ANALYSIS OF A SET-MEMBERSHIP AFFINE PROJECTION ALGORITHM IN NONSTATIONARY ENVIRONMENT

Paulo S. R. Diniz

LPS – Programa de Engenharia Elétrica, COPPE/Poli/UFRJ, Cx. P. 68504, Rio de Janeiro, RJ, BRAZIL
email: diniz@lps.ufrj.br

ABSTRACT

Set-membership (SM) adaptive filters have data-selective coefficient update leading to lower computational complexity and power consumption. The set-membership affine projection (SM-AP) algorithm has been known for not trading convergence speed with misadjustment and computation complexity. In this paper analytical results related to the SM-AP algorithm in nonstationary environments are advanced. The analysis results for the excess of mean square error (MSE) in nonstationary environments are shown to be quite accurate confirming the attractive features of the SM-AP algorithms.

1. INTRODUCTION

The affine projection (AP) algorithm, first proposed in [1], is widely discussed in the open literature due its faster convergence than the stochastic gradient algorithms, such as the LMS, and its lower computational complexity than the RLS algorithm [1]-[6]. However, the AP algorithm trades off computational complexity with convergence speed. Set-membership (SM) adaptive filtering algorithms [7]-[14] have been increasingly discussed since they reduce the computational burden while keeping low misadjustment and fast convergence. As a result, the combination of SM and AP results in computationally efficient algorithms with low misadjustment and high convergence speed, such as the SM-AP algorithms [9]. Analytical results concerning the SM-AP algorithms are scarce in the open literature [2], whereas results concerning nonstationary environments are not available so far. The objective of this paper is to propose analytical expressions for the behavior of the SM-AP algorithms in nonstationary environments

This paper is organized as follows. In Section 2 the SM-AP algorithms are briefly presented along with their energy conservation equations. Section 3 addresses the tracking performance of the SM-AP algorithms in simple nonstationary environments, where the unknown system parameters are modeled as first-order Markov processes. Section 4 presents some simulation results which confirm the validity of the analytic expressions.

2. SET-MEMBERSHIP AFFINE PROJECTION ALGORITHM

Let's define the adaptive filter output

$$y(k) = \mathbf{w}^T\mathbf{x}(k) \tag{1}$$

where $\mathbf{x}(k) = [x_0(k)\ \ x_1(k)\ \ \ldots x_N(k)]^T$ is the input signal vector, and $\mathbf{w} = [w_0\ \ w_1\ \ \ldots w_N]^T$ is the parameter vector. Assuming the availability of a reference signal sequence $d(k)$ and a sequence of input vectors $\mathbf{x}(k)$, both for $k = 0, 1, 2, \ldots, \infty$, the estimation error sequence $e(k)$ is defined as

$$e(k) = d(k) - \mathbf{w}^T\mathbf{x}(k) = d(k) - y(k) \tag{2}$$

The vectors $\mathbf{x}(k)$ and $\mathbf{w} \in \mathbf{R}^{N+1}$, where $\mathbf{R}$ represents the set of real numbers, whereas $e(k)$ and $y(k)$ represent the output error and adaptive filter output signal, respectively. The objective of the SMF is to choose $\mathbf{w}$ such that the magnitude of estimation output error is upper bounded by a prescribed quantity γ. If the value of γ is properly chosen, several valid estimates for $\mathbf{w}$ exist. That means any filter parameter is acceptable as long as the magnitude of the output estimation error is smaller than the deterministic threshold γ. The bounded error constraint results in a set of estimates rather than a single one. If γ is chosen too small there might be no solution.

In actual applications only measured data are available. Given a set of data pairs $\{\mathbf{x}(i), d(i)\}$, for $i = 0, 1, \ldots, k$, we can define $\mathcal{H}(k)$ as the set containing all vectors $\mathbf{w}$ such that the associated output error at time instant k is upper bounded in magnitude by γ. That means,

$$\mathcal{H}(k) = \{\mathbf{w} \in \mathbb{R}^{N+1} : |d(k) - \mathbf{w}^T\mathbf{x}(k)| \leq \gamma\} \tag{3}$$

The set $\mathcal{H}(k)$ is known as the *constraint set*. The boundaries of $\mathcal{H}(k)$ are hyperplanes. For the two-dimensional case, where the coefficient vector has two elements, $\mathcal{H}(k)$ represents the region between the lines where $d(k) - \mathbf{w}^T\mathbf{x}(k) = \pm\gamma$. For more than two dimensions, $\mathcal{H}(k)$ represents the region between two parallel hyperplanes in the parameter space $\mathbf{w}$.

Each data pair is associated with a constraint set. As a consequence the intersection of the constraint sets over all the available iterations $i = 0, 1, \ldots, k$, is called the *exact membership set* $\psi(k)$, formally defined as

$$\psi(k) = \bigcap_{i=0}^{k}\mathcal{H}(i) \tag{4}$$

The set $\psi(k)$ represents a polygon in the parameter space, and one of the main objectives of the SMF is to find the polygon location. In the early iterations it is likely that the constraint set reduces the size of the membership-set polygon. The polygon in $\mathbf{w}$, represented by $\psi(k)$, becomes small if the set of data pairs includes substantial innovation. This condition is usually met after a large number of iterations k, when most likely $\psi(k) = \psi(k-1)$, where $\psi(k-1)$ is already placed inside the constraint set $\mathcal{H}(k)$. In such situation, the parameters do not require updating since the current membership set is inside the constraint set, giving rise to a data-dependent selection of update.

The SM-AP algorithm minimizes the objective function by performing a coefficient update whenever $\mathbf{w}(k) \notin \psi^{L+1}(k)$ in such a way that

$$\min \|\mathbf{w}(k+1) - \mathbf{w}(k)\|^2 \tag{5}$$

subject to:

$$\mathbf{d}(k) - \mathbf{X}^T(k)\mathbf{w}(k+1) = \boldsymbol{\gamma}(k) \tag{6}$$

where $\mathbf{d}(k)$ is the desired signal vector and $\mathbf{X}(k)$ is the input signal matrix. Fig. 1 depicts a typical coefficient update related to the SM-AP algorithm for the case with two parameters, i.e., for $L = 1$ and $|\gamma_i(k)| < |\gamma|$, such that $\mathbf{w}(k+1)$ is not placed at the border of $\mathcal{H}(k)$.

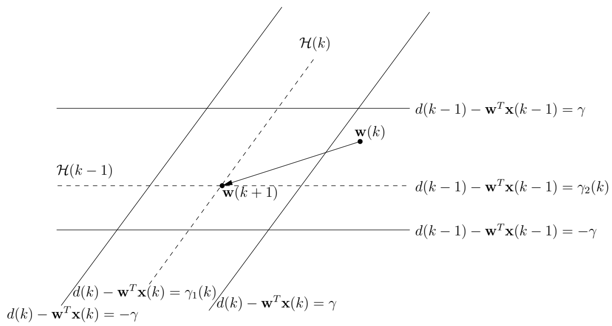

Figure 1: SM-AP algorithm coefficient update, $L = 1$.

The updating equation of the SM-AP algorithm is given by

$$\mathbf{w}(k+1) = \begin{cases} \mathbf{w}(k) + \mathbf{X}(k)\left[\mathbf{X}^T(k)\mathbf{X}(k)\right]^{-1}[\mathbf{e}(k) - \boldsymbol{\gamma}(k)] & \text{if } |e(k)| > \gamma \\ \mathbf{w}(k) & \text{otherwise} \end{cases} \quad (7)$$

where

$$\mathbf{e}(k) = [e(k)\ \varepsilon(k-1)\ \ldots\ \varepsilon(k-L)]^T \quad (8)$$

with $\varepsilon(k-i) = d(k-i) - \mathbf{x}^T(k-i)\mathbf{w}(k)$ denoting the *a posteriori* error calculated with the data pair of iteration $k-i$ using the coefficients of iteration k, for $k = 1, \ldots, L$. The general description of the SM-AP algorithm is described in detail below.

The Set-Membership Affine Projection Algorithm
Initialization
$\mathbf{x}(0) = \mathbf{w}(0) = [0\ \ldots\ 0]^T$
choose γ around $\sqrt{5}\sigma_n$ (to be explained)
$\delta =$ small constant
Do for $k \geq 0$
$\mathbf{e}(k) = \mathbf{d}(k) - \mathbf{X}^T(k)\mathbf{w}(k)$

$$\mathbf{w}(k+1) = \begin{cases} \mathbf{w}(k) + \mathbf{X}(k)\left[\mathbf{X}^T(k)\mathbf{X}(k) + \delta\mathbf{I}\right]^{-1}[\mathbf{e}(k) - \boldsymbol{\gamma}(k)] \\ \qquad \text{if } |e(k)| > \gamma \\ \mathbf{w}(k) \qquad \text{otherwise} \end{cases}$$

Whenever required, the updating equation of the set-membership affine projection algorithm has the following form

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \mathbf{X}(k)\left[\mathbf{X}^T(k)\mathbf{X}(k) + \delta\mathbf{I}\right]^{-1}[\mathbf{e}(k) - \boldsymbol{\gamma}(k)] \quad (9)$$

Before proceeding it should be considered that a coefficient update will not take place all the time. This can be addressed by associating to the coefficient updating equation of the SM-AP algorithm a probability of updating denoted by $P_{\text{up}}(k)$, with its model briefly described in Appendix I. Assuming that the desired signal is given by

$$d(k) = \mathbf{w}_o^T\mathbf{x}(k) + n(k) \quad (10)$$

the underlying updating equation can be alternatively described by

$$\Delta\mathbf{w}(k+1) = \Delta\mathbf{w}(k) + P_{\text{up}}(k)\mathbf{X}(k)\left[\mathbf{X}^T(k)\mathbf{X}(k) + \delta\mathbf{I}\right]^{-1}[\mathbf{e}(k) - \boldsymbol{\gamma}(k)] \quad (11)$$

where $\Delta\mathbf{w}(k) = \mathbf{w}(k) - \mathbf{w}_o$.

By premultiplying equation (11) by the input vector matrix, the following expressions result

$$\mathbf{X}^T(k)\Delta\mathbf{w}(k+1) = \mathbf{X}^T(k)\Delta\mathbf{w}(k) + P_{\text{up}}(k)\mathbf{X}^T(k)\mathbf{X}(k)\left[\mathbf{X}^T(k)\mathbf{X}(k) + \delta\mathbf{I}\right]^{-1}[\mathbf{e}(k) - \boldsymbol{\gamma}(k)]$$

$$-\tilde{\boldsymbol{\varepsilon}}(k) = -\tilde{\mathbf{e}}(k) + P_{\text{up}}(k)\mathbf{X}^T(k)\mathbf{X}(k)\left[\mathbf{X}^T(k)\mathbf{X}(k) + \delta\mathbf{I}\right]^{-1}[\mathbf{e}(k) - \boldsymbol{\gamma}(k)] \quad (12)$$

where

$$\tilde{\boldsymbol{\varepsilon}}(k) = -\mathbf{X}^T(k)\Delta\mathbf{w}(k+1) \quad (13)$$

is the noiseless *a posteriori* error vector and

$$\tilde{\mathbf{e}}(k) = -\mathbf{X}^T(k)\Delta\mathbf{w}(k) = \mathbf{e}(k) - \mathbf{n}(k) \quad (14)$$

is the noiseless *a priori* error vector with

$$\mathbf{n}(k) = [\ n(k)\ \ n(k-1)\ \ \ldots\ \ n(k-L)\ ]^T$$

representing the standard noise vector.

Let's now define the following quantity

$$\underline{\mathbf{e}}(k) = \mathbf{e}(k) - \boldsymbol{\gamma}(k) \quad (15)$$

With the above definition, by solving equation (12), we get

$$\left(\mathbf{X}^T(k)\mathbf{X}(k)\right)^{-1}(\tilde{\mathbf{e}}(k) - \tilde{\boldsymbol{\varepsilon}}(k)) = P_{\text{up}}(k)\left(\mathbf{X}^T(k)\mathbf{X}(k) + \delta\mathbf{I}\right)^{-1}\underline{\mathbf{e}}(k)$$

Multiplying both sides by $\mathbf{X}(k)$ and then replacing the equation above in equation (11), the resulting expression is given by

$$\Delta\mathbf{w}(k+1) - \mathbf{X}(k)\left(\mathbf{X}^T(k)\mathbf{X}(k)\right)^{-1}\tilde{\mathbf{e}}(k) = \Delta\mathbf{w}(k) - \mathbf{X}(k)\left(\mathbf{X}^T(k)\mathbf{X}(k)\right)^{-1}\tilde{\boldsymbol{\varepsilon}}(k) \quad (16)$$

Assumptions: The additional noise is white and statistically independent of the input signal; the inverse of the inherent correlation matrix is statistically independent of both the *a priori* error and the noises; the error in the coefficients during the transient is independent of the data; the *a priori* error $e(k)$ is modeled as a zero-mean Gaussian process, after convergence. From the above equation it is shown in [2] that

$$(2 - P_{\text{up}}(k))\text{tr}\left\{E[\tilde{\mathbf{e}}(k)\tilde{\mathbf{e}}^T(k)]E[\hat{\mathbf{S}}(k)]\right\} + 2(1 - P_{\text{up}}(k))\text{tr}\left\{E[\underline{\mathbf{n}}(k)\tilde{\mathbf{e}}^T(k)]E[\hat{\mathbf{S}}(k)]\right\} = P_{\text{up}}(k)\text{tr}\left\{E[\underline{\mathbf{n}}(k)\underline{\mathbf{n}}^T(k)]E[\hat{\mathbf{S}}(k)]\right\} \quad (17)$$

where $\hat{\mathbf{S}}(k)$ is the inverse of $\mathbf{X}^T(k)\mathbf{X}(k)$.

Any choice for the thresholds $\gamma_i(k)$ is valid as long as they correspond to points represented by the adaptive filter coefficients in $\mathcal{H}(k-i+1)$, i.e., $|\gamma_i(k)| \leq \gamma$. A particularly simple SM-AP version is obtained if $\gamma_i(k)$ for $i \neq 1$ corresponds to the *a posteriori* error $\varepsilon(k-i+1) = d(k-i+1) - \mathbf{w}^T(k)\mathbf{x}(k-i+1)$ and $\gamma_1(k) = e(k)/|e(k)|$. Since the coefficients were updated considering previous data pairs then at the current iteration it is true that $\mathbf{w}(k) \in \mathcal{H}(k-i+1)$, i.e., $|\varepsilon(k-i+1)| = |d(k-i+1) - \mathbf{x}^T(k-i+1)\mathbf{w}(k)| \leq \gamma$, for $i = 2, \ldots, L+1$. Therefore, by choosing $\gamma_i(k) = \varepsilon(k-i+1)$, for $i \neq 1$, all the elements the *a posteriori errors* remain constant, except for first element. The constraint value $\gamma_1(k)$ can be selected as in the SM-NLMS algorithm where $\gamma_1(k)$ is such that the solution lies at the nearest boundary of $\mathcal{H}(k)$, i.e.,

$$\gamma_1(k) = \gamma \frac{e(k)}{|e(k)|} = \gamma \text{sgn}[e(k)] \tag{18}$$

The resulting update equation is then given by

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \mathbf{X}(k)\left[\mathbf{X}^T(k)\mathbf{X}(k)\right]^{-1} \mu(k)e(k)\mathbf{u}_1 \tag{19}$$

where $\mathbf{u}_1 = [1\ 0\ \ldots\ 0]^T$,

$$e(k) = d(k) - \mathbf{w}^T(k)\mathbf{x}(k) \tag{20}$$

$$\mu(k) = \begin{cases} 1 - \frac{\gamma}{|e(k)|} & \text{if } |e(k)| > \gamma \\ 0 & \text{otherwise} \end{cases} \tag{21}$$

This algorithm minimizes the Euclidean distance $\|\mathbf{w}(k+1) - \mathbf{w}(k)\|^2$ subject to the constraint $\mathbf{w}(k+1) \in \psi^{L+1}(k)$ such that the *a posteriori* errors at iteration $k-i$, $\varepsilon(k-i)$, are kept constant for $i = 2, \ldots, L+1$. Fig. 2 illustrates a typical coefficient updating for the simplified SM-AP algorithm where it is observed that the *a posteriori* error related to previous data remains unaltered. The simplified SM-AP algorithm

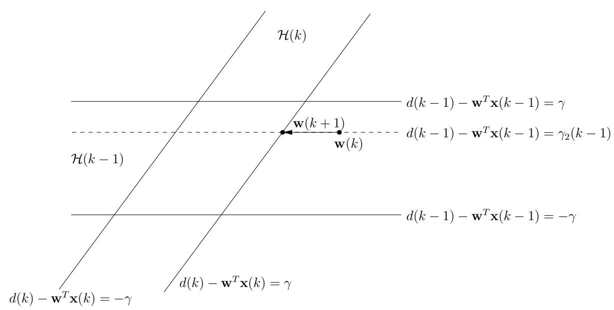

Figure 2: Simplified SM-AP algorithm coefficient update with constant *a posteriori* error, $L = 1$.

given by equation (19) will perform an update if and only if $\mathbf{w}(k) \notin \mathcal{H}(k)$, or $|e(k)| > \gamma$. After some lengthy derivations presented in [2] it is possible to verify that

$$\begin{aligned} E[\tilde{e}_0^2(k)] &\approx \frac{(L+1)P_{\text{up}}}{2 - P_{\text{up}}} \\ &\times \frac{\sigma_n^2 + \gamma^2}{1 + L\left((1-P_{\text{up}})^2 + 2P_{\text{up}}(1-P_{\text{up}})\sqrt{\frac{2}{\pi E[e_0^2(k)]}}\gamma\right)} \end{aligned} \tag{22}$$

In the expression above, γ is the SM threshold and $e_0^2(k)$ denotes the MSE. Therefore, the misadjustment for the set-membership affine projection algorithm is given by

$$\begin{aligned} M &= \frac{(L+1)P_{\text{up}}}{2 - P_{\text{up}}} \\ &\times \frac{\frac{\gamma^2}{\sigma_n^2} + 1}{1 + L\left((1-P_{\text{up}})^2 + 2P_{\text{up}}(1-P_{\text{up}})\sqrt{\frac{2}{\pi E[e_0^2(k)]}}\gamma\right)} \end{aligned} \tag{23}$$

For small $1 - P_{\text{up}}$, this equation can be approximated by

$$M = \frac{(L+1)P_{\text{up}}}{(2 - P_{\text{up}})}\left(\frac{\gamma^2}{\sigma_n^2} + 1\right) \tag{24}$$

3. BEHAVIOR IN NONSTATIONARY ENVIRONMENTS

In a nonstationary environment the error in the coefficients is described by the following vector

$$\Delta\mathbf{w}(k+1) = \mathbf{w}(k+1) - \mathbf{w}_o(k+1) \tag{25}$$

where $\mathbf{w}_o(k+1)$ is the actual time-varying vector. For this case, equation (16) becomes

$$\begin{aligned} \Delta\mathbf{w}(k+1) &= \Delta\hat{\mathbf{w}}(k) \\ &+ P_{\text{up}}(k)\mathbf{X}(k)\left(\mathbf{X}^T(k)\mathbf{X}(k) + \delta\mathbf{I}\right)^{-1}(\mathbf{e}(k) - \boldsymbol{\gamma}(k)) \end{aligned} \tag{26}$$

where $\Delta\hat{\mathbf{w}}(k) = \mathbf{w}(k) - \mathbf{w}_o(k+1)$. By premultiplying the expression above by $\mathbf{X}^T(k)$ it follows that

$$\begin{aligned} \mathbf{X}^T(k)\Delta\mathbf{w}(k+1) &= \mathbf{X}^T(k)\Delta\hat{\mathbf{w}}(k) \\ &+ P_{\text{up}}(k)\mathbf{X}^T(k)\mathbf{X}(k)\left(\mathbf{X}^T(k)\mathbf{X}(k) + \delta\mathbf{I}\right)^{-1}(\mathbf{e}(k) - \boldsymbol{\gamma}(k)) \end{aligned}$$

$$\begin{aligned} -\tilde{\boldsymbol{\varepsilon}}(k) &= -\tilde{\mathbf{e}}(k) \\ &+ P_{\text{up}}(k)\mathbf{X}^T(k)\mathbf{X}(k)\left(\mathbf{X}^T(k)\mathbf{X}(k) + \delta\mathbf{I}\right)^{-1}(\mathbf{e}(k) - \boldsymbol{\gamma}(k)) \end{aligned} \tag{27}$$

By solving the equation (27), it is possible to show that

$$\begin{aligned} &\left(\mathbf{X}^T(k)\mathbf{X}(k)\right)^{-1}[\tilde{\mathbf{e}}(k) - \tilde{\boldsymbol{\varepsilon}}(k)] = \\ &P_{\text{up}}(k)\left(\mathbf{X}^T(k)\mathbf{X}(k) + \delta\mathbf{I}\right)^{-1}(\mathbf{e}(k) - \boldsymbol{\gamma}(k)) \end{aligned} \tag{28}$$

Following the same procedure to derive equation (16), we can now substitute equation (28) in equation (26) in order to deduce that

$$\begin{aligned} &\Delta\mathbf{w}(k+1) - \mathbf{X}(k)\left(\mathbf{X}^T(k)\mathbf{X}(k)\right)^{-1}\tilde{\mathbf{e}}(k) = \Delta\hat{\mathbf{w}}(k) \\ &-\mathbf{X}(k)\left(\mathbf{X}^T(k)\mathbf{X}(k)\right)^{-1}\tilde{\boldsymbol{\varepsilon}}(k) \end{aligned} \tag{29}$$

By computing the energy on both sides of this equation it is possible to show that

$$\begin{aligned} &E\left[\|\Delta\mathbf{w}(k+1)\|^2\right] + E\left[\tilde{\mathbf{e}}^T(k)\left(\mathbf{X}^T(k)\mathbf{X}(k)\right)^{-1}\tilde{\mathbf{e}}(k)\right] \\ = &E\left[\|\Delta\hat{\mathbf{w}}(k)\|^2\right] + E\left[\tilde{\boldsymbol{\varepsilon}}^T(k)\left(\mathbf{X}^T(k)\mathbf{X}(k)\right)^{-1}\tilde{\boldsymbol{\varepsilon}}(k)\right] \\ = &E\left[\|\Delta\mathbf{w}(k) + \Delta\mathbf{w}_o(k+1)\|^2\right] \\ &+ E\left[\tilde{\boldsymbol{\varepsilon}}^T(k)\left(\mathbf{X}^T(k)\mathbf{X}(k)\right)^{-1}\tilde{\boldsymbol{\varepsilon}}(k)\right] \\ \approx &E\left[\|\Delta\mathbf{w}(k)\|^2\right] + E\left[\|\Delta\mathbf{w}_o(k+1)\|^2\right] \\ &+ E\left[\tilde{\boldsymbol{\varepsilon}}^T(k)\left(\mathbf{X}^T(k)\mathbf{X}(k)\right)^{-1}\tilde{\boldsymbol{\varepsilon}}(k)\right] \end{aligned} \tag{30}$$

where $\Delta\mathbf{w}_o(k+1) = \mathbf{w}_o(k) - \mathbf{w}_o(k+1)$, and in the last equality we have assumed that $E\left[\Delta\mathbf{w}^T(k)\Delta\mathbf{w}_o(k+1)\right] \approx 0$. This assumption is valid for simple models for the time-varying behavior of the unknown system, such as random walk model[1]. We will adopt this assumption in order to simplify our analysis.

In Appendix II we compute the covariance of $\Delta\mathbf{w}_o(k+1)$ leading to

$$\begin{aligned} E\left[\|\Delta\mathbf{w}_o(k+1)\|^2\right] &= \mathrm{tr}\{\mathrm{cov}[\Delta\mathbf{w}_o(k+1)]\} \\ &= (N+1)\left[\frac{2\kappa_{\mathbf{w}}^2}{1+\lambda_{\mathbf{w}}}\right]\sigma_{\mathbf{w}}^2 \end{aligned} \quad (31)$$

with $\kappa_{\mathbf{w}}$ defined in the Appendix II.

Solving equation (30) using equation (31) and assuming that the algorithm has converged such that

$$E\left[\|\Delta\mathbf{w}(k+1)\|^2\right] = E\left[\|\Delta\mathbf{w}(k)\|^2\right]$$

Equation (30) can be expressed as

$$\begin{aligned} &P_{\mathrm{up}}^2 E\left[\tilde{\mathbf{e}}^T(k)\left(\mathbf{X}^T(k)\mathbf{X}(k)\right)^{-1}\tilde{\mathbf{e}}(k)\right] \\ &= P_{\mathrm{up}} E\left[\tilde{\boldsymbol{\varepsilon}}^T(k)\left(\mathbf{X}^T(k)\mathbf{X}(k)\right)^{-1}\tilde{\boldsymbol{\varepsilon}}(k)\right] \\ &+(N+1)\left[\frac{2\kappa_{\mathbf{w}}^2}{1+\lambda_{\mathbf{w}}}\right]\sigma_{\mathbf{w}}^2 \end{aligned} \quad (32)$$

Leading to the equation

$$\begin{aligned} &P_{\mathrm{up}}^2 E\left[(\mathbf{e}(k)-\boldsymbol{\gamma}(k))^T\hat{\mathbf{S}}(k)\hat{\mathbf{R}}(k)\hat{\mathbf{S}}(k)(\mathbf{e}(k)-\boldsymbol{\gamma}(k))\right] \\ = \;& P_{\mathrm{up}} E\left[\tilde{\mathbf{e}}^T(k)\hat{\mathbf{S}}(k)(\mathbf{e}(k)-\boldsymbol{\gamma}(k)) + (\mathbf{e}(k)-\boldsymbol{\gamma}(k))^T\hat{\mathbf{S}}(k)\tilde{\mathbf{e}}(k)\right] \\ &+(N+1)\left[\frac{2\kappa_{\mathbf{w}}^2}{1+\lambda_{\mathbf{w}}}\right]\sigma_{\mathbf{w}}^2 \end{aligned} \quad (33)$$

By solving this equation following the same procedure as in [2], we can derive the excess of MSE only due to the time-varying unknown system.

$$\xi_{\mathrm{lag}} = \frac{(N+1)}{P_{\mathrm{up}}(2-P_{\mathrm{up}})}\left[\frac{2\kappa_{\mathbf{w}}^2}{1+\lambda_{\mathbf{w}}}\right]\sigma_{\mathbf{w}}^2 \quad (34)$$

By taking into consideration the additional noise and the time-varying parameters to be estimated, the overall excess of MSE is given by

$$\begin{aligned} \xi_{\mathrm{exc}} &= \frac{(L+1)P_{\mathrm{up}}}{2-P_{\mathrm{up}}}\frac{\sigma_n^2+\gamma^2}{1+L\left((1-P_{\mathrm{up}})^2+2P_{\mathrm{up}}(1-P_{\mathrm{up}})\sqrt{\frac{2}{\pi E[e_0^2(k)]}}\gamma\right)} \\ &+\frac{(N+1)}{P_{\mathrm{up}}(2-P_{\mathrm{up}})}\left[\frac{2\kappa_{\mathbf{w}}^2}{1+\lambda_{\mathbf{w}}}\right]\sigma_{\mathbf{w}}^2 \\ &= \frac{1}{2-P_{\mathrm{up}}}\left\{\frac{(L+1)P_{\mathrm{up}}(\sigma_n^2+\gamma^2)}{1+L\left((1-P_{\mathrm{up}})^2+2P_{\mathrm{up}}(1-P_{\mathrm{up}})\sqrt{\frac{2}{\pi E[e_0^2(k)]}}\gamma\right)}\right. \\ &\left.+\frac{N+1}{P_{\mathrm{up}}}\left[\frac{2\kappa_{\mathbf{w}}^2}{1+\lambda_{\mathbf{w}}}\right]\sigma_{\mathbf{w}}^2\right\} \end{aligned} \quad (35)$$

[1]In this model the coefficients change according to $\mathbf{w}_o(k) = \mathbf{w}_o(k-1) + \mathbf{n}_{\mathbf{w}}(k)$.

If $\kappa_{\mathbf{w}} = 1$, the expression above becomes simpler

$$\begin{aligned} \xi_{\mathrm{exc}} &= \frac{1}{2-P_{\mathrm{up}}}\left\{\frac{(L+1)P_{\mathrm{up}}(\sigma_n^2+\gamma^2)}{1+L\left((1-P_{\mathrm{up}})^2+2P_{\mathrm{up}}(1-P_{\mathrm{up}})\sqrt{\frac{2}{\pi E[e_0^2(k)]}}\gamma\right)}\right. \\ &\left.+\frac{2(N+1)}{P_{\mathrm{up}}(1+\lambda_{\mathbf{w}})}\sigma_{\mathbf{w}}^2\right\} \end{aligned} \quad (36)$$

As can be verified, the contribution due to the lag is inversely proportional to the value of P_{up}. This is an expected result since if the updates are not frequent the adaptive filter will not be able to track the variations in the unknown system.

4. SIMULATION EXAMPLES

An adaptive filtering algorithm is used to identify a system whose impulse response is given by [2]

$$[0.1\ 0.3\ 0\ -0.2\ -0.4\ -0.7\ -0.4\ -0.2]$$

using the SM-AP algorithm using $L=0$, $L=1$ and $L=2$. Table 1 lists the estimated and measured misadjustments for $L=0$, $L=1$, and $L=2$. The results were obtained for $\gamma=\sqrt{2.7\sigma_n^2}$ and $\gamma=\sqrt{5\sigma_n^2}$. The results reflect the average of three distinct experiments with different values of the input signal correlation matrix eigenvalue spread. The expected misadjustments are close to the measured ones despite the approximations in the derivation of the theoretical formula.

Table 1: Evaluation of the SM-AP Algorithm, $\gamma=\sqrt{2.7\sigma_n^2}$ and $\gamma=\sqrt{5\sigma_n^2}$

	Misadjustment					
γ	$L=0$		$L=1$		$L=2$	
	Exp.	Theory	Exp.	Theory	Exp.	Theory
$\sqrt{2.7\sigma_n^2}$	0.2591	0.3354	0.4137	0.4315	0.5305	0.5432
$\sqrt{5\sigma_n^2}$	0.1947	0.1934	0.2295	0.2292	0.3305	0.2738

In Fig. 3, it is shown the measured and theoretical values of the excess of MSE in a non-stationary environment, for the case where $\lambda_{\mathbf{w}} = 0.96$. Once again the measured and theoretical results obtained from equation (35) are as close as similar results usually found in the literature, demonstrating the validity the proposed analysis. It can be observed that the results are less accurate for larger values of γ due to a reduction in the number updatings, turning the tracking more difficult. The computational complexity of the SM-AP algorithm is similar to the original AP algorithm whenever an update is required. However, the SM-AP algorithm substantially reduces the misadjustment.

5. CONCLUDING REMARKS

This paper presented the analysis of the set-membership affine projection (SM-AP) algorithms in nonstationary environments. The closed form expressions, derived for the excess of MSE of the SM-AP algorithms in nonstationary environments, are tools for the proper setup of these algorithms in practical applications. Some simulation results were included verifying that the analytical results match well the experimental ones.

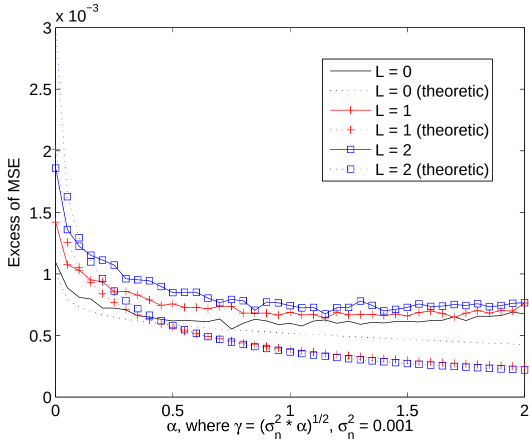

Figure 3: Excess of MSE in nonstationary environments of the SM-AP algorithms for $L = 0$, $L = 1$, and $L = 2$, eigenvalue spread equal 20.

Appendix I

The expression of the excess of MSE in the SM-AP algorithm must take into consideration how often the algorithm updates the coefficients after the transient. In [2] it is shown that probability that the adaptive filter is updated can be approximated by

$$\hat{P}_{\text{up}} \approx \max\left[2Q\left(\frac{\gamma}{\sqrt{(\sigma_n^2+\gamma^2)}}\right) + 2Q\left(\frac{\gamma}{\sqrt{5}}\right), 1\right] \quad (37)$$

where $Q[\cdot]$ is the complementary cumulative distribution function, and σ_n^2 is the additional noise variance.

Appendix II

The time-varying characteristic of the unknown system is a source of excess mean-square error. In order to take into account the excess MSE let's consider that each element of the actual coefficient vector is modeled as a first-order Markov process [17], since it leads to simple derivations. The first-order Markov process is described by

$$\mathbf{w}_o(k) = \lambda_{\mathbf{W}}\mathbf{w}_o(k-1) + \kappa_{\mathbf{W}}\mathbf{n}_{\mathbf{W}}(k) \quad (38)$$

where $\mathbf{n}_{\mathbf{W}}(k)$ is a vector whose elements are zero-mean white noise processes with variance $\sigma_{\mathbf{W}}^2$, and $\lambda_{\mathbf{W}} < 1$. The factor $\kappa_{\mathbf{W}} = (1-\lambda_{\mathbf{W}})^{\frac{p}{2}}$, for $p \geq 1$, is chosen such that $E[\mathbf{w}_o(k)\mathbf{w}_o^T(k)]$ is bounded.

The value of the excess of MSE requires the covariance of $\Delta\mathbf{w}_o(k+1) = \mathbf{w}_o(k) - \mathbf{w}_o(k+1)$, that is

$$\begin{aligned}
\text{cov}[\Delta\mathbf{w}_o(k+1)] &= E\left[(\mathbf{w}_o(k+1)-\mathbf{w}_o(k))(\mathbf{w}_o(k+1)-\mathbf{w}_o(k))^T\right] \\
&= E\left[(\lambda_{\mathbf{W}}\mathbf{w}_o(k)+\kappa_{\mathbf{W}}\mathbf{n}_{\mathbf{W}}(k)-\mathbf{w}_o(k))(\lambda_{\mathbf{W}}\mathbf{w}_o(k)+\kappa_{\mathbf{W}}\mathbf{n}_{\mathbf{W}}(k)-\mathbf{w}_o(k))^T\right] \\
&= E\left\{[(\lambda_{\mathbf{W}}-1)\mathbf{w}_o(k)+\kappa_{\mathbf{W}}\mathbf{n}_{\mathbf{W}}(k)][(\lambda_{\mathbf{W}}-1)\mathbf{w}_o(k)+\kappa_{\mathbf{W}}\mathbf{n}_{\mathbf{W}}(k)]^T\right\}
\end{aligned} \quad (39)$$

Since each element of $\mathbf{n}_{\mathbf{W}}(k)$ is a zero-mean white noise process with variance $\sigma_{\mathbf{W}}^2$, and $\lambda_{\mathbf{W}} < 1$, it follows that

$$\begin{aligned}
\text{cov}[\Delta\mathbf{w}_o(k+1)] &= \kappa_{\mathbf{W}}^2\sigma_{\mathbf{W}}^2\frac{(1-\lambda_{\mathbf{W}})^2}{1-\lambda_{\mathbf{W}}^2}\mathbf{I} + \kappa_{\mathbf{W}}^2\sigma_{\mathbf{W}}^2\mathbf{I} \\
&= \kappa_{\mathbf{W}}^2\left[\frac{1-\lambda_{\mathbf{W}}}{1+\lambda_{\mathbf{W}}}+1\right]\sigma_{\mathbf{W}}^2\mathbf{I}
\end{aligned} \quad (40)$$

Acknowledgments

The author would like to thank the financial support provided by CNPq and FAPERJ, and Markus V. S. Lima for careful reading the manuscript and for checking the simulations.

References

1. K. Ozeki and T. Umeda, "An adaptive filtering algorithm using an orthogonal projection to an affine subspace and its properties," *Electronics and Communications in Japan*, vol. 67-A, pp. 19-27, 1984.
2. P. S. R. Diniz, "Analysis of the set-membership affine projection algorithm," *Submitted for publication*, vol. , pp. -, Jan. 2009.
3. J. A. Apolinário, M. L. R. de Campos, and P. S. R. Diniz, "The binormalized data-reusing LMS algorithm," *IEEE Trans. on Signal Processing*, vol. 48, pp. 3235-3242, Nov. 2000.
4. S. G. Sankaran and A. A. (Louis) Beex, "Convergence behavior of affine projection algorithms," *IEEE Trans. on Signal Processing*, vol. 48, pp. 1086-1096, April 2000.
5. H.-C. Shin and A. H. Sayed, "Mean-square performance of a family of affine projection algorithms," *IEEE Trans. on Signal Processing*, vol. 52, pp. 90-102, Jan. 2004.
6. P. S. R. Diniz, *Adaptive Filtering: Algorithms and Practical Implementation*, Springer, New York, NY, 3rd ed., 2008.
7. E. Fogel and Y.-F. Huang, "On the value of information in system identification - bounded noise case," *Automatica*, vol. 18, pp. 229-238, March 1982.
8. S. Gollamudi, S. Nagaraj, S. Kapoor, and Y.-F. Huang, "Set-membership filtering and a set-membership normalized LMS algorithm with an adaptive step size," *IEEE Signal Processing Letters*, vol. 5, pp. 111-114, May 1998.
9. S. Werner and P. S. R. Diniz, "Set-membership affine projection algorithm," *IEEE Signal Processing Letters*, vol. 8, pp. 231-235, Aug 2001.
10. P. S. R. Diniz and S. Werner, "Set-membership binormalized data reusing LMS algorithms," *IEEE Trans. on Signal Processing*, vol. 51, pp. 124-134, Jan. 2003.
11. S. Werner, M. L. R. de Campos, and P. S. R. Diniz, "Partial-update NLMS algorithm with data-selective updating," *IEEE Trans. on Signal Processing*, vol. 52, pp. 938-949, April 2004.
12. P. S. R. Diniz, R. P. Braga, and S. Werner, "Set-membership affine projection algorithm for echo cancellation," *Proc. IEEE Intern. Symp. on Circuits and Systems*, Island of Kos, Greece, pp. 405-408, May 2006.
13. S. Werner, P. S. R. Diniz, and J. E. W. Moreira, "Set-membership affine projection algorithm with variable data-reuse factor," *Proc. IEEE Intern. Symp. on Circuits and Systems*, Island of Kos, Greece, pp. 261-264, May 2006.
14. J. F. Galdino, J. A. Apolinário, Jr., and M. L. R. de Campos, "A set-membership NLMS algorithm with time-varying error bound," *Proc. IEEE Intern. Symp. on Circuits and Systems*, Island of Kos, Greece, pp. 277-280, May 2006.
15. N. R. Yousef and A. H. Sayed, "A unified approach to the steady-state and tracking analyses of adaptive filters," *IEEE Trans. on Signal Processing*, vol. 49, pp. 314-324, Feb. 2001.
16. R. Price, "A useful theorem for nonlinear devises having Gaussian Inputs," *IEEE Trans. on Information Theory*, vol. IT-4, pp. 69-72, June 1958.
17. A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, McGraw Hill, New York, NY, 3rd ed., 1991.