x}_i = (x_i, x_{i-1}, \ldots, x_{i-M+1})^H$, passes through the system giving an output $y_i = \mathbf{x}_i^H \mathbf{w}_i$. This output is observed, but in this process it usually appears a measurement noise, v_i, which will be considered as additive. Thus, each input $\mathbf{x}_i$ gives an output $d_i = \mathbf{x}_i^H \mathbf{w}_i + v_i$. We want to find $\hat{\mathbf{w}}_i$ to estimate $\mathbf{w}_i$. This adaptive filter receives the same input, leading to an output estimation error $e_i = d_i - \mathbf{x}_i^H \hat{\mathbf{w}}_i$.

When data blocks are used, we can define the $M \times K$ data matrix $\mathbf{X}_i = [\mathbf{x}_i\, \mathbf{x}_{i-1} \cdots \mathbf{x}_{i-K+1}]$, the $K \times 1$ desired vector $\mathbf{d}_i = [d_i, d_{i-1}, \cdots, d_{i-K+1}]^T$, the $K \times 1$ noise vector $\mathbf{v}_i = [v_i\, v_{i-1} \cdots \cdots v_{i-K+1}]^T$, the $K \times 1$ estimation error vector $\mathbf{e}_i = \mathbf{d}_i - \mathbf{X}_i^H \hat{\mathbf{w}}_i$, the system error vector $\tilde{\mathbf{w}}_i = \mathbf{w}_i - \hat{\mathbf{w}}_i$ and the *a priori* error vector $\mathbf{e}_{a,i} = \mathbf{X}_i^H \tilde{\mathbf{w}}_i$.
The APA was first introduced in [2], and follows the recursion:

$$\hat{\mathbf{w}}_{i+1} = \hat{\mathbf{w}}_i + \mu \mathbf{X}_i \left(\mathbf{X}_i^H \mathbf{X}_i\right)^{-1} \mathbf{e}_i \qquad (1)$$

where μ is a scalar known as step size, included to control the changes along the selected direction. Moreover, setting $K = 1$ in (1), leads to the popular NLMS algorithm.

The first motivation for using APA is to make an improvement on the convergence speed with an acceptable increase in the computational cost. Sankaran and Beex have shown in [6] that $0 < \mu < 1$ and $1 < \mu < 2$ are both stable, but the first choice has less steady state error with the same convergence speed. They also showed in [7] that the tracking ability of APA is maximized when μ is close to 1. On the other hand, when highly colored input data are presented, the matrix inversion in (1) becomes very difficult as its condition number grows critically. Using this numerical stability argument, a positive *regularization* term is usually added.

This work was supported in part by Universidad de Buenos Aires, project UBACYT I005.

These are some of the reasons why we propose to set $\mu = 1$ and use a time varying regularization parameter to control the changes along the selected direction, so that the APA update becomes:

$$\hat{\mathbf{w}}_{i+1} = \hat{\mathbf{w}}_i + \mathbf{X}_i \left(\beta_i \mathbf{I}_K + \mathbf{X}_i^H \mathbf{X}_i\right)^{-1} \mathbf{e}_i. \quad (2)$$

This rule gives an "effective step size" in $(0, 1)$ for any positive β_i, so there is no upper bound that could make the algorithm unstable. In the following, we define:

$$\mathbf{S}_i = \left(\beta_i \mathbf{I}_K + \mathbf{X}_i^H \mathbf{X}_i\right)^{-1}. \quad (3)$$

Despite the fact of assuming that the system is LTI in the next section, we will consider the nonstationary environment when we perform the optimization of the regularization parameter.

3. OPTIMAL REGULARIZATION CHOICE

As follows from the analysis performed in [3], any positive sequence $\{\beta_i\}$ would guarantee the robust behavior of the APA. From the many options that can be chosen, we will look for the one that maximize the speed of convergence. To do so, we choose for each i the β_i that minimizes $E\left[||\tilde{\mathbf{w}}_{i+1}||^2\right] - E\left[||\tilde{\mathbf{w}}_i||^2\right]$. We assume that the measurement noise is a zero mean white noise independent of the input data. We also use the usual assumption [4][6][8][9] that $\mathbf{e}_{a,i}$ and $\mathbf{v}_i$ are independent. As a consequence, the sequence β_i is independent, i.e. independence between β_i at different time steps. At first, we assume a stationary environment and then analyze the nonstationary one.
From the APA recursion (2):

$$E\left[||\tilde{\mathbf{w}}_{i+1}||^2\right] - E\left[||\tilde{\mathbf{w}}_i||^2\right] = E\left[\mathbf{e}_i^H \mathbf{S}_i \mathbf{X}_i^H \mathbf{X}_i \mathbf{S}_i \mathbf{e}_i\right] - E\left[\mathbf{e}_{a,i}^H \mathbf{S}_i \mathbf{e}_i\right] - E\left[\mathbf{e}_i^H \mathbf{S}_i \mathbf{e}_{a,i}\right]. \quad (4)$$

As $\mathbf{e}_i = \mathbf{e}_{a,i} + \mathbf{v}_i$, this leads to:

$$E\left[||\tilde{\mathbf{w}}_{i+1}||^2\right] - E\left[||\tilde{\mathbf{w}}_i||^2\right] = E\left[\mathbf{e}_i^H \mathbf{S}_i \mathbf{X}_i^H \mathbf{X}_i \mathbf{S}_i \mathbf{e}_i\right] - 2E\left[\mathbf{e}_i^H \mathbf{S}_i \mathbf{e}_i\right] + 2E\left[\mathbf{v}_i^H \mathbf{S}_i \mathbf{v}_i\right]. \quad (5)$$

Now, we perform a singular value decomposition (SVD) of the input matrix, i.e. $\mathbf{X}_i = \mathbf{V}_i \mathbf{\Sigma}_i \mathbf{U}_i^H$. By differentiating (5) partially towards β_i, its optimum value is the one that solve:

$$E\left(\mathbf{e}_i^H \mathbf{U}_i \mathbf{M}_i \mathbf{U}_i^H \mathbf{e}_i\right) = E\left(\mathbf{v}_i^H \mathbf{U}_i \mathbf{O}_i \mathbf{U}_i^H \mathbf{v}_i\right) \quad (6)$$

where $\mathbf{M}_i$ and $\mathbf{O}_i$ are diagonal matrices:

$$(\mathbf{M}_i)_{kk} = \frac{\beta_i}{\left(\left(\rho_i^k\right)^2 + \beta_i\right)^3} \qquad (\mathbf{O}_i)_{kk} = \frac{1}{\left(\left(\rho_i^k\right)^2 + \beta_i\right)^2}$$

with $\left(\rho_i^k\right)^2$ and $\mathbf{U}_i \in \mathbb{C}^{K \times K}$ being the eigenvalues and eigenvector matrix of $\mathbf{X}_i^H \mathbf{X}_i$.

Suppose now that the system is nonstationary and its dynamic is given by:

$$\mathbf{w}_{i+1} = \mathbf{w}_i + \mathbf{n}_i \quad (7)$$

where $\mathbf{n}_i$ is a zero mean white noise vector independent of the system. Following the same procedure as before, we can obtain (6) as in the stationary case. This fact does not depend on the input statistics nor on the system noise power.

3.1. Choice of β_i under simplifying assumptions

Nevertheless we have just made the usual assumptions in APA analysis, solving (6) depends on the SVD of $\mathbf{X}_i^H \mathbf{X}_i$. As we do not have this information, we perform an *heuristic* approximation by replacing for $1 \leq k \leq K$:

$$\left(\rho_i^k\right)^2 \approx \sigma_x^2 M \quad (8)$$

which is the average of the eigenvalues. Under this condition, the matrices $\mathbf{M}_i$ and $\mathbf{O}_i$ can be expressed as a constant times the identity matrix for each i. Hence, solving (6) leads to:

$$\beta_i = \frac{K\,\sigma_v^2\,\sigma_x^2\,M}{E\left[||\mathbf{e}_i||^2\right] - K\,\sigma_v^2}. \quad (9)$$

The denominator is large at the beginning of the adaptation, leading to a small β_i. As the error decreases, the denominator grows, slowing the adaptation and allowing the APA to have a small misadjustment. Intuitively, β_i is also expected to be proportional to the variables that appear on the numerator.

Despite the approximation (8) could seem inaccurate, we can arrive to equation (9) with different assumptions. As $E\left[\mathbf{X}_i^H \mathbf{X}_i\right] = M\,\mathbf{R}_x^K$, where $\mathbf{R}_x^K$ is the $K - th$ order autocorrelation matrix of the (stationary) input signal, if M is large and $K << M$ (usual condition in applications), then it is reasonable to assume:

$$\mathbf{X}_i^H \mathbf{X}_i \approx M\,\mathbf{R}_x^K. \quad (10)$$

Thus, the eigenvalues and eigenvectors of $\mathbf{R}_x^K$ are not stochastic anymore. In general, $\mathbf{R}_x^K$ will be required and the expression for β_i could be quite complicated. However, we can analyze two special cases.

For white input ($\mathbf{R}_x^K = \sigma_x^2\,\mathbf{I}_K$), replacing (10) in (5) gives the result:

$$E\left[||\tilde{\mathbf{w}}_{i+1}||^2\right] - E\left[||\tilde{\mathbf{w}}_i||^2\right] = \frac{2}{M\sigma_x^2 + \beta_i} K\,\sigma_v^2 - \frac{(M\sigma_x^2 + 2\,\beta_i)}{(M\sigma_x^2 + \beta_i)^2} E\left[||\mathbf{e}_i||^2\right]. \quad (11)$$

Minimizing this expression with respect to β_i brings the optimal regularization choice (9).

If the input is a highly correlated AR1 (first order autoregressive process with pole close to 1), it can be proved that (9) is the solution to the β_i optimization [10].

Although the β_i choice (9) was first derived heuristically, the two extreme cases analyzed here encourage us to implement it. For these reasons we propose a new *Variable Regularized APA* (VR-APA) by using the update equation (2) with β_i computed from (9). The quantity $E\left[||\mathbf{e}_i||^2\right]$ is estimated by time averaging $||\mathbf{e}_i||^2$.

4. SIMULATION RESULTS

For implementing the proposed algorithm, some issues should be taken into account. The quantity $E\left[||\mathbf{e}_i||^2\right]$ is estimated by averaging $||\mathbf{e}_i||^2$ over a sliding window. This estimation could result in a lower magnitude than $K\sigma_v^2$, specially when $K > 1$, which would make $\beta_i < 0$. When the denominator of (9) becomes negative, we set $\beta_i = \beta_{MAX}$, which is the maximum value allowed for beta and is defined by:

$$\beta_{MAX} = \frac{M\sigma_x^2}{\delta} \quad (12)$$

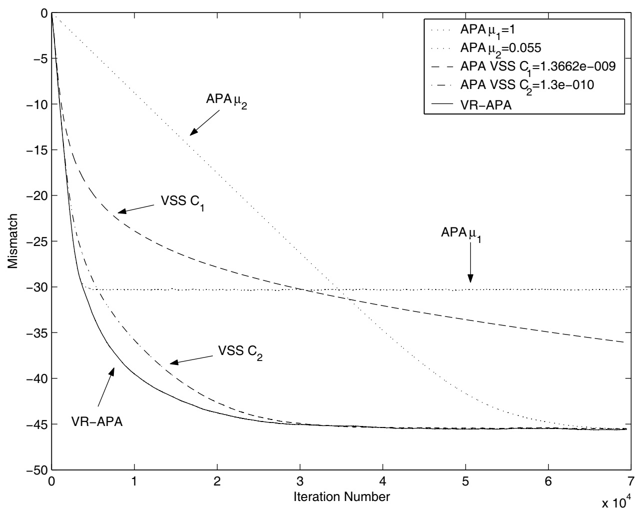

Fig. 1. Mismatch (in dB) for white input signal. $M = 512$, $K = 1$, $\delta = 0.05$, $SNR = 30$ dB.

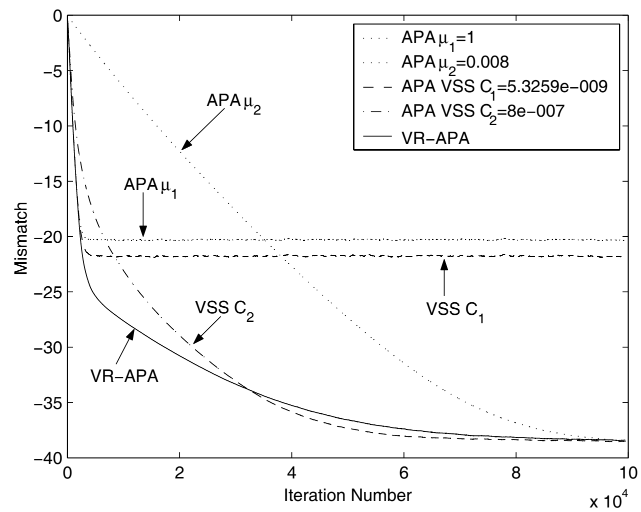

Fig. 2. Mismatch (in dB) for AR1(0.95). $M = 512$, $K = 8$, $\delta = 0.08$, $SNR = 30$ dB.

where δ is a parameter of design that could be set for having a certain level of steady state error (see [10] for details).

The system is taken from a measured impulse response, truncated to $M = 256$ or $M = 512$. The adaptive filter length is set equal to M in each case.

We use the *mismatch*, i.e. $\|\tilde{\mathbf{w}}_i\|^2/\|\mathbf{w}_i\|^2$, as a measure of performance. The plots are the result of ensemble averaging over 100 independent trials. A zero mean Gaussian white noise is added to the system output, so that the signal to noise ratio is 30 dB, except for the nonstationary environments where 60 dB is used. The output variance is set to $\sigma_y^2 = 0.1$ and $\sigma_y^2 = 1$ respectively.

The performance of the proposed algorithm is compared with other strategies. We simulate the standard APA with $\mu_1 = 1$ and other with the μ_2 that gives the same steady state mismatch as the one of the VR-APA. In both cases, a fixed regularization factor is set to $\beta = 20\,\sigma_x^2$. We also show the performance of the variable step size APA (VSS-APA) introduced by Shin *et al.* [9]. Its update is calculated as in (1), where:

$$\mu_i = \mu_{max} \frac{\|\hat{\mathbf{p}}_i\|^2}{\|\hat{\mathbf{p}}_i\|^2 + C}$$

$$\hat{\mathbf{p}}_i = \alpha\, \hat{\mathbf{p}}_{i-1} + (1-\alpha)\, \mathbf{X}_i \left(\mathbf{X}_i^H \mathbf{X}_i\right)^{-1} \mathbf{e}_i$$

and $C = \sigma_v^2 \mathrm{Tr}\left\{E\left[\left(\mathbf{X}_i^H \mathbf{X}_i\right)^{-1}\right]\right\}$. Although the authors proposed that C could be approximated by K/SNR, we believe that a better approximation is:

$$C \approx \frac{K\,\sigma_v^2}{M\,\sigma_x^2}. \qquad (13)$$

We set $\mu_{max} = 1$ and $\alpha = 0.99$. One simulation uses C_1 given by (13) while another employs the C_2 that gives the same steady state mismatch as the one of the VR-APA.

Stationary systems

Fig. 1 shows the simulated results for white input excitation. The proposed VR-APA outperforms the standard APA. With respect to the VSS-APA, the authors in [9] claim that the performance is not highly sensitive to the parameter C. Here, it can be seen that even small changes lead to a 3.5 dB difference in the steady state mismatch. Moreover, C_2 could be bigger or smaller than C_1 and it can differ in more than one order of magnitude.

We also explore the performance for an AR1 input with pole in 0.95. In Fig. 2, $C_2 \approx 150\, C_1$ gives more than 16 dB improvement. The VR-APA shows a good performance, with fast initial convergence and low steady state.

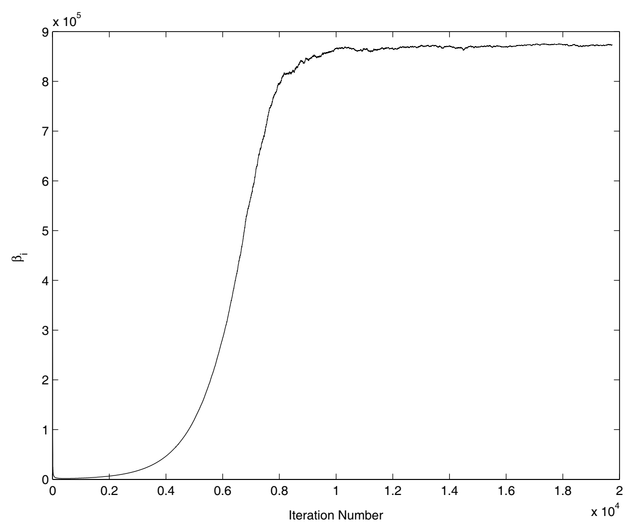

Fig. 3. Variation of β_i across time for AR1(0.9). $M = 256$, $K = 2$, $\delta = 0.08$, $SNR = 30$ dB.

The evolution of β_i across time is presented in Fig. 3 for an AR1(0.9) with $K = 2$. Initially, it is small so fast convergence is obtained. As the estimation of $E\left[\|\mathbf{e}_i\|^2\right]$ becomes smaller, β_i increases until it reaches β_{MAX}.

Although we have analyzed the cases of white and AR1 inputs with long system responses, we also observed that the VR-APA outperforms the other schemes even when these hypothesis are not satisfied (not shown) [10].

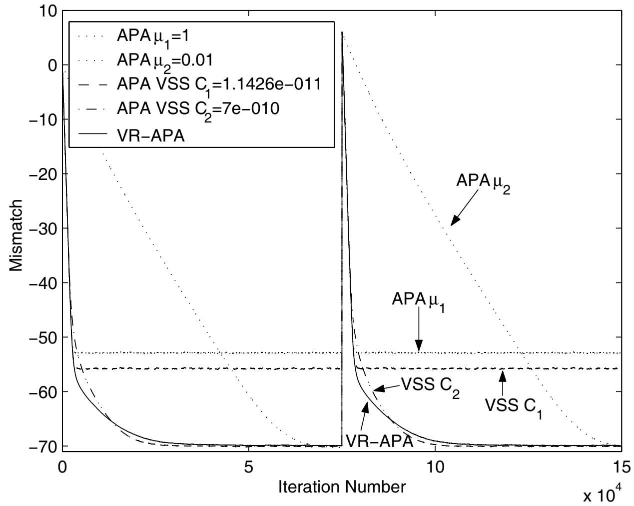

Fig. 4. Mismatch (in dB) for AR1(0.9). Sudden change from $\mathbf{w}$ to $-\mathbf{w}$. $M = 256$, $K = 8$, $\delta = 0.05$, $SNR = 60$ dB.

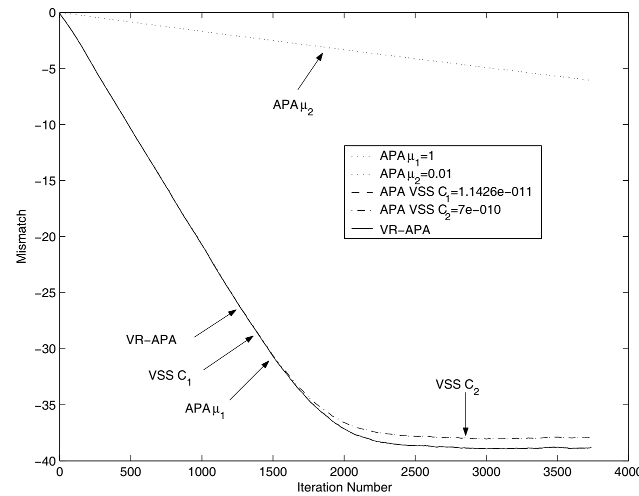

Fig. 5. Mismatch (in dB) for AR1(0.9). First order Markov system. $M = 256$, $K = 8$, $\delta = 0.05$, $\sigma_n^2 = 8.96 \times 10^{-13}$, $SNR = 60$ dB.

Nonstationary systems

First we test the recovery to a sudden change in the system. In Fig. 4 we show the result of suddenly multiplying by -1 the system response. The algorithm can learn the new system without loosing performance with respect to the one with the initial system.

Now we study the performance under a first order Markov system. To do so, we start with the parameters of the stationary systems simulated in Fig. 4. Then, we generate a Gaussian white noise vector $\mathbf{n}_i$ that is added to the system according to (7). Each component of the noise vector has a power σ_n^2.

The VR-APA can adapt well to this situation without increasing so much β_i, so that good tracking performance is accomplished. Despite the fact that the VR-APA has the same performance as the VSS-APA with C_1 and the standard one with $\mu_1 = 1$, the proposed algorithm have the same initial convergence speed and about 15 dB improvement in steady state for the stationary case, as can be seen in the first half of Fig. 4. The standard APA with small μ_2 has a poor performance as previously noted in [7].

5. CONCLUSIONS

In this work we proposed a modified update for the APA family, which includes the explicit regularization factor. We performed an analysis for optimizing β_i to have maximum speed of convergence. The general expression depends on the input statistics. Nevertheless, we proved that the same expression holds for stationary and nonstationary (random walk) systems. A closed formula for β_i was derived and shows to be optimal for white and highly colored AR1 inputs.

The proposed VR-APA showed great performance under different scenarios even when compared with standard and VSS APAs. The variable regularization factor can control well the system update and at the same time allow a robust performance against perturbations (not only generated by numerical instabilities).

6. REFERENCES

[1] S. Haykin, *Adaptive Filter Theory*, 3rd ed., Upper Saddle River, NJ: Prentice Hall, 2002.

[2] K. Ozeki and T. Umeda, "An adaptive filtering algorithm using an orthogonal projection to an affine subspace and its properties," *Electron. Commun. Jpn.*, vol. 67-A, no. 5, pp. 19–27, May 1984.

[3] H. G. Rey, L. Rey Vega, S. Tressens and B. Cernuschi Frias, "Analysis of explicit regularization in affine projection algorithms: robustness and optimal choice," in *Proc. EUSIPCO-04*, Vienna, Austria, Sept. 2004, pp. 1089–1092.

[4] V. Myllylä and G. Schmidt, "Pseudo-optimal regularization for affine projection algorithms," in *Proc. ICASSP-02*, Orlando, USA, May 2002, pp. 1917–1920.

[5] K. Mayyas and T. Aboulnasr, "A fast weighted subband adaptive algorithm," in *Proc. ICASSP-99*, Phoenix, USA, Mar. 1999, pp. 1249–1252.

[6] S. G. Sankaran and A. A. L. Beex, "Convergence behavior of affine projection algorithms," *IEEE Trans. Signal Processing*, vol. 48, no. 4, pp. 1086–1096, Apr. 2000.

[7] S. G. Sankaran and A. A. L. Beex, "Tracking analysis results for NLMS and APA," in *Proc. ICASSP-02*, Orlando, USA, May 2002, pp. 1105–1108.

[8] H. C. Shin and A. H. Sayed , "Mean-square performance of a family of affine projection algorithms," *IEEE Trans. Signal Processing*, vol. 52, no. 1, pp. 90-102, Jan. 2004.

[9] H. C. Shin and A. H. Sayed , "Variable step-size NLMS and affine projection algorithms," *IEEE Signal Processing Lett.*, vol. 11, no. 2, pp. 132–135, Feb. 2004.

[10] H. G. Rey, L. Rey Vega, S. Tressens and J. Benesty , "Variable explicit regularization in Affine Projection Algorithm: robustness issues and optimal choice," *Submitted to IEEE Trans. Signal Processing*, 2005.