ANALYSIS OF EXPLICIT REGULARIZATION IN AFFINE PROJECTION ALGORITHMS: ROBUSTNESS AND OPTIMAL CHOICE

Hernán G. Rey†, *Leonardo Rey Vega*†, *Sara Tressens*† *and Bruno Cernuschi Frías*‡

‡ Facultad de Ingeniería, Universidad de Buenos Aires and CONICET
† Facultad de Ingeniería, Univ. de Buenos Aires
Paseo Colón 850 (1063), Buenos Aires, Argentina
{hrey@fi.uba.ar}{stres@fi.uba.ar}{bcf@ieee.org}

ABSTRACT

The affine projection algorithms (APA) are a family of adaptive filters with decorrelating properties. In this paper we propose the use of a time-varying explicit regularization factor instead of the classic step size control. We show that APA results stable and has robust performance against disturbances and model uncertainties. This fact is shown by the H$^\infty$ optimality and error energy bounds we provide. Finally, we find the optimal regularization choice for maximum speed of convergence. We consider first the case under independence assumptions and then with gaussian hypothesis.

1. INTRODUCTION

An adaptive filtering problem could be understood as one of identifying an unknown system using input-output data pairs. These situations appears very frequently in engineering applications. 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) [5]. It performs the actualization of the system estimation based on multiple input vectors. Although it has more computational cost than the widely used *Normalized Least Mean Squares* (NLMS) algorithm [2], it also presents better decorrelation properties. This allows us to improve the speed of convergence of the NLMS algorithm as the correlation of the input data increases [8].

Particularly, we study the role of the regularization factor in this algorithm. Its use has been justified in the bibliography – specially if the input data is highly correlated – by invoking noise power reduction and numerical stability arguments [6].

We propose a modified step size with a time varying regularization, β_i, which is sustained from new 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. Second, β_i is related with robustness issues. By invoking the theory of H$^\infty$ estimation in Krein spaces [3], we can prove that the APA with this modified update is H$^\infty$ optimal. As a consequence, it presents a robust behavior against perturbations and model uncertainties, in the sense that small perturbations lead to small estimation errors. This is an important characteristic in real implementations.

This work has been supported in part by *Universidad de Buenos Aires* (TI-09, TI-39 and I-025), and *Consejo Nacional de Investigaciones Científicas y Técnicas*, CONICET (PIP-4030).

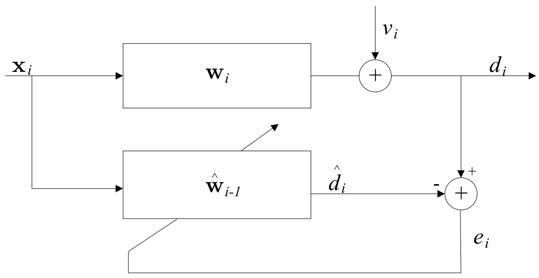

Figure 1: An adaptive filtering problem.

However, the robustness guaranteed by the H$^\infty$ approach is only true along a time interval, with possibly infinite length. That is why we look for energy relations that allow a local robust behavior (robust at each time instant). Following the ideas introduced by Rupp and Sayed [7], we find local and global error energy bounds for the APA family using our modified update.

After analyzing the use of the explicit regularization, we propose to optimize its choice for maximum speed of convergence. We work first using independence assumptions and second with gaussian hypothesis with unspecified coloring. Due to the difficulty of finding closed expressions, we assume certain conditions on the correlation matrix of the weight error vector. In this case, an intuitive result is found: the regularization parameter should be proportional to the noise power and inversely proportional to the weight error power at the previous time instant. A similar result was reported in [4], but using white input and large filter order hypothesis. Although the expressions could be complex, they give information about the relationship between the value of the regularization parameter and the convergence behavior of the algorithm.

The remainder of this work is organized as follows. In section 2 we introduced the APA family and the first approach that supports our modified update. Section 3 states the main results about the H$^\infty$ optimality of the APA family and the local and global error energy bounds. We leave to section 4 the optimization of the regularization factor.

2. THE APA FAMILY

Let $\mathbf{w}_i = \left(w_i^0, w_i^1, \ldots, w_i^{M-1}\right)^T \in \mathbb{C}^{M\times 1}$ be an unknown linear FIR system. The associated adaptive filtering problem is represented in Fig. 1. The input signal at time i, $\mathbf{x}_i = \left(x^i, x^{i-1}, \ldots, x^{i-M+1}\right)^T \in \mathbb{C}^{M\times 1}$, pass through the system giving an output $\mathbf{w}_i^H \mathbf{x}_i \in \mathbb{C}$. This output is observed, but in this process it usually appears a measurement noise, $v_i \in \mathbb{C}$, which will be considered as additive. Thus, each input $\mathbf{x}_i$ gives an output $d_i = \mathbf{w}_i^H \mathbf{x}_i + v_i$.

The idea is to find $\hat{\mathbf{w}}_i$ to estimate $\mathbf{w}_i$. This filter receives the same input, leading to an output estimation error $e_i = d_i - \hat{\mathbf{w}}_i^H \mathbf{x}_i$. The way in which $\hat{\mathbf{w}}_i$ is computed depends on the cost function to be optimized. In the sequel, we assume that the system is LTI (Linear Time Invariant).

When data blocks are used, we can define the data matrix $\mathbf{X}_i = [\mathbf{x}_i \, \mathbf{x}_{i-1} \cdots \mathbf{x}_{i-L}] \in \mathbb{C}^{M\times(L+1)}$, the desired vector $\mathbf{d}_i = [d_i \, d_{i-1} \cdots d_{i-L}]^T \in \mathbb{C}^{(L+1)\times 1}$ and the error vector $\mathbf{e}_i^* = \mathbf{d}_i^* - \mathbf{X}_i^H \hat{\mathbf{w}}_{i-1} \in \mathbb{C}^{(L+1)\times 1}$.
The APA was first introduced in [5], as the recursion:

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

where μ_i is a scalar known as step size, included to control the changes along the selected direction. Moreover, setting $L = 0$ in (1), leads to the popular NLMS algorithm.

An interesting interpretation of the APA comes from looking for the least squares solution that minimizes the square norm of $\delta\hat{\mathbf{w}}_i = \hat{\mathbf{w}}_i - \hat{\mathbf{w}}_{i-1}$, subject to the constrain $\mathbf{e}_{p,i}^* = \mathbf{d}_i^* - \mathbf{X}_i^H \hat{\mathbf{w}}_i = 0$. This shows that the APA sets to zero the last $L+1$ *a posteriori* errors.

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 [8] 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.

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 argument, a positive *regularization* term is usually added.

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:

$$\delta\hat{\mathbf{w}}_i = \mathbf{X}_i \left(\beta_i \mathbf{I}_L + \mathbf{X}_i^H \mathbf{X}_i\right)^{-1} \mathbf{e}_i^* \quad i = 0, 1, \ldots, N \tag{2}$$

This rule gives an "effective step size" in the interval $(0,1)$ for any positive β_i. So there is no superior bound that could make the algorithm unstable.

3. ROBUST BEHAVIOR OF THE APA FAMILY

There are other very important reasons for preferring the modified update, and they will be stated in this section. The main consequence is that using (2) makes APA more robust to perturbations and model uncertainties. This is a very important issue when dealing with highly noisy environments, time variant model parameters or when the tracking capability becomes a major problem.

We show the robust behavior of the APA family from two approaches: H$^\infty$ optimality and error energy bounds.

3.1 H$^\infty$ Optimality of the APA Family

Perturbations are something that any algorithm has to lead with in a real world implementation. They have many different sources: parameters variation with time, initial conditions errors, measurement noise, modeling errors, numerical precision, etc. The robustness is a concept associated with the sensibility of an algorithm to the perturbations. If we follow a deterministic framework, an algorithm is robust if it does not amplify the perturbations energy.

In the middle of the 90's, Hassibi *et al.* present the relationship between H$^\infty$ optimal filters and Kalman filtering in Krein spaces [3]. We can apply this theory for showing that the APA family is H$^\infty$ optimal. More precisely, it can be proved that the APA as in (2) is the solution to:

$$\gamma_{p,opt}^2 = \inf_{\mathscr{F}_p} \sup_{\mathbf{w}_0, \{\mathbf{v}_i\} \in h_2} \frac{\sum_{j=0}^{i} \beta_j^{-1} ||\mathbf{e}_{p,j}||^2}{||\mathbf{w}_0 - \hat{\mathbf{w}}_0||^2 + \sum_{j=0}^{i} \beta_j^{-1} ||\mathbf{v}_j||^2}. \tag{3}$$

The value $\gamma_{f,opt}^2$ represents a bound on the energy transfer from the perturbations to the estimation errors. The necessary conditions for the result to be true are *exciting* input signals, i.e., $\lim_{N\to\infty} \sum_{i=0}^{N} \mathbf{x}_i^H \mathbf{x}_i = \infty$, and $\beta_i > 0$.

The fact that β_i acts as a weighting sequence for the energy of the error $\mathbf{e}_i$, and thus for $\mathbf{e}_{p,i}$ and $\mathbf{v}_i$, is an indication that its importance is not restricted to the numerical instability problem.

When the hypothesis of exciting input was used, we assumed that the value of i in (3) – which represents the interval length where the energy is computed – is *sufficiently large*. As a consequence, the problem could become of infinite horizon. We can say nothing about what happens to the energy of the estimation errors just after a noise peak appears at a precise time instant. This "local behavior" needs a different approach, which will be treated in the next subsection.

3.2 Local and Global Error Energy Bounds for APA

We wonder if the APA family guarantees that the energy of the perturbations will never exceed the energy of the estimation errors for all time instants. To do so, we follow the approach introduced by Rupp and Sayed [7] for LMS type of algorithms. We define the weight error vector $\tilde{\mathbf{w}}_i = \mathbf{w} - \hat{\mathbf{w}}_i$, the *a priori* error $\mathbf{e}_{a,i}^* = \mathbf{X}_i^H \tilde{\mathbf{w}}_{i-1}$ and the weighting matrix $\mathbf{S}_i = \left(\beta_i \mathbf{I}_L + \mathbf{X}_i^H \mathbf{X}_i\right)^{-1}$.

If we choose $\beta_i > 0$ for all i, then $\mathbf{S}_i$ is positive definite. Using the APA recursion (2) we show that the following local bounds for the *a posteriori* and *a priori* errors holds:

Theorem 1 *At each time instant $i \in [0,N]$, the following energy bounds apply to the a posteriori and a priori errors respectively:*

$$\frac{\|\tilde{\mathbf{w}}_i\|^2 + \beta_i^{-1} \left\|\mathbf{e}_{p,i}\right\|^2}{\|\tilde{\mathbf{w}}_{i-1}\|^2 + \beta_i^{-1} \|\mathbf{v}_i\|^2} \begin{cases} < 1 & \text{if} \quad \mathbf{e}_i \neq 0 \\ = 1 & \text{if} \quad \mathbf{e}_i = 0 \end{cases} \tag{4}$$

$$\frac{\|\tilde{\mathbf{w}}_i\|^2 + \mathbf{e}_{a,i}^H \mathbf{S}_i^* \mathbf{e}_{a,i}}{\|\tilde{\mathbf{w}}_{i-1}\|^2 + \mathbf{v}_i^H \mathbf{S}_i^* \mathbf{v}_i} \begin{cases} < 1 & \text{if} \quad \mathbf{e}_i \neq 0 \\ = 1 & \text{if} \quad \mathbf{e}_i = 0 \end{cases} \tag{5}$$

Proof: We use some error relations based on its definitions and show that relation

$$\beta_i \left(\|\tilde{\mathbf{w}}_{i-1}\|^2 - \|\tilde{\mathbf{w}}_i\|^2\right) + \|\mathbf{v}_i\|^2 - \|\mathbf{e}_{p,i}\|^2,$$

in the case of (4) and

$$\left(\|\tilde{\mathbf{w}}_{i-1}\|^2 - \|\tilde{\mathbf{w}}_i\|^2\right) + \mathbf{v}_i^H \mathbf{S}_i^* \mathbf{v}_i - \mathbf{e}_{a,i}^H \mathbf{S}_i^* \mathbf{e}_{a,i},$$

for (5), are positive. Both facts require only $\beta_i > 0$. ■

A first interpretation of these bounds is that **for all time instants** i, the energy of the estimation errors **never** exceeds the energy of the perturbations.

As the local bounds are valid for all i, if a time interval of length N is taken, global error bounds are obtained as a generalization. In this case:

$$\frac{\|\tilde{\mathbf{w}}_N\|^2 + \sum_{i=0}^{N} \beta_i^{-1} \|\mathbf{e}_{p,i}\|^2}{\|\tilde{\mathbf{w}}_{-1}\|^2 + \sum_{i=0}^{N} \beta_i^{-1} \|\mathbf{v}_i\|^2} \begin{cases} < 1 \text{ if } \exists i \in [0,N] : \mathbf{e}_i \neq 0 \\ = 1 \text{ if } \forall i \in [0,N] : \mathbf{e}_i = 0 \end{cases} \quad (6)$$

$$\frac{\|\tilde{\mathbf{w}}_N\|^2 + \sum_{i=0}^{N} \mathbf{e}_{a,i}^H \mathbf{S}_i^* \mathbf{e}_{a,i}}{\|\tilde{\mathbf{w}}_{-1}\|^2 + \sum_{i=0}^{N} \mathbf{v}_i^H \mathbf{S}_i^* \mathbf{v}_i} \begin{cases} < 1 \text{ if } \exists i \in [0,N] : \mathbf{e}_i \neq 0 \\ = 1 \text{ if } \forall i \in [0,N] : \mathbf{e}_i = 0 \end{cases} \quad (7)$$

The H^∞ and the error bounds approaches are concerned with the robustness of the APA family. As the global error bounds are valid for any interval length, we would like the two results to be equivalent in the infinite horizon case.

As $\gamma_{p,opt}^2 = 1$ in (3), the two approaches will be asymptotically equivalent if:

$$\lim_{N\to\infty} \hat{\mathbf{w}}_N = \mathbf{w}. \quad (8)$$

By following a similar procedure to the one used in [7] for LMS type algorithms, this (deterministic) convergence may be proved.

There is one more thing we would like to point out. For the first approach we proved that the APA is the algorithm that minimizes the energy relation (3), over all possible estimation strategies. For the second one, the bound (6) shows that actually, the APA family can reach a more tight energy relation, as $\|\tilde{\mathbf{w}}_N\|^2$ is always positive.

This difference should not despise the H^∞ approach. The local and global bounds are satisfied by the APA family as a consequence of its way of finding the system estimation. On the other hand, the theory of linear estimation in Krein spaces is a powerful and elegant tool for the design and implementation of H^∞ optimal filters.

There is still the issue of choosing the sequence β_i, which will be solved in the next section.

4. OPTIMAL REGULARIZATION CHOICE

In this section we propose to maximize the speed of convergence by choosing for each i the value β_i which minimizes $E[||\tilde{\mathbf{w}}_i||^2]$. As the weight error recursion is

$$\tilde{\mathbf{w}}_i = (\mathbf{I}_M - \tilde{\mathbf{P}}_i)\tilde{\mathbf{w}}_{i-1} - \mathbf{X}_i(\mathbf{X}_i^H\mathbf{X}_i + \beta_i\mathbf{I}_L)^{-1}\mathbf{v}_i^*,$$

with $\tilde{\mathbf{P}}_i = \mathbf{X}_i(\mathbf{X}_i^H\mathbf{X}_i + \beta_i\mathbf{I}_L)^{-1}\mathbf{X}_i^H$, if the noise is independent of the data matrix, then:

$$E[||\tilde{\mathbf{w}}_i||^2] = E[\tilde{\mathbf{w}}_{i-1}^H(\mathbf{I}_M - \tilde{\mathbf{P}}_i)^2\tilde{\mathbf{w}}_{i-1}] + \mathrm{tr}(\mathbf{R}_{\mathbf{v}_i}\tilde{\mathbf{K}}_i), \quad (9)$$

where $\tilde{\mathbf{K}}_i = E[(\mathbf{X}_i^H\mathbf{X}_i + \beta_i\mathbf{I}_L)^{-1}\mathbf{X}_i^H\mathbf{X}_i(\mathbf{X}_i^H\mathbf{X}_i + \beta_i\mathbf{I}_L)^{-1}]$ and $\mathbf{R}_{\mathbf{v}_i}$ is the correlation matrix of the noise.

4.1 Analysis under the Independence Assumption

If $\tilde{\mathbf{w}}_{i-1}$ is independent of $\tilde{\mathbf{P}}_i$ and the noise is white (with zero mean and variance $\sigma_{v_i}^2$), we have from (9):

$$\begin{aligned} E[||\tilde{\mathbf{w}}_i||^2] = E[||\tilde{\mathbf{w}}_{i-1}||^2 + \mathrm{tr}(\mathbf{R}_{\tilde{\mathbf{w}}_{i-1}}E[\tilde{\mathbf{P}}_i^2]) \\ - 2\,\mathrm{tr}(\mathbf{R}_{\tilde{\mathbf{w}}_{i-1}}E[\tilde{\mathbf{P}}_i]) + \sigma_{v_i}^2\mathrm{tr}(\tilde{\mathbf{K}}_i), \end{aligned} \quad (10)$$

with $\mathbf{R}_{\tilde{\mathbf{w}}_{i-1}}$ being the correlation matrix of the weight error.

Now, we perform a *Singular Value Decomposition* (SVD) on the data matrix, i.e., $\mathbf{X}_i = \mathbf{U}_i\mathbf{\Sigma}_i\mathbf{V}_i^H$. So $\tilde{\mathbf{P}}_i = \mathbf{U}_i\mathbf{\Delta}_{\tilde{\mathbf{P}}_i}\mathbf{U}_i^H$, where $\mathbf{\Delta}_{\tilde{\mathbf{P}}_i}$ is a diagonal matrix with nonzero elements of the form:

$$(\mathbf{\Delta}_{\tilde{\mathbf{P}}_i})_{kk} = \frac{(\rho_i^k)^2}{(\rho_i^k)^2 + \beta_i},$$

and ρ_i^k are the singular values of $\mathbf{X}_i$.

It could also be seen that $\tilde{\mathbf{K}}_i = \mathbf{V}_i\mathbf{\Delta}_{\tilde{\mathbf{K}}_i}\mathbf{V}_i^H$, with the central diagonal matrix elements represented by:

$$(\mathbf{\Delta}_{\tilde{\mathbf{K}}_i})_{kk} = \frac{(\rho_i^k)^2}{((\rho_i^k)^2 + \beta_i)^2}.$$

By making the replacements in (10), differentiating partially towards β_i – assuming uncorrelation with respect to β_k for $k \neq i$ – and setting it to zero, we find:

$$\mathrm{tr}(\mathbf{R}_{\tilde{\mathbf{w}}_{i-1}}E[\mathbf{U}_i\mathbf{\Delta}_{l_i}\mathbf{U}_i^H]) = \sigma_{v_i}^2\mathrm{tr}(E[\mathbf{V}_i\mathbf{\Delta}_{r_i}\mathbf{V}_i^H]), \quad (11)$$

where we have defined the diagonal matrices:

$$(\mathbf{\Delta}_{l_i})_{kk} = \frac{\beta_i(\rho_i^k)^2}{((\rho_i^k)^2 + \beta_i)^3}, \qquad (\mathbf{\Delta}_{r_i})_{kk} = \frac{(\rho_i^k)^2}{((\rho_i^k)^2 + \beta_i)^3}. \quad (12)$$

Solving (11) for the general case requires information of the correlation matrix of the filter estimation error and the singular values of the input data matrix. If $\mathbf{R}_{\tilde{\mathbf{w}}_{i-1}} = \sigma_{\tilde{\mathbf{w}}_{i-1}}^2\mathbf{I}_M$, from (11), it could be seen that:

$$\beta_i = \frac{\sigma_{v_i}^2}{\sigma_{\tilde{\mathbf{w}}_{i-1}}^2}. \quad (13)$$

This result is intuitively correct. When the background noise becomes large, so should be β_i in order to attenuate its amplification. On the other hand, when the filter estimation error grows, the regularization parameter should decrease, allowing the algorithm to quickly reestimate the system.
We want to remark that it also holds:

$$\sigma_{e_{a,i}}^2 = \mathrm{tr}(\mathbf{R}_{\tilde{\mathbf{w}}_{i-1}}E[\mathbf{x}_i\mathbf{x}_i^H]) = \sigma_{\tilde{\mathbf{w}}_{i-1}}^2 M\sigma_{x_i}^2,$$

and replacing it in (13), leads to:

$$\beta_i = \frac{\sigma_{v_i}^2 M\sigma_{x_i}^2}{\sigma_{e_{a,i}}^2}, \quad (14)$$

which is the result presented in [4]. In that work, white input excitation and large filter orders were assumed.

4.2 Analysis for Gaussian Hypothesis

We replace the SVD decomposition in (9). Thus,

$$\begin{aligned} E[||\tilde{\mathbf{w}}_i||^2] = E[||\tilde{\mathbf{w}}_{i-1}||^2 + E[\tilde{\mathbf{w}}_{i-1}^H\mathbf{U}_i\mathbf{\Delta}_{\tilde{\mathbf{P}}_i}^2\mathbf{U}_i^H\tilde{\mathbf{w}}_{i-1}] \\ -2E[\tilde{\mathbf{w}}_{i-1}^H\mathbf{U}_i\mathbf{\Delta}_{\tilde{\mathbf{P}}_i}\mathbf{U}_i^H\tilde{\mathbf{w}}_{i-1}] + E[\mathbf{v}_i^H\mathbf{V}_i\mathbf{\Delta}_{\tilde{\mathbf{K}}_i}\mathbf{V}_i^H\mathbf{v}_i] \end{aligned} \quad (15)$$

By defining the k-*th* normalized singular vector as $\underline{\mathbf{u}}_i^k = (\mathbf{\Delta}_{\tilde{\mathbf{P}}_i})_{kk}^{1/2}\mathbf{u}_i^k$, we can write:

$$E[\tilde{\mathbf{w}}_{i-1}^H \mathbf{U}_i \mathbf{\Delta}_{\tilde{\mathbf{P}}_i} \mathbf{U}_i^H \tilde{\mathbf{w}}_{i-1}] = E[\tilde{\mathbf{w}}_{i-1}^H \sum_{k=1}^{L+1} (\mathbf{\Delta}_{\tilde{\mathbf{P}}_i})_{kk} \mathbf{u}_i^k (\mathbf{u}_i^k)^H \tilde{\mathbf{w}}_{i-1}]$$
$$= \sum_{k=1}^{L+1} E[\tilde{\mathbf{w}}_{i-1}^H \underline{\mathbf{u}}_i^k (\underline{\mathbf{u}}_i^k)^H \tilde{\mathbf{w}}_{i-1}]. \quad (16)$$

If $\tilde{\mathbf{w}}_{i-1}$ and $\underline{\mathbf{u}}_i^k$ are assumed to be gaussian distributed, we can apply the gaussian moment factoring theorem to each term of (16), leading to:

$$E[\tilde{\mathbf{w}}_{i-1}^H \mathbf{U}_i \mathbf{\Delta}_{\tilde{\mathbf{P}}_i} \mathbf{U}_i^H \tilde{\mathbf{w}}_{i-1}] = \sum_{k=1}^{L+1} E[\tilde{\mathbf{w}}_{i-1}^H \underline{\mathbf{u}}_i^k] E[(\underline{\mathbf{u}}_i^k)^H \tilde{\mathbf{w}}_{i-1}]$$
$$+ \operatorname{tr}(\mathbf{R}_{\tilde{\mathbf{w}}_{i-1}} E[\mathbf{U}_i \mathbf{\Delta}_{\tilde{\mathbf{P}}_i} \mathbf{U}_i^H])$$

Following a similar process for the last term of (15) and assuming independence between the input and noise sequences, we obtain,

$$E[\mathbf{v}_i^H \mathbf{V}_i \mathbf{\Delta}_{\tilde{\mathbf{K}}_i} \mathbf{V}_i^H \mathbf{v}_i] = \operatorname{tr}(\mathbf{R}_{\mathbf{v}_i} E[\mathbf{V}_i \mathbf{\Delta}_{\tilde{\mathbf{K}}_i} \mathbf{V}_i^H])$$

By making the replacements in (15), differentiating partially towards β_i and setting it to zero, it results that the optimum regularization parameter satisfies:

$$2\operatorname{tr}(\mathbf{R}_{\tilde{\mathbf{w}}_{i-1}} E[\mathbf{U}_i \mathbf{\Delta}_{l_i} \mathbf{U}_i^H]) - 2\operatorname{tr}(\mathbf{R}_{\mathbf{v}_i} E[\mathbf{V}_i \mathbf{\Delta}_{r_i} \mathbf{V}_i^H])$$
$$= \Bigg[\sum_{k=1}^{L+1} E[(\mathbf{\Delta}_{\tilde{\mathbf{K}}_i})_{kk} \tilde{\mathbf{w}}_{i-1}^H \mathbf{u}_i^k] E[(\mathbf{\Delta}_{\tilde{\mathbf{P}}_i})_{kk} (\mathbf{u}_i^k)^H \tilde{\mathbf{w}}_{i-1}]$$
$$+ E[(\mathbf{\Delta}_{\tilde{\mathbf{P}}_i})_{kk} \tilde{\mathbf{w}}_{i-1}^H \mathbf{u}_i^k] E[(\mathbf{\Delta}_{\tilde{\mathbf{K}}_i})_{kk} (\mathbf{u}_i^k)^H \tilde{\mathbf{w}}_{i-1}]\Bigg]$$
$$- \Bigg[\sum_{k=1}^{L+1} E[(\mathbf{\Delta}_{r_i})_{kk}^{1/2} \tilde{\mathbf{w}}_{i-1}^H \mathbf{u}_i^k] E[(\mathbf{\Delta}_{\tilde{\mathbf{P}}_i})_{kk}^{1/2} (\mathbf{u}_i^k)^H \tilde{\mathbf{w}}_{i-1}]$$
$$+ E[(\mathbf{\Delta}_{\tilde{\mathbf{P}}_i})_{kk}^{1/2} \tilde{\mathbf{w}}_{i-1}^H \mathbf{u}_i^k] E[(\mathbf{\Delta}_{r_i})_{kk}^{1/2} (\mathbf{u}_i^k)^H \tilde{\mathbf{w}}_{i-1}]\Bigg] \quad (17)$$

where $\mathbf{\Delta}_{l_i}$ and $\mathbf{\Delta}_{r_i}$ were previously defined in (12). It becomes evident the difficulty of finding a close expression for β_i from the optimum condition given by (17). If $\mathbf{R}_{\mathbf{v}_i}$ and $\mathbf{R}_{\tilde{\mathbf{w}}_{i-1}}$ are diagonal, then:

$$\operatorname{tr}(\mathbf{R}_{\tilde{\mathbf{w}}_{i-1}} E[\mathbf{U}_i \mathbf{\Delta}_{l_i} \mathbf{U}_i^H]) - \operatorname{tr}(\mathbf{R}_{\mathbf{v}_i} E[\mathbf{V}_i \mathbf{\Delta}_{r_i} \mathbf{V}_i^H])$$
$$= (\beta_i \sigma_{\tilde{\mathbf{w}}_{i-1}}^2 - \sigma_{v_i}^2) \sum_{k=1}^{L+1} E[(\mathbf{\Delta}_{r_i})_{kk}],$$

and when we replace it in (17), we find,

$$(\beta_i \sigma_{\tilde{\mathbf{w}}_{i-1}}^2 - \sigma_{v_i}^2) \sum_{k=1}^{L+1} E[(\mathbf{\Delta}_{r_i})_{kk}]$$
$$= \operatorname{Re}\left\{\sum_{k=1}^{L+1} E[(\mathbf{\Delta}_{\tilde{\mathbf{K}}_i})_{kk} \tilde{\mathbf{w}}_{i-1}^H \mathbf{u}_i^k] E[(\mathbf{\Delta}_{\tilde{\mathbf{P}}_i})_{kk} \tilde{\mathbf{w}}_{i-1}^H \mathbf{u}_i^k]^H\right\}$$
$$- \operatorname{Re}\left\{\sum_{k=1}^{L+1} E[(\mathbf{\Delta}_{r_i})_{kk}^{1/2} \tilde{\mathbf{w}}_{i-1}^H \mathbf{u}_i^k] E[(\mathbf{\Delta}_{\tilde{\mathbf{P}}_i})_{kk}^{1/2} \tilde{\mathbf{w}}_{i-1}^H \mathbf{u}_i^k]^H\right\} \quad (18)$$

where $\operatorname{Re}\{\cdot\}$ means taking the real part. Now we write $\tilde{\mathbf{w}}_{i-1}$ as a linear combination of the singular vectors $\mathbf{u}_i^k$ [1],

$$\tilde{\mathbf{w}}_{i-1} = \sum_{j=1}^{M} \alpha_i^j \mathbf{u}_i^j.$$

As the vectors form an orthonormal basis and if we assume that α_i^k has zero mean and is independent of ρ_i^k, we can rewrite the right side of (18) as,

$$\sum_{k=1}^{L+1} |E[\alpha_i^k]|^2 \Big[E[(\mathbf{\Delta}_{\tilde{\mathbf{K}}_i})_{kk}] E[(\mathbf{\Delta}_{\tilde{\mathbf{P}}_i})_{kk}]$$
$$- E[(\mathbf{\Delta}_{r_i})_{kk}^{1/2}] E[(\mathbf{\Delta}_{\tilde{\mathbf{P}}_i})_{kk}^{1/2}]\Big] = 0.$$

The complex expressions (11) and (17) are optimal for each analyzed case. When we assume that $\mathbf{R}_{\tilde{\mathbf{w}}_{i-1}}$ is diagonal, both of them become equivalent to the simple result (13), where $\sigma_{\tilde{\mathbf{w}}_{i-1}}^2$ can be estimated by the method of delayed coefficients [2].

5. CONCLUSIONS

A modified update was proposed for the APA family, which includes the explicit regularization factor. The "effective step size" of the algorithm is between 0 and 1 for all $\beta > 0$, so there is no upper bound that could cause instabilities.

Particularly, the explicit regularization factor does not only help on dealing with numerical precision problems, but also allows a robust performance against all possible perturbations. This was justified from the H^∞ optimality and the local error energy obtained.

We also performed an analysis for optimizing β_i to have maximum speed of convergence. First we used independence assumptions and then we dealed with the gaussian case. Complex expressions were derived, which results in a simple one if certain conditions on the weight error correlation matrix are assumed.

REFERENCES

[1] N.J. Bershad, D. Lincharger and S. McLaughlin, "A Stochastic Analysis of the Affine Projection Algorithm for Gaussian Autoregressive Inputs", *Proc. ICASSP-01*, Salt Lake City, USA, May. 2001, pp. 3837–3840.

[2] S. Haykin, *Adaptive Filter Theory*, Fourth Edition, Prentice Hall, Upper Saddle River, New Jersey, 2002.

[3] B. Hassibi, A.H. Sayed and T. Kailath, "Linear Estimation in Krein Spaces - Parts I and II", *IEEE Trans. Automat. Contr.*, Vol. 41, No. 1, pp. 18–49, Jan. 1996.

[4] V. Myllylä and G. Schmidt, "Pseudo-Optimal Regularization for Affine Projection Algorithms", *Proc. ICASSP-02*, Orlando, USA, May 2002, pp. 1917–1920.

[5] 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, 1984.

[6] G. Rombouts and M. Moonen , "Avoiding Explicit Regularisation in Affine Projection Algorithms for Acoustic Echo Cancellation", *Proc. ProRISC-99*, Mierlo, The Netherlands, Nov. 1999, pp. 395–398.

[7] M. Rupp and A.H. Sayed , "A Time-Domain Feedback Analysis of Filtered-Error Adaptive Gradient Algorithms", *IEEE Trans. on Signal Processing*, Vol. 44, No. 6, pp. 1428–1439, Jun. 1996.

[8] S.G. Sankaran and A.A.L. Beex, "Convergence Behavior of Affine Projection Algorithms", *IEEE Trans. on Signal Processing*, Vol. 48, No. 4, pp. 1086–1096, Apr. 2000.