STABILITY AND MSE ANALYSES OF AFFINE PROJECTION ALGORITHMS FOR SPARSE SYSTEM IDENTIFICATION

Markus V.S. Lima[†], *Iker Sobron*[†*], *Wallace A. Martins*[†], *Paulo S.R. Diniz*[†]

[†]Universidade Federal do Rio de Janeiro
COPPE & DEL-Poli/UFRJ
P.O. Box 68504, Rio de Janeiro, RJ, 21941-972, Brazil
[*]University of the Basque Country (UPV/EHU), Spain

ABSTRACT

We analyze two algorithms, viz. the affine projection algorithm for sparse system identification (APA-SSI) and the quasi APA-SSI (QAPA-SSI), regarding their stability and steady-state mean-squared error (MSE). These algorithms exploit the sparsity of the involved signals through an approximation of the l^0 norm. Such approach yields faster convergence and reduced steady-state MSE, as compared to algorithms that do not take the sparse nature of the signals into account. In addition, modeling sparsity via such approximation has been consistently verified to be superior to the widely used l^1 norm in several scenarios. In this paper, we show how to properly set the parameters of the two aforementioned algorithms in order to guarantee convergence, and we derive closed-form theoretical expressions for their steady-state MSE. A key conclusion from the proposed analysis is that the MSE of these two algorithms is a monotonically decreasing function of the sparsity degree. Simulation results are used to validate the theoretical findings.

***Index Terms*—** Affine projection, l^0 norm, sparsity, sparse system identification, adaptive filtering.

1. INTRODUCTION

Sparse signals and systems are reasonable models for a multitude of real-world problems, including medical imaging, wireless communications, radar processing, just to mention a few examples [1, 2, 3]. Many of those applications rely on the proper identification of the underlying sparse system in order to achieve their particular goals. In this context, adaptive algorithms play a major role since they yield online processing with reduced computational complexity as compared to batch-based offline algorithms. The family of affine projection (AP) adaptive algorithms [4, 5] is a case in point since it encompasses many online solutions commonly used in practice.

The first AP-based solutions that have addressed the problem of sparse system identification (SSI) are the so-called *proportionate* algorithms. Some examples of these algorithms, which do not employ data reuse, are the proportionate normalized least-mean-square (PNLMS) [6], the PNLMS++ [7], improved PNLMS (IPNLMS) [8], improved μ-law PNLMS (IMPNLMS) [9], among others [10] [11]. Regarding the algorithms that reuse previous data, we have the proportionate AP algorithm (PAPA) and the improved PAPA (IPAPA) [12]. As compared to standard AP algorithms, those proportionate solutions achieve higher convergence speed and lower steady-state MSE by taking into account the sparsity of the system.

An alternative approach to exploit sparsity is the use of regularization based on a sparsity-promoting norm, such as the l^0 and l^1 norms.[1] From the optimization viewpoint, this approach adds a penalty function to the original objective function and a stochastic gradient algorithm is derived, such as the zero-attracting AP algorithm (ZA-APA) and reweighted ZA-APA (RZA-APA) [18], whose penalty functions are related to the l^1 norm of the parameter vector. In [19], we proposed two new algorithms, namely the *affine projection algorithm for sparse system identification* (APA-SSI) and the *quasi APA-SSI* (QAPA-SSI), whose penalty functions are approximations of the l^0 norm of the parameter vector. As shown in [19], APA-SSI and QAPA-SSI outperform the aforementioned competing methods in many sparse scenarios and are able to exploit sparsity even in scenarios with low sparsity degree in which the l^1 norm could not improve the performance of ZA-APA and RZA-APA in comparison to that of the AP algorithm (APA).

After presenting the proposals in [19], some important theoretical issues have been raised: (i) *convexity* of the resulting objective function; (ii) *stability* of the proposed algorithms; and (iii) MSE *analysis* of the algorithms. The first point, related to convexity, has been addressed in [17] and is directly connected with the smoothness of the function that approximates the l^0 norm of the parameter vector. Addressing the second and third points is the main goal of this paper. Indeed, we show how to set the free parameters of APA-SSI and QAPA-SSI in order to yield a stable algorithm in the sense of not drifting away from the actual (but unknown) sparse system impulse response. In addition, we also present theoretical expressions for the excess MSE (EMSE) of APA-SSI and QAPA-SSI which allow us to draw important conclusions, such as the theoretical guarantees that those algorithms perform better than the standard AP algorithm in sparse scenarios.

Notation: For a given iteration k, the adaptive filter coefficient vector and input vector are denoted by $\mathbf{w}(k), \mathbf{x}(k) \in \mathbb{R}^{N+1}$, respectively. The desired signal, output signal, and error signal are respectively denoted by $d(k), y(k), e(k) \in \mathbb{R}$ and the following relations hold: $y \triangleq \mathbf{x}^T(k)\mathbf{w}(k)$ and $e(k) \triangleq d(k) - y(k)$. For AP algorithms, $L \in \mathbb{Z}_+$ previously used data are also employed together with the current data. For these algorithms we define the input matrix $\mathbf{X}(k) \triangleq [\mathbf{x}(k)\, \mathbf{x}(k-1)\ \ldots\ \mathbf{x}(k-L)] \in \mathbb{R}^{(N+1)\times(L+1)}$, the de-

This work has been financially supported in part by the FAPERJ, CNPq, and CAPES, and in part by the University of the Basque Country UPV/EHU (UFI 11/30), by the Department of Education, Universities and Research of the Basque Government, by the Basque Government (IT-683-13 and SAIOTEK), and by the Spanish Ministry of Economy and Competitiveness under the project HEDYT-GBB (TEC2012-33302).

[1] In addition, set estimation [13, 14] and compressive sensing theories have been combined leading to a different class of algorithms [15, 16, 17].

sired vector $\mathbf{d}(k) \triangleq [d(k)\, d(k-1)\, \ldots\, d(k-L)]^T \in \mathbb{R}^{L+1}$, and the error vector $\mathbf{e}(k) \triangleq \mathbf{d}(k) - \mathbf{X}^T(k)\mathbf{w}(k)$. In addition, $\|\cdot\|_2 = \|\cdot\|$ is the Euclidean norm and $\|\cdot\|_0$ is the l^0 norm.

2. THE ALGORITHMS

In this section, we present the two algorithms proposed in [19], viz. the APA-SSI and the QAPA-SSI. Both of them were devised for identification of sparse systems and have an important feature: they promote sparsity via function $F_\beta(\mathbf{w}(k))$, which is an approximation of the l^0 norm of $\mathbf{w}(k)$, where $\beta \in \mathbb{R}_+$ controls the trade-off between smoothness (low values of β) and quality of approximation (high values of β) [17].

The optimization problem related to APA-SSI is:

$$\begin{aligned} &\text{minimize } \|\mathbf{w}(k+1) - \mathbf{w}(k)\|_2^2 + \alpha F_\beta(\mathbf{w}(k+1)) \\ &\text{subject to } \mathbf{d}(k) - \mathbf{X}^T(k)\mathbf{w}(k+1) = \mathbf{0}, \end{aligned} \tag{1}$$

where $\alpha \in \mathbb{R}_+$ is a parameter that determines the weight given to the penalty function $F_\beta(\mathbf{w}(k+1))$.

Thus, the APA-SSI is characterized by the following recursion:

$$\begin{aligned} \mathbf{w}(k+1) =& \mathbf{w}(k) + \mu\mathbf{X}(k)\mathbf{S}(k)\mathbf{e}(k) \\ &+ \mu\frac{\alpha}{2}\left[\mathbf{X}(k)\mathbf{S}(k)\mathbf{X}^T(k) - \mathbf{I}\right]\mathbf{f}_\beta\left(\mathbf{w}(k)\right), \end{aligned} \tag{2}$$

in which $\mathbf{S}(k) \triangleq \left(\mathbf{X}^T(k)\mathbf{X}(k) + \delta\mathbf{I}\right)^{-1}$, $0 < \delta \ll 1$ is the regularization factor, μ is the step size, and $\mathbf{f}_\beta(\mathbf{w}(k)) \triangleq \nabla F_\beta(\mathbf{w}(k))$. On the other hand, the updating rule of the QAPA-SSI is given by

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \mu\mathbf{X}(k)\mathbf{S}(k)\mathbf{e}(k) - \mu\frac{\alpha}{2}\mathbf{f}_\beta\left(\mathbf{w}(k)\right), \tag{3}$$

which encompasses the l_0-NLMS algorithm proposed in [20] and can be regarded as an approximation of the APA-SSI recursion where the term $\mathbf{X}(k)\mathbf{S}(k)\mathbf{X}^T(k)$ is simply neglected. For more details about the APA-SSI and QAPA-SSI, see [19].

Since this paper is mainly about theoretical analyses, the choice of $F_\beta(\mathbf{w}(k))$ does not matter for further discussions and any function satisfying $F_\beta(\mathbf{w}(k)) \overset{\beta\to\infty}{\longrightarrow} \|\mathbf{w}(k)\|_0$ is valid. Examples of F_β can be found in [19, 17].

3. ANALYSES

In this section, stability and MSE analyses of the APA-SSI and QAPA-SSI are performed in a unified manner by considering the following updating rule:

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \mu\mathbf{X}(k)\mathbf{S}(k)\mathbf{e}(k) - \mu\frac{\alpha}{2}\mathbf{P}\mathbf{f}_\beta\left(\mathbf{w}(k)\right), \tag{4}$$

in which the APA-SSI and QAPA-SSI are obtained by making $\mathbf{P} = \mathbf{P}_{\mathbf{X}}^{\perp} \triangleq \mathbf{I} - \mathbf{X}(k)\mathbf{S}(k)\mathbf{X}^T(k)$ and $\mathbf{P} = \mathbf{I}$, respectively.

3.1. Stability

Let us start by considering that the unknown sparse system can be modeled as a linear time-invariant system with impulse response $\mathbf{w}_* \in \mathbb{R}^{N+1}$. The aim of this subsection is to characterize the values of μ and α which ensure that $\mathbf{w}(k+1)$ is closer to $\mathbf{w}_*$ than $\mathbf{w}(k)$ is. Thus, by defining

$$\Delta\mathbf{w}(k) \triangleq \mathbf{w}_* - \mathbf{w}(k) \tag{5}$$

and the function $f : \mathbb{R} \times \mathbb{R} \longrightarrow \mathbb{R}$ as

$$f(\alpha, \mu) \triangleq \|\Delta\mathbf{w}(k+1)\|^2 - \|\Delta\mathbf{w}(k)\|^2, \tag{6}$$

our aim is to determine the values of α and μ that make $f(\alpha, \mu) < 0$.

By expanding $\|\Delta\mathbf{w}(k+1)\|^2$ using (5) and (4), we obtain:

$$\begin{aligned} f(\alpha,\mu) =& \underbrace{\|\mathbf{P}\mathbf{f}_\beta(\mathbf{w}(k))\|^2\,\frac{\mu^2}{4}}_{\triangleq a(k)}\alpha^2 \\ &+ \underbrace{\Big(\mu\Delta\mathbf{w}^T(k)\mathbf{P}\mathbf{f}_\beta(\mathbf{w}(k)) - \mu^2\mathbf{e}^T(k)\mathbf{S}(k)\mathbf{X}^T(k)\mathbf{P}\mathbf{f}_\beta(\mathbf{w}(k))\Big)}_{\triangleq b(k)}\alpha \\ &+ \underbrace{\Big(\mu^2\|\mathbf{X}(k)\mathbf{S}(k)\mathbf{e}(k)\|^2 - 2\mu\Delta\mathbf{w}^T(k)\mathbf{X}(k)\mathbf{S}(k)\mathbf{e}(k)\Big)}_{\triangleq c(k)}, \end{aligned} \tag{7}$$

Since the algorithms should work properly even in the case when sparsity is not exploited, i.e., when $\alpha = 0$, then the condition $f(\alpha, \mu) < 0$ becomes $c(k) < 0$, where

$$\begin{aligned} c(k) &= \mu^2\|\mathbf{X}(k)\mathbf{S}(k)\mathbf{e}(k)\|^2 - 2\mu\Delta\mathbf{w}^T(k)\mathbf{X}(k)\mathbf{S}(k)\mathbf{e}(k) \\ &= \mu^2\mathbf{e}^T(k)\mathbf{S}(k)\mathbf{e}(k) - 2\mu\Delta\mathbf{w}^T(k)\mathbf{X}(k)\mathbf{S}(k)\mathbf{e}(k) \\ &= \mathbf{e}^T(k)\mathbf{S}(k)\mathbf{e}(k)\left[\mu^2 - 2\mu\right] \\ &= \underbrace{\mathbf{e}^T(k)\mathbf{S}(k)\mathbf{e}(k)}_{>0}\left[\mu(\mu-2)\right]. \end{aligned} \tag{8}$$

In the second equality we used the symmetry of $\mathbf{S}(k)$ and considered that $\mathbf{S}(k) = \left(\mathbf{X}^T(k)\mathbf{X}(k)\right)^{-1}$, i.e., we considered $\delta = 0$ since it was artificially introduced to prevent numerical issues due to matrix inversion. In the third equality we considered a noiseless scenario: in such kind of scenario $\mathbf{d}(k) = \mathbf{X}^T(k)\mathbf{w}_*$ (without additive measurement noise) and $\mathbf{y}(k) = \mathbf{X}^T(k)\mathbf{w}(k)$ so that $\mathbf{e}(k) = \mathbf{d}(k) - \mathbf{y}(k) = \mathbf{X}^T(k)\Delta\mathbf{w}(k)$. In the last equality we used the fact that $\mathbf{S}(k)$ is positive definite so that $\mathbf{e}^T(k)\mathbf{S}(k)\mathbf{e}(k) > 0$. Therefore, in order to have $c(k) < 0$ we must use $0 < \mu < 2$.

Now, considering that μ is appropriately set so that $c(k) < 0$, we may think of (7) as a function of α only. For a fixed μ, $f(\alpha, \mu)$ is a second-order polynomial on α whose discriminant $d(k)$ is

$$d(k) = \underbrace{b^2(k)}_{\geq 0} - 4\underbrace{a(k)}_{\geq 0}\underbrace{c(k)}_{<0} \geq 0, \tag{9}$$

where the above inequality follows from the definitions of $a(k)$ and $b(k)$, and the choice of μ that guarantees $c(k) < 0$. Therefore, there exist two roots, viz. $\alpha_{\min}(k)$ and $\alpha_{\max}(k)$, and α should be chosen as $\alpha_{\min}(k) \leq \alpha \leq \alpha_{\max}(k)$.[2] In addition, notice that $\alpha = 0$ is always a valid solution because $f(0, \mu) = c(k) < 0$ for $0 < \mu < 2$.

Up to this point, we have addressed the choice of α and μ from the theoretical viewpoint. In practice, when noise is present the range of μ should be reduced to $0 < \mu < 1$ so that noise enhancement is mitigated and condition $f(\alpha, \mu) < 0$ is met. In addition, since we are assuming that the unknown system $\mathbf{w}_*$ is sparse, then it makes no sense to use negative values of α, thus $0 \leq \alpha < \alpha_{\max}(k)$, for every k. Therefore, α should be a nonnegative real number with small amplitude (usually, $\alpha \ll 1$).

[2] In practice, $a(k) > 0$ due to floating point arithmetic, implying $\alpha_{\min}(k) \neq \alpha_{\max}(k)$. Indeed, after convergence none of the entries of $\mathbf{w}(k)$ will be exactly equal to 0 and, therefore, none of the entries of $\mathbf{f}_\beta(\mathbf{w}(k))$ is 0. Thus, for the two definitions of $\mathbf{P}$ that interest us we have $\mathbf{P}\mathbf{f}_\beta(\mathbf{w}(k)) \neq \mathbf{0}$. In addition, when $\mathbf{P}\mathbf{f}_\beta(\mathbf{w}(k)) = \mathbf{0}$, we have $a(k) = b(k) = 0$, implying that any value of α can be used.

3.2. Steady-state MSE

In this subsection, we derive an expression for the steady-state MSE of the two involved algorithms using the energy-conservation method [21]. Due to lack of space, several mathematical derivations are skipped so that we can focus on the resulting expressions, their interpretations and relationships with the analysis presented in [22].

Let us start by subtracting $\mathbf{w}_*$ from both sides of (4) so that

$$\Delta\mathbf{w}(k+1) = \Delta\mathbf{w}(k) - \mu\mathbf{X}(k)\mathbf{S}(k)\mathbf{e}(k) + \mu\frac{\alpha}{2}\mathbf{P}\mathbf{f}_\beta\left(\mathbf{w}(k)\right), \tag{10}$$

which, after the premultiplication by $\mathbf{X}^T(k)$, gives us

$$\tilde{\boldsymbol{\epsilon}}(k) = \tilde{\mathbf{e}}(k) - \mu\mathbf{R}(k)\mathbf{S}(k)\mathbf{e}(k) + \mu\frac{\alpha}{2}\mathbf{X}^T(k)\mathbf{P}\mathbf{f}_\beta\left(\mathbf{w}(k)\right), \tag{11}$$

where $\mathbf{R}(k) \triangleq \mathbf{X}^T(k)\mathbf{X}(k)$ is assumed to be full rank, and

$$\tilde{\boldsymbol{\epsilon}}(k) = \mathbf{X}^T(k)\Delta\mathbf{w}(k+1) = \boldsymbol{\epsilon}(k) - \mathbf{n}(k), \tag{12}$$
$$\tilde{\mathbf{e}}(k) = \mathbf{X}^T(k)\Delta\mathbf{w}(k) = \mathbf{e}(k) - \mathbf{n}(k) \tag{13}$$

are the noiseless a posteriori error vector and the noiseless a priori error vector, respectively. Thus, after premultiplying (11) by $\mathbf{R}^{-1}(k)$, we end up with

$$\mu\mathbf{S}(k)\mathbf{e}(k) = \mathbf{R}^{-1}(k)\left[\tilde{\mathbf{e}}(k) - \tilde{\boldsymbol{\epsilon}}(k)\right] + \mu\frac{\alpha}{2}\mathbf{R}^{-1}(k)\mathbf{X}^T(k)\mathbf{P}\mathbf{f}_\beta\left(\mathbf{w}(k)\right), \tag{14}$$

which can be employed in (10), thus allowing one to obtain (after some lengthy manipulations) the expression

$$\Delta\mathbf{w}(k+1) + \mathbf{X}(k)\mathbf{R}^{-1}(k)\tilde{\mathbf{e}}(k) = \Delta\mathbf{w}(k) + \mathbf{X}(k)\mathbf{R}^{-1}(k)\tilde{\boldsymbol{\epsilon}}(k) + \mu\frac{\alpha}{2}\left[\mathbf{I} - \mathbf{X}(k)\mathbf{R}^{-1}(k)\mathbf{X}^T(k)\right]\mathbf{P}\mathbf{f}_\beta\left(\mathbf{w}(k)\right). \tag{15}$$

Therefore, by evaluating the energies at both sides of (15), one can prove the following relation:

$$\|\Delta\mathbf{w}(k+1)\|^2 + \tilde{\mathbf{e}}^T(k)\mathbf{R}^{-1}(k)\tilde{\mathbf{e}}(k) = \|\Delta\mathbf{w}(k)\|^2 + \tilde{\boldsymbol{\epsilon}}^T(k)\mathbf{R}^{-1}(k)\tilde{\boldsymbol{\epsilon}}(k) + (\mu\alpha)\Delta\mathbf{w}^T(k)\bar{\mathbf{P}}\mathbf{P}\mathbf{f}_\beta\left(\mathbf{w}(k)\right) + \frac{(\mu\alpha)^2}{4}\mathbf{f}_\beta^T\left(\mathbf{w}(k)\right)\mathbf{P}^T\bar{\mathbf{P}}\mathbf{P}\mathbf{f}_\beta\left(\mathbf{w}(k)\right), \tag{16}$$

where $\bar{\mathbf{P}} \triangleq \mathbf{I} - \mathbf{X}(k)\mathbf{R}^{-1}(k)\mathbf{X}^T(k)$. It is worth noticing that, for the two kinds of matrix $\mathbf{P}$ corresponding to the APA-SSI and QAPA-SSI, the following relations hold: (i) $\bar{\mathbf{P}}\mathbf{P} = \bar{\mathbf{P}}$ and (ii) $\mathbf{P}^T\bar{\mathbf{P}}\mathbf{P} = \bar{\mathbf{P}}$. Thus, applying such relations, the expected value operator, and assuming that the algorithms have reached steady-state so that the relation $\mathrm{E}\left[\|\Delta\mathbf{w}(k+1)\|^2\right] = \mathrm{E}\left[\|\Delta\mathbf{w}(k)\|^2\right]$ is valid, then (16) becomes

$$\mathrm{E}\left[\tilde{\mathbf{e}}^T(k)\mathbf{R}^{-1}(k)\tilde{\mathbf{e}}(k)\right] = \mathrm{E}\left[\tilde{\boldsymbol{\epsilon}}^T(k)\mathbf{R}^{-1}(k)\tilde{\boldsymbol{\epsilon}}(k)\right] + (\mu\alpha)\mathrm{E}\left[\Delta\mathbf{w}^T(k)\bar{\mathbf{P}}\mathbf{f}_\beta\left(\mathbf{w}(k)\right)\right] + \frac{(\mu\alpha)^2}{4}\mathrm{E}\left[\mathbf{f}_\beta^T\left(\mathbf{w}(k)\right)\bar{\mathbf{P}}\mathbf{f}_\beta\left(\mathbf{w}(k)\right)\right]. \tag{17}$$

After many manipulations, we obtain for the APA-SSI

$$(2-\mu)\mathrm{E}\left[\tilde{\mathbf{e}}^T(k)\mathbf{R}^{-1}(k)\tilde{\mathbf{e}}(k)\right] = \mu\mathrm{E}\left[\mathbf{n}^T(k)\mathbf{S}(k)\mathbf{n}(k)\right] + \alpha\mathrm{E}\left[\Delta\mathbf{w}^T(k)\bar{\mathbf{P}}\mathbf{f}_\beta\left(\mathbf{w}(k)\right)\right], \tag{18}$$

whereas for the QAPA-SSI we have

$$(2-\mu)\mathrm{E}\left[\tilde{\mathbf{e}}^T(k)\mathbf{R}^{-1}(k)\tilde{\mathbf{e}}(k)\right] = \mu\mathrm{E}\left[\mathbf{n}^T(k)\mathbf{S}(k)\mathbf{n}(k)\right] + \alpha\mathrm{E}\left[\Delta\mathbf{w}^T(k)\mathbf{f}_\beta\left(\mathbf{w}(k)\right)\right]. \tag{19}$$

In order to continue the derivation of the EMSE, we shall assume that, in the steady-state, $\mathbf{X}(k)$ is independent of $\tilde{\mathbf{e}}(k)$ and $\mathrm{E}\left[\tilde{\mathbf{e}}(k)\tilde{\mathbf{e}}^T(k)\right] = \mathrm{E}\left[|\tilde{e}_0(k)|^2\right]\mathbf{S}_1$, where $\tilde{e}_0(k)$ is the first entry of $\tilde{\mathbf{e}}(k)$, $\mathbf{S}_1 = \mathrm{diag}\{\left[1,\ (1-\mu)^2,\ \ldots,\ (1-\mu)^{2L}\right]\}$, and we assumed that $\alpha < \mu$, so that the terms containing factors of $\alpha\mu$ and μ^2 were neglected. Those assumptions are inspired by a similar procedure used in [22]. Now we can approximate the expected values of (18) and (19) as follows:

$$\begin{aligned}(2-\mu)\mathrm{E}\left[\tilde{\mathbf{e}}^T(k)\mathbf{R}^{-1}(k)\tilde{\mathbf{e}}(k)\right] &= \\ (2-\mu)\mathrm{tr}\{\mathrm{E}\left[\tilde{\mathbf{e}}(k)\tilde{\mathbf{e}}^T(k)\mathbf{R}^{-1}(k)\right]\} &= \\ (2-\mu)\mathrm{E}\left[|\tilde{e}_0(k)|^2\right]\mathrm{tr}\{\mathbf{S}_1\mathrm{E}\left[\mathbf{R}^{-1}(k)\right]\}&.\end{aligned} \tag{20}$$

In the same way, denoting the noise variance by σ^2, we can rewrite the first term of the right-hand side of (18) or (19) as

$$\mu\mathrm{E}\left[\mathbf{n}^T(k)\mathbf{S}(k)\mathbf{n}(k)\right] = \mu\sigma^2\mathrm{tr}\{\mathrm{E}\left[\mathbf{S}\right]\}. \tag{21}$$

Replacing (20) and (21) in (18) and (19), we can obtain the EMSE value, $\mathrm{E}\left[|\tilde{e}_0(k)|^2\right]$, for the APA-SSI algorithm as

$$\mathrm{E}\left[|\tilde{e}_0(k)|^2\right] = \frac{\mu\sigma^2\mathrm{tr}\{\mathrm{E}\left[\mathbf{S}\right]\} + \alpha\mathrm{E}\left[\Delta\mathbf{w}^T(k)\bar{\mathbf{P}}\mathbf{f}_\beta\left(\mathbf{w}(k)\right)\right]}{(2-\mu)\mathrm{tr}\{\mathbf{S}_1\mathrm{E}\left[\mathbf{R}^{-1}(k)\right]\}}, \tag{22}$$

and for the QAPA-SSI as

$$\mathrm{E}\left[|\tilde{e}_0(k)|^2\right] = \frac{\mu\sigma^2\mathrm{tr}\{\mathrm{E}\left[\mathbf{S}\right]\} + \alpha\mathrm{E}\left[\Delta\mathbf{w}^T(k)\mathbf{f}_\beta\left(\mathbf{w}(k)\right)\right]}{(2-\mu)\mathrm{tr}\{\mathbf{S}_1\mathrm{E}\left[\mathbf{R}^{-1}(k)\right]\}}. \tag{23}$$

Now, if one compares (22) and (23) with the results described in [22], one can conclude that:

$$\text{EMSE} = \text{‘EMSE of AP’} + \text{‘Sparsity Modelling Term’}, \tag{24}$$

where the first term ‘EMSE of AP’, given by $\mu\sigma^2\mathrm{tr}\{\mathrm{E}\left[\mathbf{S}\right]\}/(2-\mu)\mathrm{tr}\{\mathbf{S}_1\mathrm{E}\left[\mathbf{R}^{-1}(k)\right]\}$, is the same expression that was derived in [22]. As for the ‘Sparsity Modelling Term’, it models the way the particular algorithm takes sparsity into account.

In order to analyze the role of the ‘Sparsity Modelling Term’, let us first consider the QAPA-SSI. The argument of the expected value of $\alpha\mathrm{E}\left[\Delta\mathbf{w}^T(k)\mathbf{f}_\beta\left(\mathbf{w}(k)\right)\right]$ (see (23)) can be written as

$$\Delta\mathbf{w}^T(k)\mathbf{f}_\beta\left(\mathbf{w}(k)\right) = \sum_{n=0}^{N}\Delta w_n(k)f_\beta\left(w_n(k)\right), \tag{25}$$

where $\Delta w_n(k) = w_{*,n} - w_n(k)$ and $f_\beta(w_n(k)) \triangleq \frac{\partial F_\beta(\mathbf{w}(k))}{\partial w_n(k)}$ are the nth components of $\Delta\mathbf{w}(k)$ and $\mathbf{f}_\beta(\mathbf{w}(k))$, respectively. When dealing with sparse scenarios, two possibilities can occur: (i) if $w_{*,n} \neq 0$, then assuming β is properly set (i.e., $F_\beta(\cdot)$ is a good approximation of $\|\cdot\|_0$) [17], then there is no correction to be applied to the standard AP recursion, so that we have $f_\beta(w_n(k)) \approx 0$ after convergence since $w_n(k)$ is likely to be nonzero as well. In other words, our sparsity model does not affect these coefficients

Table 1. Steady-state MSE (dB): experimental *vs.* theoretical.

$\mathcal{M}_s$ (%)	APA-SSI with $L=0$				QAPA-SSI with $L=0$				APA-SSI with $L=2$				QAPA-SSI with $L=2$			
	$\mu=0.05$		$\mu=0.1$		$\mu=0.05$		$\mu=0.1$		$\mu=0.05$		$\mu=0.1$		$\mu=0.05$		$\mu=0.1$	
	Exp.	Theor.	Exp.	Theor.	Exp.	Theor.	Exp.	Theor.	Exp.	Theor.	Exp.	Theor.	Exp.	Theor.	Exp.	Theor.
0	-19.85	-19.88	-19.75	-19.77	-19.85	-19.88	-19.75	-19.77	-19.61	-19.88	-19.25	-19.73	-19.61	-19.88	-19.25	-19.73
20	-19.90	-19.91	-19.79	-19.82	-19.90	-19.91	-19.79	-19.82	-19.67	-19.92	-19.32	-19.80	-19.68	-19.92	-19.33	-19.81
40	-19.92	-19.93	-19.84	-19.87	-19.92	-19.93	-19.84	-19.87	-19.71	-19.96	-19.42	-19.88	-19.72	-19.97	-19.44	-19.89
60	-19.94	-19.96	-19.90	-19.92	-19.94	-19.96	-19.90	-19.92	-19.77	-20.00	-19.50	-19.95	-19.79	-20.02	-19.53	-19.98
80	-19.97	-19.98	-19.94	-19.97	-19.97	-19.99	-19.94	-19.97	-19.82	-20.05	-19.58	-20.02	-19.84	-20.07	-19.62	-20.06

and the component $\Delta w_n(k) f_\beta(w_n(k)) \approx 0$; (ii) if $w_{*,n}=0$, then $\Delta w_n(k) = -w_n(k) = -\text{sign}(w_n(k))|w_n(k)|$. In addition, as shown in [19], $f_\beta(w_n(k))$ has the general form $\text{sign}(w_n(k))\gamma$, where γ is a positive constant. Therefore, $\Delta w_n(k)$ and $f_\beta(w_n(k))$ have opposite signs, thus leading to $\Delta w_n(k) f_\beta(w_n(k)) < 0$. As a result, the sum in (25) contains only (approximately) zero or negative terms, thus implying that $\Delta \mathbf{w}^T(k)\mathbf{f}_\beta(\mathbf{w}(k)) \leq 0$. Note that the equality is achieved for dispersive impulse responses in which $\|\mathbf{w}_*\|_0 = N+1$, i.e., no null component, which means that the QAPA-SSI achieves the same steady-state MSE of the standard AP algorithm for non-sparse environments. Furthermore, as $\|\mathbf{w}_*\|_0$ decreases toward 0, more negative components of the form $\Delta w_n(k) f_\beta(w_n(k))$ appear in (25), which means that the steady-state MSE of the QAPA-SSI decreases with the sparsity degree.

The same conclusions seem to be valid for the APA-SSI, as the simulation results indicate. However, it requires a much more involving proof. The numerator of the 'Sparsity Modelling Term' of the APA-SSI can be rewritten as a sum of two terms: $\alpha \text{E}\left[\Delta \mathbf{w}^T(k)\mathbf{f}_\beta(\mathbf{w}(k))\right] + \alpha \text{E}\left[\Delta \mathbf{w}^T(k)\mathbf{X}(k)\mathbf{R}^{-1}(k)\mathbf{X}^T(k)\mathbf{f}_\beta(\mathbf{w}(k))\right]$, where the first term is the same as that explained in (25). Since the rank of $\mathbf{X}(k)\mathbf{R}^{-1}(k)\mathbf{X}^T(k)$ is $L+1$ at most, then only $L+1$ components out of $N+1$ entries of vectors $\Delta\mathbf{w}(k)$ and $\mathbf{f}_\beta(\mathbf{w}(k))$ are considered in the second term. Hence, intuitively, the first term seems to be dominant.

Based on the previous interpretations for the EMSE expressions, it is very interesting to observe that what matters to the MSE is the *number of coefficients equal to 0*, and not how they are distributed along the vector $\mathbf{w}_*$. In other words, what matters is the sparsity degree. Indeed, in [17] we show simulation results that confirm this observation.

Finally, we would like to mention that the expressions in (22) and (23) are only for theoretical purposes because one cannot compute them in practice. Indeed, such expressions depend implicitly on $\mathbf{w}_*$, which is not known, and on the expected values related to the other variables, which might be unknown beforehand.

4. SIMULATION RESULTS

In this section, we provide simulation results aiming at validating the EMSE expressions in (22) and (23) and also to illustrate the steady-state MSE performance of the APA-SSI and QAPA-SSI as the degree of sparsity varies. Therefore, we now introduce our measure of the sparsity degree:

$$\mathcal{M}_s \triangleq \frac{(N+1) - \|\mathbf{w}_*\|_0}{(N+1)} \times 100, \tag{26}$$

i.e., $\mathcal{M}_s \in [0, 100)$ represents the percentage of coefficients that are equal to 0.

The experiments consist in identifying the impulse response of an unknown system $\mathbf{w}_*$, which is modelled as an FIR filter of order 15. In the 0th experiment we have all 16 taps of $\mathbf{w}_*$ equal to 1, i.e., $\mathcal{M}_s = 0$. In the nth experiment, $n \in \{1, 2, \ldots, 15\}$, we turn n coefficients of $\mathbf{w}_*$ to 0, whereas $16-n$ coefficients remain equal to 1. The input signal is drawn from a zero-mean Gaussian distribution with variance equal to 1. The additive measurement noise is uncorrelated with the input signal and is assumed to be white and Gaussian with variance $\sigma^2 = 10^{-2}$.

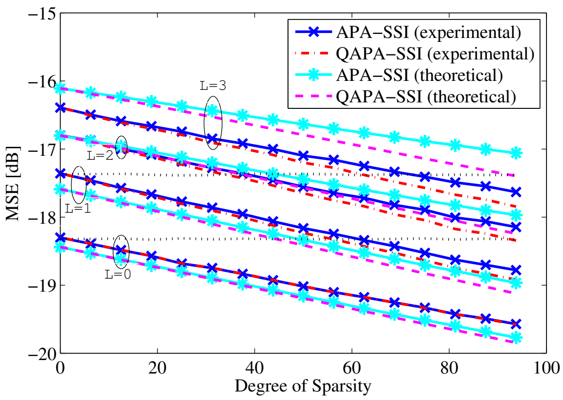

Fig. 1. Steady-state MSE *vs.* $\mathcal{M}_s$ with $\mu = 0.6$. For the sake of clarity, the APA is represented in dotted line only for $L \in \{0, 1\}$.

As for the adaptive filter, its order is also $N = 15$, it is initialized as $\mathbf{w}(0) = \mathbf{0}$, and the algorithm parameters are set as: $\alpha = 10^{-3}$, $\beta = 5$, $\delta = 10^{-12}$, and $\mu \in \{0.05, 0.1, 0.6\}$.

The results shown in Fig. 1 and Table 1 validate the theoretical steady-state MSE expressions as well as they corroborate our previous observations/interpretations. Indeed, it is clear from such results that the APA-SSI and QAPA-SSI can reduce their steady-state MSEs as the degree of sparsity increases, for a fixed μ. Alternatively, for a desired steady-state MSE, these algorithms can use a higher value of μ, thus increasing convergence speed. In addition, as it happens with the APA, the accuracy of the theoretical expressions decreases as L increases [22].

It is worth noticing that for $\mathcal{M}_s = 0\%$, the APA-SSI and QAPA-SSI exhibit similar performance, as compared to the APA, but even when $\mathcal{M}_s$ is low, these algorithms can still benefit from this low sparsity degree, as these results indicate. This is a very important feature. For instance, when $\mathcal{M}_s \approx 50\%$, AP-based algorithms using the l^1 norm to model sparsity did not exhibit any improvement as compared to the AP algorithm, as shown in [19].

We also observed that, the MSE curves of APA-SSI and QAPA-SSI decrease slower with $\mathcal{M}_s$ when L increases toward N. Such behavior is theoretically explained by the fact that the term 'EMSE of AP' increases with L, thus becoming dominant.

5. CONCLUSION

In this paper, we analyzed two important features of the APA-SSI and QAPA-SSI: stability and steady-state MSE. Indeed, we explained how to set their parameters μ and α in order to guarantee convergence, and we derived theoretical expressions for their EMSE. Such expressions are composed of two terms: the EMSE of the AP algorithm and a term that takes the sparsity degree into account.

6. REFERENCES

[1] M. Elad, *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*, Springer, New York, USA, 2010.

[2] Y. Eldar and G. Kutyniok, *Compressed Sensing: Theory and Applications*, Cambridge University Press, Cambridge, UK, 2012.

[3] S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing*, Springer, New York, USA, 2013.

[4] 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, no. 5, pp. 19–27, 1984.

[5] P.S.R. Diniz, *Adaptive Filtering: Algorithms and Practical Implementation*, Springer, New York, USA, 4th edition, 2013.

[6] D.L. Duttweiler, "Proportionate normalized least-mean-squares adaptation in echo cancelers," *IEEE Transactions on Speech and Audio Processing*, vol. 8, no. 5, pp. 508–518, September 2000.

[7] S.L. Gay, "An efficient, fast converging adaptive filter for network echo cancellation," in *Thirty-Second Asilomar Conference on Signals, Systems amp, Computers (ACSSC 1998)*, 1998, vol. 1, pp. 394–398.

[8] J. Benesty and S.L. Gay, "An improved PNLMS algorithm," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2002)*, Dallas, USA, May 2002, vol. 2, pp. 1881–1884.

[9] L. Liu, M. Fukumoto, and S. Saiki, "An improved mu-law proportionate NLMS algorithm," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2008)*, 2008, pp. 3797–3800.

[10] P.A. Naylor, J. Cui, and M. Brookes, "Adaptive algorithms for sparse echo cancellation," *Signal Processing*, vol. 86, no. 6, pp. 1182–1192, June 2006.

[11] A.W.H. Khong and P.A. Naylor, "Efficient use of sparse adaptive filters," in *Fortieth Asilomar Conference on Signals, Systems and Computers (ACSSC 2006)*, 2006, pp. 1375–1379.

[12] O. Hoshuyama, R.A. Goubran, and A. Sugiyama, "A generalized proportionate variable step-size algorithm for fast changing acoustic environments," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2004)*, 2004, vol. 4, pp. 161–164.

[13] P.L. Combettes, "The foundations of set theoretic estimation," *Proceedings of the IEEE*, vol. 81, no. 2, pp. 182–208, February 1993.

[14] M.V.S. Lima and P.S.R. Diniz, "Steady-state MSE performance of the set-membership affine projection algorithm," *Circuits, Systems, and Signal Processing*, vol. 32, no. 4, pp. 1811–1837, January 2013.

[15] Y. Kopsinis, K. Slavakis, and S. Theodoridis, "Online sparse system identification and signal reconstruction using projections onto weighted l_1 balls," *IEEE Transactions on Signal Processing*, vol. 59, no. 3, pp. 936–952, March 2011.

[16] K. Slavakis, Y. Kopsinis, and S. Theodoridis, "Adaptive algorithm for sparse system identification using projections onto weighted l_1 balls," in *International Conference on Acoustics Speech and Signal Processing (ICASSP 2010)*, Dallas, USA, March 2010, vol. 17, pp. 3742–3745.

[17] M.V.S. Lima, W.A. Martins, T.N. Ferreira, and P.S.R. Diniz, "Sparsity-aware data-selective adaptive filters," *IEEE Transactions on Signal Processing*, 2013 (submitted).

[18] R. Meng, R.C. de Lamare, and V.H. Nascimento, "Sparsity-aware affine projection adaptive algorithms for system identification," in *Sensor Signal Processing for Defence (SSPD 2011)*, London, U.K., September 2011, pp. 1–5.

[19] M.V.S. Lima, W.A. Martins, and P.S.R. Diniz, "Affine projection algorithms for sparse system identification," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2013)*, Vancouver, Canada, May 2013, pp. 5666–5670.

[20] Y. Gu, J. Jin, and S. Mei, "l_0 norm constraint LMS algorithm for sparse system identification," *IEEE Signal Processing Letters*, vol. 16, no. 9, pp. 774–777, September 2009.

[21] N.R. Yousef and A.H. Sayed, "A unified approach to the steady-state and tracking analyses of adaptive filters," *IEEE Transactions on Signal Processing*, vol. 49, no. 2, pp. 314–324, February 2001.

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