AFFINE PROJECTION ALGORITHMS FOR SPARSE SYSTEM IDENTIFICATION

Markus V.S. Lima, Wallace A. Martins, Paulo S.R. Diniz

Federal University of Rio de Janeiro
COPPE & DEL-Poli/UFRJ
P.O. Box 68504, Rio de Janeiro, RJ, 21941-972, Brazil

ABSTRACT

We propose two versions of affine projection (AP) algorithms tailored for sparse system identification (SSI). Contrary to most adaptive filtering algorithms devised for SSI, which are based on the l^1 norm, the proposed algorithms rely on homotopic l^0 norm minimization, which has proven to yield better results in some practical contexts. The first proposal is obtained by direct minimization of the AP cost function with a penalty function based on the l^0 norm of the coefficient vector, whereas the second algorithm is a simplified version of the first proposal. Simulation results are presented in order to evaluate the performance of the proposed algorithms considering three different homotopies to the l^0 norm as well as competing algorithms.

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

1. INTRODUCTION

System identification (SI) has been a widely studied topic due to its importance in applied sciences, especially in Engineering [1]. Many practical applications involve unknown systems that can be properly modeled as having finite memory, justifying the fact that finite-duration impulse response (FIR) filters are more used than infinite-duration impulse response (IIR) filters.

In the SI context one frequently faces two problems: (i) lack of a priori information/specifications and (ii) time-varying characteristics of the unknown system. Adaptive filtering algorithms have become a popular tool to cope with those problems in SI. In particular, the least mean square (LMS) and recursive least squares (RLS) algorithms [2] are the most widely known. Indeed, the LMS is quite used due to its computational simplicity, whereas the RLS provides faster convergence but its application might be sometimes hindered due to its relatively high computational complexity. The affine projection (AP) algorithm, on the other hand, is well-known for representing a tradeoff between complexity and convergence speed [2, 3, 4].

Although the aforementioned adaptive filtering algorithms have already proven their values, their related convergence speeds degrade as the impulse response[1] of the unknown system becomes longer [5, 6]. Fortunately, in several practical cases (e.g., in echo cancellation), long impulse responses are also sparse/compressible, i.e., most of their components have values close to zero. The sparsity inherent to such systems can be exploited in order to increase the convergence speed of adaptive filtering algorithms, as shown in [5, 6, 7, 8].

Sparsity is directly revealed by the l^0 norm,[2] but minimization of such a norm is a very difficult task (NP-hard problem). Borrowing some results from the compressed sensing literature, in particular the equivalence between l^0 and l^1 minimizations for sufficiently high dimensional problems, some algorithms using (directly or indirectly) the l^1 norm have been derived, such as the ones proposed in [6, 7, 8, 10, 11, 12, 13, 14, 15, 16]. Indeed, l^1 minimization leads to a convex problem, which is much more tractable. However, some papers [17, 18] have shown that the results achieved via l^0 norm minimization outperform the ones obtained using the l^1 norm in practical conditions. Performance differences occur since some practical applications do not meet exactly the constraints required to characterize the equivalence between l^0 and l^1 norm minimizations. Roughly speaking, those constraints impose that the proportion of nonzero coefficients must be very small as compared to the proportion of zero/small-value coefficients in the associated parameter vector [18].

In this paper, we tackle the sparse SI (SSI) problem via homotopic l^0 norm minimization, as presented in an image processing context [17]. The same idea has been previously used to derive the l^0 norm-constraint LMS algorithm [19], but considering just one type of l^0 norm homotopy, which is based on the Laplace function. This paper, on the other hand, presents two versions of the AP algorithm employing three different homotopies to the l^0 norm. In addition, we describe the connection between one of the proposed algorithms and the algorithm proposed in [19], and then we show simulation results aimed at evaluating the utilized homotopies as well as comparing their performances against some competing algorithms [6].

This paper is organized as follows. Section 2 provides a brief review of the AP algorithm while also establishes the notation. Section 3 presents the two proposed l^0-norm-based algorithms, explaining how homotopies are used to enable the approximation of the gradient of an l^0-norm function. The competing algorithms are discussed in Section 4. In Section 5, the performance of the proposed algorithms is evaluated in different setups: (i) varying the l^0-norm homotopies and (ii) varying the degree of sparsity of the unknown system. The conclusions are drawn in Section 6.

2. AP ALGORITHM

Before describing the AP algorithm, let us define some important variables in the context of adaptive filtering. Let $\mathbf{x}(k), \mathbf{w}(k) \in \mathbb{R}^{N+1}$ be the input vector (sometimes called regressor) and the coefficient vector for a given iteration $k \in \mathbb{Z}$, respectively. Let the

Prof. Martins is also with the graduate program PPEEL/CEFET-RJ.

Thanks to CNPq, CAPES, and FAPERJ for funding this work.

[1]The term "*the* impulse response", which is often found in the adaptive-filtering literature, is loose since the system could be time-varying. The reader should keep in mind that the idea here is that the unknown system has long memory.

[2]The l^0 norm of a vector is the number of coefficients within this vector that are different from zero. In fact, it is not a true norm [9].

reference (or desired) signal be denoted by $d(k) \in \mathbb{R}$, the output of the filter be defined as $y(k) \triangleq \mathbf{x}^T(k)\mathbf{w}(k)$, and the error signal be given as $e(k) \triangleq d(k) - y(k)$.

The AP algorithm, originally proposed in [3], reuses previous input vectors to increase convergence speed. Assuming we have available the last $L+1 \in \mathbb{N}$ data pairs of input vectors and desired signals, 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 desired vector $\mathbf{d}(k) \triangleq [d(k)\,d(k-1)\,\ldots\,d(k-L)]^T \in \mathbb{R}^{L+1}$, and the error vector $\mathbf{e}(k) \triangleq \mathbf{d}(k) - \mathbf{X}^T(k)\mathbf{w}(k)$.

Geometrically, the updating process of the AP algorithm yields a coefficient vector $\mathbf{w}(k+1)$ which is as close as possible to $\mathbf{w}(k)$ (minimum perturbation criterion in the Euclidean distance sense) and whose associated a posteriori error is equal to zero, i.e., $\mathbf{d}(k) - \mathbf{X}^T(k)\mathbf{w}(k+1) = \mathbf{0}$. Mathematically, the optimization problem solved by the AP algorithm is

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

which leads to the updating equation of the AP algorithm:

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \mathbf{X}(k)\left(\mathbf{X}^T(k)\mathbf{X}(k)\right)^{-1}\mathbf{e}(k). \tag{2}$$

In practice, a regularization factor $\mathbb{R}_+ \ni \delta \ll 1$ is usually employed to avoid numerical instability due to matrix inversion. In addition, a convergence factor (step-size) $\mathbb{R}_+ \ni \mu < 1$ is also used to relax the constraint of the optimization problem, thus enabling smaller steps in the updating process. Therefore, the AP algorithm is commonly implemented as [2]

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \mu\mathbf{X}(k)\left(\mathbf{X}^T(k)\mathbf{X}(k) + \delta\mathbf{I}\right)^{-1}\mathbf{e}(k). \tag{3}$$

3. PROPOSED AP-BASED ALGORITHMS

In this section we propose two algorithms that take into account the l^0 norm of the coefficient vector: the *AP algorithm for sparse system identification* (AP-SSI) and the *quasi AP-SSI* (QAP-SSI). The AP-SSI is obtained by adding a penalty function based on the l^0 norm of $\mathbf{w}(k+1)$ to the standard AP cost function, whereas the aim of the QAP-SSI is to reduce the computational complexity of the AP-SSI and also to enable a more general updating process, as it will be further explained.

3.1. AP-SSI

In this subsection, we derive the AP-SSI algorithm. As previously mentioned, AP-SSI adds a penalty function based on the l^0 norm of $\mathbf{w}(k+1)$ to the AP optimization problem in order to promote sparsity at each iteration. Thus, the AP-SSI optimization problem is given by

$$\begin{aligned}&\text{minimize } \|\mathbf{w}(k+1) - \mathbf{w}(k)\|_2^2 + \alpha\|\mathbf{w}(k+1)\|_0 \\ &\text{subject to } \mathbf{d}(k) - \mathbf{X}^T(k)\mathbf{w}(k+1) = \mathbf{0},\end{aligned} \tag{4}$$

where $\alpha \in \mathbb{R}_+$ is a nonnegative parameter that determines the weight given to the l^0 norm penalty.

In order to solve this optimization problem, we form the Lagrangean $\mathcal{L}$ as

$$\begin{aligned}\mathcal{L} =& \|\mathbf{w}(k+1) - \mathbf{w}(k)\|_2^2 + \alpha\|\mathbf{w}(k+1)\|_0 \\ &+ \boldsymbol{\lambda}^T(k)\left[\mathbf{d}(k) - \mathbf{X}^T(k)\mathbf{w}(k+1)\right],\end{aligned} \tag{5}$$

differentiate it with respect to $\mathbf{w}(k+1)$ and $\boldsymbol{\lambda}(k)$, and equal the resulting expressions to zero (i.e., $\nabla\mathcal{L} = \mathbf{0}$), thus yielding

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \mathbf{X}(k)\frac{\boldsymbol{\lambda}(k)}{2} - \frac{\alpha}{2}\nabla\|\mathbf{w}(k+1)\|_0, \tag{6}$$

$$\mathbf{d}(k) = \mathbf{X}^T(k)\mathbf{w}(k+1), \tag{7}$$

respectively. Then, the left-multiplication of eq. (6) by $\mathbf{X}^T(k)$ and the substitution of eq. (7) into the resulting equation generates (assuming that $\mathbf{X}^T(k)\mathbf{X}(k)$ is invertible)

$$\begin{aligned}\frac{\boldsymbol{\lambda}(k)}{2} =& \left(\mathbf{X}^T(k)\mathbf{X}(k)\right)^{-1}\mathbf{e}(k) \\ &+ \frac{\alpha}{2}\left(\mathbf{X}^T(k)\mathbf{X}(k)\right)^{-1}\mathbf{X}^T(k)\nabla\|\mathbf{w}(k+1)\|_0.\end{aligned} \tag{8}$$

Substituting eq. (8) into eq. (6) leads to the updating equation of AP-SSI

$$\begin{aligned}\mathbf{w}(k+1) =& \mathbf{w}(k) + \mathbf{X}(k)\left(\mathbf{X}^T(k)\mathbf{X}(k)\right)^{-1}\mathbf{e}(k) \\ &+ \frac{\alpha}{2}\left[\mathbf{X}(k)\left(\mathbf{X}^T(k)\mathbf{X}(k)\right)^{-1}\mathbf{X}^T(k) - \mathbf{I}\right] \\ &\times \nabla\|\mathbf{w}(k+1)\|_0.\end{aligned} \tag{9}$$

3.2. QAP-SSI

Eq. (9) takes into account the projection of $\nabla\|\mathbf{w}(k+1)\|_0$ onto the subspace which is orthogonal to the space spanned by $\mathbf{X}(k)$. Such a property is important to guarantee that the a posteriori error vector is indeed zero. In effect, left-multiplication of eq. (9) by $\mathbf{X}^T(k)$ cancels the term which depends on $\nabla\|\mathbf{w}(k+1)\|_0$, so that $\mathbf{X}^T(k)\mathbf{w}(k+1) = \mathbf{d}(k)$, as expected. If one does not impose such a greedy constraint by allowing a nonzero a posteriori error vector, then a more flexible updating process could be achieved. A simple way of implementing this constraint relaxation, while taking advantage of the proposal derived in Subsection 3.1, is to directly use the vector $\nabla\|\mathbf{w}(k+1)\|_0$ rather than its associated projection. Thus, the proposed QAP-SSI updating equation is

$$\begin{aligned}\mathbf{w}(k+1) =& \mathbf{w}(k) + \mathbf{X}(k)\left(\mathbf{X}^T(k)\mathbf{X}(k)\right)^{-1}\mathbf{e}(k) \\ &- \frac{\alpha}{2}\nabla\|\mathbf{w}(k+1)\|_0\,.\end{aligned} \tag{10}$$

It is worth mentioning that, in comparison with the AP-SSI, the QAP-SSI is less complex and achieves a more flexible updating process since there is no projection applied to the term $\nabla\|\mathbf{w}(k+1)\|_0$. Indeed, observe that left-multiplication of eq. (10) by $\mathbf{X}^T(k)$ leads to $\mathbf{X}^T(k)\mathbf{w}(k+1) = \mathbf{d}(k) - \frac{\alpha}{2}\mathbf{X}^T(k)\nabla\|\mathbf{w}(k+1)\|_0$, thus implying that the a posterior error vector is no longer zero. In fact, QAP-SSI is not a true AP algorithm, i.e., it does not generate $\mathbf{w}(k+1)$ as a projection of $\mathbf{w}(k)$ onto some space.

Note that the current forms of eqs. (9) and (10) cannot be implemented in practice, since the term $\nabla\|\mathbf{w}(k+1)\|_0$ has two issues: (i) requires a proper definition for the derivative of the l^0 norm and (ii) depends on $\mathbf{w}(k+1)$. Some tricks to overcome such undesirable features as well as practical updating equations are presented in Subsections 3.3 and 3.4.

3.3. Homotopic l^0 norm

It remains to adapt the proposed updating equations of Subsections 3.1 and 3.2 so that they can be employed in practical adaptive filters, which have an inherent online nature. Indeed, eqs. (9) and (10) depend on the derivative of the l^0 norm of $\mathbf{w}(k+1)$ since they are related to the minimization of the l^0 norm of the coefficient vector (sparsity constraint), thus leading to an NP-hard problem (in plain English, it is very difficult to work with such a norm, especially in online applications). Therefore, when online algorithms are required, it is common practice to approximate the l^0 norm by some continuous and differentiable function, allowing one to use gradient-based methods straightforwardly. In this subsection, we show typical approximations for the l^0 norm of an arbitrary vector $\mathbf{z} \in \mathbb{R}^{N+1}$ and compute their derivatives with respect to $\mathbf{z} \triangleq [z_0 \; z_1 \; \dots \; z_N]^T$.

The following approximations for the l^0 norm are valid:

$$\|\mathbf{z}\|_0 \approx \sum_{i=0}^{N} \left(1 - \mathrm{e}^{-\beta|z_i|}\right), \tag{11}$$

$$\|\mathbf{z}\|_0 \approx \sum_{i=0}^{N} \left(1 - \mathrm{e}^{-\frac{1}{2}\beta^2 z_i^2}\right), \tag{12}$$

$$\|\mathbf{z}\|_0 \approx \sum_{i=0}^{N} \left(1 - \frac{1}{\beta|z_i|+1}\right), \tag{13}$$

where eqs. (11) and (13) are based on the well-known Laplace [20] and Geman-McClure [21] functions, respectively, whereas eq. (12) is also used in [22]. Experimental results presented in [22], whose aim is to compare different approximations for the l^0 norm, show that the Geman-McClure function leads to a fast implementation and accurate reconstruction in the compressed sensing (CS) context, considering a fixed-point CS processor.

The idea here is that, when the nonnegative parameter $\beta \in \mathbb{R}_+$ which appears in eqs. (11), (12), and (13) is sufficiently large, all of these approximations essentially count the number of nonzero values in the vector $\mathbf{z}$. Indeed, for all approximations, one adds $N+1$ terms of the form $1 - F_\beta(z_i)$, where $F_\beta(z_i)$ can be $\mathrm{e}^{-\beta|z_i|}$ or $\mathrm{e}^{-\frac{1}{2}\beta^2 z_i^2}$ or $\frac{1}{\beta|z_i|+1}$. It is straightforward to see that, when $z_i = 0$, one has $F_\beta(z_i) = 1$, for whichever choice of $F_\beta(z_i)$. On the other hand, if $z_i \neq 0$, then $|z_i| > 0$ so that one has $F_\beta(z_i) \to 0$ as long as $\beta \to \infty$, for whichever choice of $F_\beta(z_i)$. These facts eventually mean that, when $z_i = 0$, the term $1 - F_\beta(z_i)$ in the above sums will be zero, whereas when $z_i \neq 0$, that term will be approximately 1 for large β. The term homotopic comes from the fact that those approximations of the l^0 norm are continuous functions of the parameter β and can be continuously distorted to yield the actual l^0 norm of $\mathbf{z}$.

The corresponding derivatives $f_\beta(z_i) \triangleq \frac{\mathrm{d}\|\mathbf{z}\|_0}{\mathrm{d}z_i}$ of the above approximations are

$$f_\beta(z_i) \approx \beta \mathrm{sign}(z_i) \mathrm{e}^{-\beta|z_i|}, \tag{14}$$

$$f_\beta(z_i) \approx \beta^2 z_i \mathrm{e}^{-\frac{1}{2}\beta^2 z_i^2}, \tag{15}$$

$$f_\beta(z_i) \approx \frac{\beta \mathrm{sign}(z_i)}{(\beta|z_i|+1)^2}. \tag{16}$$

Observe that the functions $f_\beta(z_i)$ in eqs. (14) and (16) have a discontinuity at $z_i = 0$, whereas eq. (15) is continuous everywhere.

Thus, we can define the gradient of $\|\mathbf{z}\|_0$ with respect to $\mathbf{z}$ as

$$\nabla\|\mathbf{z}\|_0 = \mathbf{f}_\beta(\mathbf{z}) \triangleq [f_\beta(z_0) \; f_\beta(z_1) \; \dots \; f_\beta(z_N)]^T. \tag{17}$$

3.4. Implemented updating equations

Now that we have available good approximations for the gradient of the l^0 norm, we can rewrite eqs. (9) and (10) in a way that is amenable to online implementations.

The implemented updating equation for the AP-SSI is

$$\begin{aligned}\mathbf{w}(k+1) =& \mathbf{w}(k) + \mu \mathbf{X}(k)\mathbf{S}(k)\mathbf{e}(k) \\ &+ \mu\frac{\alpha}{2}\left[\mathbf{X}(k)\mathbf{S}(k)\mathbf{X}^T(k) - \mathbf{I}\right]\mathbf{f}_\beta(\mathbf{w}(k)),\end{aligned} \tag{18}$$

in which $\mathbf{S}(k) \triangleq \left(\mathbf{X}^T(k)\mathbf{X}(k) + \delta\mathbf{I}\right)^{-1}$, δ is a regularization factor, and μ is the step-size factor. Note that, since we do not know a priori the exact value of $\|\mathbf{w}(k+1)\|_0$ at the kth iteration, and since we are also minimizing $\|\mathbf{w}(k+1) - \mathbf{w}(k)\|_2^2$, then it is reasonable to assume that $\|\mathbf{w}(k+1)\|_0 \approx \|\mathbf{w}(k)\|_0$, so that $\nabla\|\mathbf{w}(k+1)\|_0 \approx \nabla\|\mathbf{w}(k)\|_0 = \mathbf{f}_\beta(\mathbf{w}(k)) = [f_\beta(w_0(k)) \; \dots \; f_\beta(w_N(k))]^T$, as given in eq. (17). In this case, $f_\beta(w_i(k))$ can be computed using any approximation described in Subsection 3.3.

Similarly, the implemented update for the QAP-SSI is

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \mu\mathbf{X}(k)\mathbf{S}(k)\mathbf{e}(k) - \mu\frac{\alpha}{2}\mathbf{f}_\beta(\mathbf{w}(k)). \tag{19}$$

It is worth pointing out that this proposed QAP-SSI updating equation encompasses the l_0-NLMS algorithm in [19]. Indeed, the l_0-NLMS algorithm can be achieved by setting the QAP-SSI in the following way: (i) $L = 0$, (ii) $f_\beta(w_0(k))$ as a first-order approximation via Taylor series of eq. (14).

4. COMPETING ALGORITHMS

Aiming at describing how our contribution is related to prior work in the field, this section briefly describes the two algorithms proposed in [6], viz. the zero-attracting affine projection algorithm (ZA-APA) and the reweighted ZA-APA (RZA-APA), The ZA-APA algorithm was derived by directly minimization of the AP cost function plus a penalty function based on the l^1 norm of $\mathbf{w}(k)$. In order to highlight the similarities between AP-SSI and ZA-APA algorithms, we write the updating equation of the ZA-APA algorithm as[3]

$$\begin{aligned}\mathbf{w}(k+1) =& \mathbf{w}(k) + \mu\mathbf{X}(k)\mathbf{S}(k)\mathbf{e}(k) \\ &+ \mu\frac{\alpha}{2}\left[\mathbf{X}(k)\mathbf{S}(k)\mathbf{X}^T(k) - \mathbf{I}\right]\mathbf{sign}(\mathbf{w}(k)),\end{aligned} \tag{20}$$

where $\mathbf{sign}(\mathbf{w}(k))$ is the element-wise sign function.

Similarly, the RZA-APA updating equation can be written as

$$\begin{aligned}\mathbf{w}(k+1) =& \mathbf{w}(k) + \mu\mathbf{X}(k)\mathbf{S}(k)\mathbf{e}(k) \\ &+ \mu\frac{\alpha}{2}\left[\mathbf{X}(k)\mathbf{S}(k)\mathbf{X}^T(k) - \mathbf{I}\right]\mathbf{P}(\mathbf{w}(k)),\end{aligned} \tag{21}$$

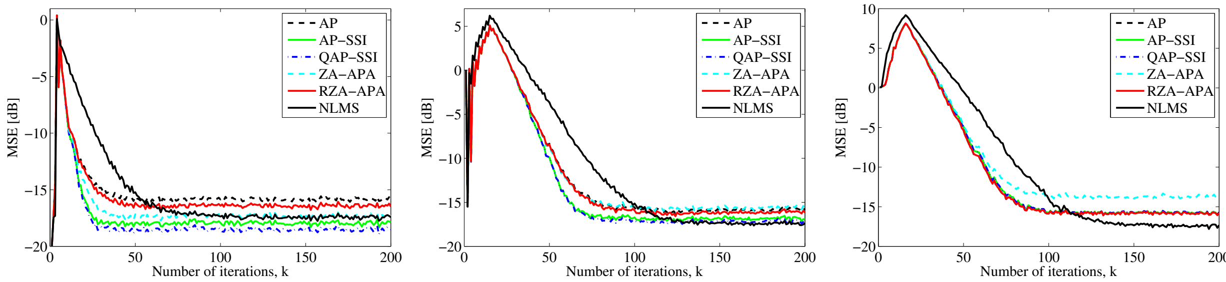
where $\mathbf{P}(\mathbf{w}(k)) \in \mathbb{R}^{N+1}$ is defined as

$$\mathbf{P}(\mathbf{w}(k)) \triangleq \mathbf{sign}(\mathbf{w}(k)) \div (\mathbf{1} + \epsilon|\mathbf{w}(k)|), \tag{22}$$

where the symbol $\div$ stands for the element-wise division of the first vector by the second vector, $\mathbf{1}$ is a vector whose $N+1$ entries are equal to 1, $|\mathbf{w}(k)|$ is the element-wise absolute value, and $\epsilon \in \mathbb{R}$ is called the shrinkage magnitude [6].

One can easily see that both ZA-APA and RZA-APA algorithms differ from the proposed AP-SSI algorithm only in the definition of the vector that multiplies the term $\mu\frac{\alpha}{2}\left[\mathbf{X}(k)\mathbf{S}(k)\mathbf{X}^T(k) - \mathbf{I}\right]$. Our proposal employs an homotopic approximation for the derivative of the l^0 norm of $\mathbf{w}(k+1)$. This key difference is enough to yield better results in the context of sparse systems, as the simulation results of Section 5 indicate.

[3] The only difference between eq. (20) and the ZA-APA of [6] is a regularization factor δ which appears in the definition of $\mathbf{S}(k)$ after eq. (18). In addition, we incorporated the step-size factor μ in every term that is added to $\mathbf{w}(k)$, which essentially implies that our α is a scaled version of the α in [6].



(a) Exp. 1. (b) Exp. 2. (c) Exp. 3.

Fig. 2. Degree of sparsity.

5. SIMULATION RESULTS

We present the simulation scenarios in Subsection 5.1 and then, in Subsection 5.2, we evaluate the AP-SSI and QAP-SSI considering all the homotopies presented in Subsection 3.3. In Subsection 5.3, using the homotopy that leads to the best results, we compare the proposed algorithms *versus* the ones described in Section 4.

5.1. Scenarios

The simulation scenarios we consider are the same three experiments proposed in [6]. Those scenarios allow us to assess the performance of the proposed algorithms for different degrees of sparsity. The experiments consist of identifying an unknown system composed of 16 coefficients, whose taps are set as follows: (i) Exp. 1: 4th tap equal to 1, others equal to 0; (ii) Exp. 2: odd taps equal to 1, even taps equal to 0; and (iii) Exp. 3: all taps equal to 1.

Regarding the adaptive filter parameters, the number of coefficients is 16 and the following algorithms are tested: the proposed ones (AP-SSI and QAP-SSI), the proposals of [6] (ZA-APA and RZA-APA), and the classical ones (AP and NLMS) to serve as benchmarks for comparisons. The algorithms were set so that they have a similar convergence speed.[4] Thus, we use step-size $\mu = 0.9$, regularization factor $\delta = 10^{-12}$, reuse factor $L = 4$, $\beta = 5$ (following the suggestion of [19]), and, in accordance with the suggested values in [6], we use $\alpha = 5 \times 10^{-3}$ and $\epsilon = 100$. In addition, the reference signal $d(k)$ is assumed to be corrupted by an additive white Gaussian measurement noise with variance $\sigma_n^2 = 0.01$.

5.2. Performance evaluation: homotopy

The AP-SSI and QAP-SSI were tested in several scenarios using the l^0 norm homotopies given in eqs. (11) to (13). Throughout all scenarios we tested, the following observations always hold: (i) convergence speed was similar for all homotopies; (ii) eq. (12) leads to the worst results in terms of steady-state mean-square error (MSE); and (iii) approximations based on eqs. (11) and (13) exhibit almost identical performances. Fig. 1 depicts an example considering Exp. 1.

In the results shown in Subsection 5.3, the proposed algorithms employ the Geman-McClure homotopy given in eq. (13) since the implementation cost of eq. (16) is lower than the one of eq. (14).

[4]This observation is valid for the AP-based algorithms. By using the same step-size μ of the AP algorithm, the NLMS algorithm will be slower.

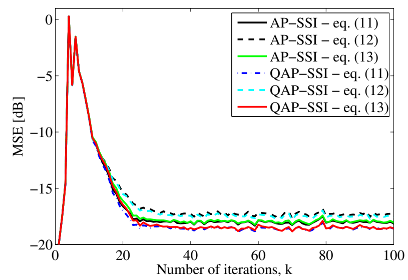


Fig. 1. Comparing the homotopies for the two proposed algorithms considering the scenario of Exp. 1 described in Subsection 5.3.

5.3. Performance evaluation: degree of sparsity

Fig. 2 depicts the MSE results for Exps. 1, 2, and 3. It can be observed that the convergence speeds are similar for all AP-based methods. Fig. 2(a) shows that by exploiting the sparsity of the underlying unknown system, all four algorithms (AP-SSI, QAP-SSI, ZA-APA, and RZA-APA) outperform the AP algorithm. Actually, the proposed AP-SSI and QAP-SSI algorithms achieved the best results. As depicted in Figs. 2(b) and (c), as the unknown system becomes less sparse, the performances of the algorithms which explicitly take sparsity into account in their formulations become worse and converge to the performance of the AP algorithm when there is no sparsity (see Fig. 2(c)). However, one may note that when the sparsity factor is 50% (Exp. 2), the performance of the methods based on l^1 norm is not very different from that of the AP algorithm, whereas the proposed methods lead to a better performance.

6. CONCLUSION

This paper proposed two novel data-reusing adaptive filtering algorithms which allow efficient identification of sparse systems. The key feature of the proposed algorithms, namely the AP-SSI and QAP-SSI, is that they consider the sparsity feature by using l^0 norm, whose capability of promoting sparsity in some practical contexts without impairing performance is proven to be superior to other norms, such as the l^1 norm. The inherent difficulties of working directly with the l^0 norm are circumvented by the use of homotopic approximations of the derivative of the l^0 norm of the filter coefficient vector. Simulation results indicate that the proposed algorithms outperform other competing algorithms, especially when the degree of sparsity is relatively high, and suggest the use of the Geman-McClure homotopy.

7. REFERENCES

[1] T. Hinamoto and S. Maekawa, "Extended theory of learning identification," *Electrical Engineering in Japan*, vol. 95, no. 5, pp. 101–107, October 1975.

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

[3] 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.

[4] S.L. Gay and S. Tavathia, "The fast affine projection algorithm," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP-95)*, Detroit, MI, USA, May 1995, vol. 5, pp. 3023–3026.

[5] M.R. Petraglia and D.B. Haddad, "New adaptive algorithms for identification of sparse impulse responses – analysis and comparisons," in *7th International Symposium on Wireless Communication Systems (ISWCS 2010)*, York, UK, September 2010, pp. 384–388.

[6] 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.

[7] 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.

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

[9] Y. Eldar and G. Kutyniok, *Compressed Sensing: Theory and Applications*, Cambridge University Press, Cambridge, UK, 1st edition, 2012.

[10] S. Werner, J.A. Apolinário Jr., and P.S.R. Diniz, "Set-membership proportionate affine projection algorithms," *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2007, no. 1, pp. 034242, 2007.

[11] S. Werner, J.A. Apolinário Jr., P.S.R. Diniz, and T.I. Laakso, "A set-membership approach to normalized proportionate adaptation algorithms," in *European Signal Processing Conference (EUSIPCO 2005)*, Antalya, Turkey, September 2005, pp. 1–4.

[12] P. Di Lorenzo, S. Barbarossa, and A.H. Sayed, "Sparse diffusion LMS for distributed adaptive estimation," in *International Conference on Acoustics, Speech and Signal Processing (ICASSP 2012)*, March 2012, pp. 3281–3284.

[13] Y. Kopsinis, K. Slavakis, S. Theodoridis, and S. McLaughlin, "Reduced complexity online sparse signal reconstruction using projections onto weighted l_1 balls," in *International Conference on Digital Signal Processing (DSP)*, July 2011, pp. 1–8.

[14] 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, TX, USA, March 2010, vol. 17, pp. 3742–3745.

[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] S. Chouvardas, K. Slavakis, Y. Kopsinis, and S. Theodoridis, "A sparsity promoting adaptive algorithm for distributed learning," *IEEE Transactions on Signal Processing*, vol. 60, no. 10, pp. 5412–5425, October 2012.

[17] J. Trzasko and A. Manduca, "Highly undersampled magnetic resonance image reconstruction via homotopic l_0-minimization," *IEEE Transactions on Medical Imaging*, vol. 28, no. 1, pp. 106–121, January 2009.

[18] L. Mancera and J. Portilla, "L0-norm-based sparse representation through alternate projections," in *IEEE International Conference on Image Processing (ICIP 2006)*, Atlanta, GA, USA, October 2006, pp. 2089–2092.

[19] 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.

[20] P. Huber, *Robust Statistics*, Wiley, New York, NY, USA, 1st edition, 1981.

[21] D. Geman and G. Reynolds, "Nonlinear image recovery with half-quadratic regularization," *IEEE Transactions on Image Processing*, vol. 4, no. 7, pp. 932–946, July 1995.

[22] C.A. Sing-Long, C.A. Tejos, and P. Irarrazaval, "Evaluation of continuous approximation functions for the l_0-norm for compressed sensing," in *International Soc. Mag. Reson. Med. (ISMRM 2009)*, 2009, vol. 17, p. 4585.