A Simple Set-Membership Affine Projection Algorithm for Sparse System Modeling

Hamed Yazdanpanah, Paulo S. R. Diniz, Markus V. S. Lima
Federal University of Rio de Janeiro (UFRJ)
Rio de Janeiro, RJ, Brazil, Zip 21941-972, P.O. Box 68504
{hamed.yazdanpanah, diniz, markus.lima}@smt.ufrj.br

Abstract—**In this paper, we derive two algorithms, namely the Simple Set-Membership Affine Projection (S-SM-AP) and the improved S-SM-AP (IS-SM-AP), in order to exploit the sparsity of an unknown system while focusing on having low computational complexity. To achieve this goal, the proposed algorithms apply a discard function on the weight vector to disregard the coefficients close to zero during the update process. In addition, the IS-SM-AP algorithm reduces the overall number of computations required by the adaptive filter even further by replacing small coefficients with zero. Simulation results show similar performance when comparing the proposed algorithm with some existing state-of-the-art sparsity-aware algorithms while the proposed algorithms require lower computational complexity.**

Index Terms—**adaptive filtering, set-membership filtering, sparsity, discard function, computational complexity.**

I. Introduction

Adaptive filtering applied to signals originating from time-varying systems find applications in a wide diversity of areas such as communications, control, radar, acoustics, and speech processing. Nowadays, it is well known that many types of signal or system parameters admit sparse representation in a certain domain. However, classical adaptive algorithms such as the least-mean square (LMS), the normalized LMS (NLMS), the affine projection (AP), and the recursive least squares (RLS) do not take into consideration the sparsity in the signal or system models.

Recently, it has been understood that by appropriately exploiting signal sparsity, significant improvement in convergence rate and steady-state performance can be achieved. As a consequence, many extensions of the classical algorithms were proposed aiming at exploiting sparsity. One of the most widely used approaches consists in updating each filter coefficient using a step-size proportional to its magnitude, which led to the development of a family of algorithms known as *proportionate* [1]–[5]. Another interesting approach to exploit sparsity is to include a *sparsity-promoting penalty* function into the original optimization problem of classical algorithms. Within this approach, most algorithms employ the l_1 norm as the sparsity-promoting penalty [6], [7], but recently an approximation to the l_0 norm has shown some advantages [8]–[11]. In addition, these two approaches were combined and tested in [12], [13] yielding interesting results. Observe that in all of the aforementioned approaches something is being included/added to the classical algorithms, thus entailing an increase in their computational complexity.

In this paper we use a different strategy to exploit sparsity. Instead of including additional features in the algorithm, as the techniques presented in the previous paragraph do, we actually discard some coefficients, thus reducing the computational burden. This idea is motivated by the existence of some uncertainty in the coefficients in practical applications. Indeed, a measured sparse impulse response of a system/signal presents a few coefficients concentrating most of the energy, whereas the other coefficients are close to zero, but not exactly equal to zero [8]. Thus, if we have some prior information about the uncertainty in those coefficients, then we can replace the coefficients which are "lower than" this uncertainty with zero (i.e., discard the coefficients) in order to save computational resources.

In addition to this new way of exploiting sparsity, we also employ the set-membership filtering (SMF) approach [14], [15] in order to generate the Simple Set-Membership Affine Projection (S-SM-AP) algorithm, which is essentially the combination of the set-membership affine projection (SM-AP) algorithm [16] with our strategy to exploit sparsity. The SMF approach is used just to reduce the computational burden even further, since the filter coefficients are updated only when the estimation error is higher than the pre-determined upper bound.

The rest of this paper is organized as follows. Section II briefly describes the SMF concept. The proposed S-SM-AP algorithm is derived in Section III. Simulations are presented in Section IV and Section V contains the conclusions.

Notation: The real field is represented by $\mathbb{R}$. The order of the adaptive filter is denoted as N. Scalars are denoted by lowercase letters. Boldface lowercase (uppercase) letters represent vectors (matrices). Superscript ^{T} represents transpose operation.

II. Set-Membership Filtering (SMF)

The purpose of the SMF is to obtain the system model parameter $\mathbf{w}$ such that the magnitude of the estimation error is upper bounded by a prescribed parameter $\overline{\gamma} \in \mathbb{R}_+$. The value of $\overline{\gamma}$ can change with the specific application. If $\overline{\gamma}$ is suitably selected, then there are many valid estimates for $\mathbf{w}$. Suppose that $\mathcal{S}$ denotes the set of all possible input-desired data pairs $(\mathbf{x}, d)$ of interest. Define the feasibility set Θ as

$$\Theta = \bigcap_{(\mathbf{x},d)\in\mathcal{S}} \{\mathbf{w} \in \mathbb{R}^{N+1} : |d - \mathbf{w}^T\mathbf{x}| \leq \overline{\gamma}\}. \qquad (1)$$

978-0-9928-6265-7/16/$31.00 ©2016 IEEE

Also, let's define the constraint set $\mathcal{H}(k)$ consisting of all vectors $\mathbf{w}$ such that their estimation errors at time instant k are upper bounded in magnitude by $\overline{\gamma}$, i.e.,

$$\mathcal{H}(k) = \{\mathbf{w} \in \mathbb{R}^{N+1} : |d(k) - \mathbf{w}^T\mathbf{x}(k)| \leq \overline{\gamma}\}, \quad (2)$$

where the quantities $\mathbf{x}(k)$, $\mathbf{w}$, and $d(k)$ are input vector, weight vector, and desired signal, respectively. The membership set $\psi(k)$ is defined as

$$\psi(k) = \bigcap_{i=0}^{k} \mathcal{H}(i). \quad (3)$$

The idea of iterative techniques based on SMF is to adapt the coefficient vector such that it will always remain within the membership set, since this set converges to the feasibility set as $k \to \infty$. Due to practical constraints that prevent the computation of $\psi(k)$, we calculate a point estimate using, for example, the information provided by the constraint set $\mathcal{H}(k)$ like in the set-membership NLMS algorithm [14], or several previous constraint sets as is done in the set-membership affine projection algorithm [16].

III. A Simple Set-Membership Affine Projection Algorithm

In Subsection III-A, we propose a Simple Set-Membership Affine Projection (S-SM-AP) algorithm that exploits the sparsity of the involved system in order to obtain lower computational complexity than the existing sparsity-aware algorithms. For this purpose, the strategy consists in not updating the coefficients of the sparse filter which are close to zero. Then, in Subsection III-B, we include a discussion of some characteristics of the proposed algorithm. Finally, in Subsection III-C, we introduce an improved version of the proposed algorithm aiming at reducing computational burden even further.

A. Derivation of the S-SM-AP algorithm

Let us define the *discard function* $f_\epsilon : \mathbb{R} \to \mathbb{R}$ for the positive constant ϵ as follows

$$f_\epsilon(w) = \left\{ \begin{array}{ll} w & \text{if } |w| > \epsilon \\ 0 & |w| \leq \epsilon \end{array} \right. . \quad (4)$$

That is, function f_ϵ discards the values of w which are close to zero. The parameter ϵ defines what is considered as close to zero and, therefore, should be chosen based on some *a priori* information about the relative importance of a coefficient to the sparse system. Figure 1 depicts the function $f_\epsilon(w)$ for $\epsilon = 10^{-4}$. Note that the function $f_\epsilon(w)$ is not differentiable at $\pm\epsilon$, however, we need to differentiate this function in order to derive the S-SM-AP algorithm. To address this issue, we define the derivative of $f_\epsilon(w)$ at $+\epsilon$ and $-\epsilon$ as equal to the left and the right derivatives, respectively. Thus, the derivative of $f_\epsilon(w)$ at $\pm\epsilon$ is zero. Define the *discard vector function* $\mathbf{f}_\epsilon : \mathbb{R}^{N+1} \to \mathbb{R}^{N+1}$ as $\mathbf{f}_\epsilon(\mathbf{w}) = [f_\epsilon(w_0), \cdots, f_\epsilon(w_N)]^T$.

The S-SM-AP algorithm updates the coefficients whose absolute values are larger than ϵ whenever the error is such that $|e_0(k)| = |d(k) - \mathbf{w}^T(k)\mathbf{x}(k)| > \overline{\gamma}$. Let $\psi^{L+1}(k)$ denote the intersection of the last $L+1$ constraint sets and state

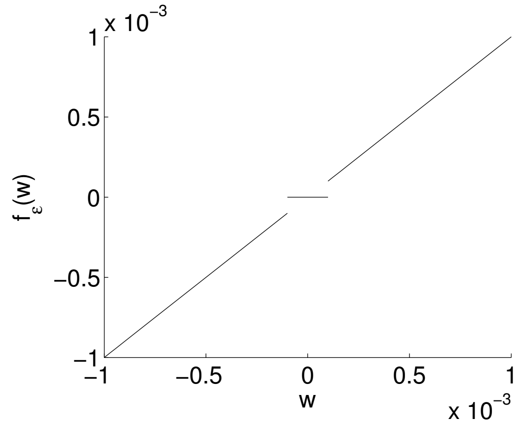

Figure 1. Discard function $f_\epsilon(w)$ for $\epsilon = 10^{-4}$.

the following optimization criterion for the vector update whenever $\mathbf{w}(k) \notin \psi^{L+1}(k)$

$$\begin{aligned} &\min \frac{1}{2}\|\mathbf{f}_\epsilon(\mathbf{w}(k+1)) - \mathbf{w}(k)\|^2 \\ &\text{subject to} \\ &\mathbf{d}_{\text{ap}}(k) - \mathbf{X}_{\text{ap}}^T(k)\mathbf{w}(k+1) = \bar{\boldsymbol{\gamma}}(k). \end{aligned} \quad (5)$$

where

$\mathbf{d}_{\text{ap}}(k) \in \mathbb{R}^{(L+1)\times 1}$	contains the desired output from the $L+1$ last time instants;
$\bar{\boldsymbol{\gamma}}(k) \in \mathbb{R}^{(L+1)\times 1}$	specifies the point in $\psi^{L+1}(k)$;
$\mathbf{X}_{\text{ap}}(k) \in \mathbb{R}^{(N+1)\times(L+1)}$	contains the corresponding input vectors, i.e.,

$$\begin{aligned} \mathbf{d}_{\text{ap}}(k) &= [d(k)\ d(k-1)\ \cdots\ d(k-L)]^T, \\ \mathbf{e}_{\text{ap}}(k) &= [e_0(k)\ e_1(k)\ \cdots\ e_L(k)]^T, \\ \bar{\boldsymbol{\gamma}}(k) &= [\overline{\gamma}_0(k)\ \overline{\gamma}_1(k)\ \cdots\ \overline{\gamma}_L(k)]^T, \\ \mathbf{X}_{\text{ap}}(k) &= [\mathbf{x}(k)\ \mathbf{x}(k-1)\ \cdots\ \mathbf{x}(k-L)], \end{aligned} \quad (6)$$

with $\mathbf{x}(k)$ being the input-signal vector

$$\mathbf{x}(k) = [x(k)\ x(k-1)\ \cdots\ x(k-N)]^T \quad (7)$$

and $\mathbf{e}_{\text{ap}}(k) = \mathbf{d}_{\text{ap}}(k) - \mathbf{X}_{\text{ap}}^T(k)\mathbf{w}(k)$ being the error vector.

In order to solve this optimization problem, we construct the Lagrangian $\mathbb{L}$ as

$$\begin{aligned} \mathbb{L} = &\frac{1}{2}\|\mathbf{f}_\epsilon(\mathbf{w}(k+1)) - \mathbf{w}(k)\|^2 \\ &+ \boldsymbol{\lambda}_{\text{ap}}^T(k)[\mathbf{d}_{\text{ap}}(k) - \mathbf{X}_{\text{ap}}^T(k)\mathbf{w}(k+1) - \bar{\boldsymbol{\gamma}}(k)], \end{aligned} \quad (8)$$

where $\boldsymbol{\lambda}_{\text{ap}}(k) \in \mathbb{R}^{L+1}$ is a vector of Lagrange multipliers. After differentiating the above equation with respect to $\mathbf{w}(k+1)$ and setting the result equal to zero, we obtain

$$\mathbf{f}_\epsilon(\mathbf{w}(k+1)) = \mathbf{w}(k) + \mathbf{F}_\epsilon^{-1}(\mathbf{w}(k+1))\mathbf{X}_{\text{ap}}(k)\boldsymbol{\lambda}_{\text{ap}}(k), \quad (9)$$

where $\mathbf{F}_\epsilon(\mathbf{w}(k+1))$ is the Jacobian matrix of $\mathbf{f}_\epsilon(\mathbf{w}(k+1))$. In Equation (9) by employing a similar strategy as the PASTd (projection approximation subspace tracking with deflation) [17], we replace $\mathbf{f}_\epsilon(\mathbf{w}(k+1))$ and $\mathbf{F}_\epsilon^{-1}(\mathbf{w}(k+1))$ with $\mathbf{w}(k+1)$ and $\mathbf{F}_\epsilon^{-1}(\mathbf{w}(k))$, respectively, in order to form the recursion, then we obtain

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \mathbf{F}_\epsilon^{-1}(\mathbf{w}(k))\mathbf{X}_{\text{ap}}(k)\boldsymbol{\lambda}_{\text{ap}}(k). \quad (10)$$

If we substitute the above equation in the constraint relation (5), then we will find $\boldsymbol{\lambda}_{\rm ap}(k)$ as follows

$$\boldsymbol{\lambda}_{\rm ap}(k) = \mathbf{A}(k)(\mathbf{e}_{\rm ap}(k) - \bar{\boldsymbol{\gamma}}(k)), \qquad (11)$$

where

$$\mathbf{A}(k) = (\mathbf{X}_{\rm ap}^T(k)\mathbf{F}_\epsilon^{-1}(\mathbf{w}(k))\mathbf{X}_{\rm ap}(k))^{-1}. \qquad (12)$$

Replacing (11) into (10) leads to the following updating equation

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \mathbf{p}(k), \qquad (13)$$

where

$$\mathbf{p}(k) = \mathbf{F}_\epsilon^{-1}(\mathbf{w}(k))\mathbf{X}_{\rm ap}(k)\mathbf{A}(k)(\mathbf{e}_{\rm ap}(k) - \bar{\boldsymbol{\gamma}}(k)). \qquad (14)$$

Note that $\mathbf{F}_\epsilon(\mathbf{w}(k))$ is not an invertible matrix and, therefore, we apply the Moore-Penrose pseudoinverse (generalization of the inverse matrix) instead of the standard inverse. However, $\mathbf{F}_\epsilon(\mathbf{w}(k))$ is a diagonal matrix with diagonal entries equal to zero or one. Indeed, for the components of $\mathbf{w}(k)$ whose absolute values are larger than ϵ, their corresponding entries on the diagonal matrix $\mathbf{F}_\epsilon(\mathbf{w}(k))$ are one, whereas the remaining entries are zero. Hence, the pseudoinverse of $\mathbf{F}_\epsilon(\mathbf{w}(k))$ is again $\mathbf{F}_\epsilon(\mathbf{w}(k))$. As a result, the update equation of the S-SM-AP algorithm is as follows

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \mathbf{q}(k), \qquad (15)$$

where

$$\mathbf{q}(k) = \mathbf{F}_\epsilon(\mathbf{w}(k))\mathbf{X}_{\rm ap}(k)\mathbf{B}(k)(\mathbf{e}_{\rm ap}(k) - \bar{\boldsymbol{\gamma}}(k)), \qquad (16)$$

$$\mathbf{B}(k) = [\mathbf{X}_{\rm ap}^T(k)\mathbf{F}_\epsilon(\mathbf{w}(k))\mathbf{X}_{\rm ap}(k)]^{-1}. \qquad (17)$$

Note that, in practice, we apply a regularization factor in (17) in order to avoid numerical problems in the matrix inversion.

B. Discussion of the S-SM-AP algorithm

1) Computational Complexity: The update equation of the S-SM-AP algorithm is similar to the update equation of the SM-AP algorithm, but the former updates only the subset of coefficients of $\mathbf{w}(k)$ whose absolute values are larger than ϵ. As a result, the role of matrix $\mathbf{F}_\epsilon(\mathbf{w}(k))$ is to discard some coefficients of $\mathbf{w}(k)$, thus reducing the computational complexity when compared to the SM-AP algorithm.

The computational complexity for each update of the weight vector of the SM-PAPA [4], the SSM-AP [8], and the proposed S-SM-AP algorithms are listed in Table I. The filter order and the memory length factors are N and L, respectively. It should be noted that the number of operations in Table I are presented for the full update of all coefficients. In other words, for the S-SM-AP algorithm we have presented the worst case scenario which is equivalent to setting $\epsilon = 0$,[1] while in practice we are updating only the coefficients with absolute values larger than a pre-determined positive constant. Also, it is notable that the number of divisions in the S-SM-AP algorithm is less than the SM-PAPA and SSM-AP algorithms. This is quite significant, as divisions require more computational complexity than other operations.

[1]In this case, the complexity of the S-SM-AP and SM-AP algorithms are exactly the same.

2) Initialization: Unlike classical algorithms in which the initialization of the weight vector is often chosen as $\mathbf{w}(0) = \mathbf{0}$, this same procedure cannot be applied to the proposed algorithm. Indeed, for the S-SM-AP algorithm, each of the coefficients should be initialized as $|w_i(0)| > \epsilon$ for $i = 0, 1, \ldots, N$.

3) Relation with other algorithms: The similarities and differences between the proposed algorithm and the SM-AP algorithm were already addressed when we discussed the complexity of these algorithms. Now, one should observe that the update equation of the S-SM-AP algorithm is similar to the one of the set-membership partial update affine projection (SM-PUAP) algorithm [15], in which our matrix $\mathbf{F}_\epsilon(\mathbf{w}(k))$ is replaced by a diagonal matrix $\mathbf{C}$ also with entries equal to 1 or 0, but there is no specific form to set/select $\mathbf{C}$. Therefore, the proposed algorithm can be considered as a particular case of the SM-PUAP in which there is a mathematically defined way (based on the sparsity of the unknown system) to select the coefficients that are relevant and the ones that will be discarded. Regarding the memory requirements of the proposed algorithm, they are exactly the same as in the AP algorithm, i.e., determined by the data-reuse factor L.

C. The Improved S-SM-AP (IS-SM-AP) algorithm

As we can see in the update equation of the S-SM-AP algorithm, if a coefficient of the weight vector falls inside the interval $[-\epsilon, +\epsilon]$, then in the next update this coefficient does not update since it is eliminated by the discard function. On the other hand, the coefficients $w_i(k)$ inside the interval $[-\epsilon, +\epsilon]$ are close to zero and the best intuitive approximation for them is zero (the center of the interval). Besides, making these coefficients $w_i(k)$ equal to zero implies in a reduction of computational complexity, because it reduces the number of operations required to compute the output of the adaptive filter $y(k) = \mathbf{x}^T(k)\mathbf{w}(k)$.[2] For this purpose, we multiply $\mathbf{w}(k)$ by $\mathbf{F}_\epsilon(\mathbf{w}(k))$, and obtain the Improved S-SM-AP (IS-SM-AP) algorithm as follows

$$\mathbf{w}(k+1) = \mathbf{F}_\epsilon(\mathbf{w}(k))\mathbf{w}(k) + \mathbf{q}(k). \qquad (18)$$

IV. SIMULATIONS

Here, we have applied the IS-SM-AP, the SSM-AP [8], the SM-PAPA [8], and the NLMS algorithms to identify three unknown sparse systems of order 14.[3] The first one is an arbitrary sparse system $\mathbf{w}_o$, the second one is a block sparse system $\mathbf{w}_o'$, and the third one is a symmetric-block sparse system $\mathbf{w}_o''$. The coefficients of these three systems are presented in Table II. The input is a BPSK (binary phase-shift keying) signal with variance $\sigma_x^2 = 1$. The signal-to-noise ratio (SNR) is set to be 20 dB, i.e., the noise variance is $\sigma_n^2 = 0.01$. The data-reuse factor is $L = 1$, the bound on the estimation error is set to be $\bar{\gamma} = \sqrt{5\sigma_n^2}$, and the

[2]This additional reduction in the number of operations becomes more important as the filter order increases. For instance, in acoustic echo cancellation systems, in which the adaptive filter has some thousands of coefficients [18], [19], this simple strategy implies in significant computational savings.

[3]The results for the S-SM-AP algorithm are not shown here because they are almost equal to the results of the IS-SM-AP algorithm, but the latter has the advantage of requiring fewer computations.

Table I
NUMBER OF OPERATIONS FOR SM-PAPA, SSM-AP, AND S-SM-AP ALGORITHMS

Algorithm	Addition & Subtraction	Multiplication	Division
SM-PAPA	$N^2 + (L^2 + 4L + 5)N +$ $(2L^3 + 5L^2 + 7L + 5)$	$(L^2 + 5L + 7)N +$ $(2L^3 + 6L^2 + 9L + 8)$	$2N + (2L^2 + 4L + 4)$
SSM-AP	$(L^2 + 6L + 7)N +$ $(2L^3 + 6L^2 + 9L + 7)$	$(L^2 + 6L + 9)N +$ $(2L^3 + 7L^2 + 12L + 11)$	$N + (2L^2 + 4L + 3)$
S-SM-AP	$\frac{1}{2}(L^2 + 5L + 6)N$ $\frac{1}{2}(L^3 + 4L^2 + 11L + 8)$	$\frac{1}{2}(L^2 + 5L + 6)N$ $\frac{1}{2}(L^3 + 6L^2 + 11L + 8)$	L^2

Table II
THE COEFFICIENTS OF UNKNOWN SYSTEMS $\mathbf{w}_o$, $\mathbf{w}'_o$, AND $\mathbf{w}''_o$

$\mathbf{w}_o$	**24e-4**	2e-8	**-23e-4**	-3e-7	**5e-1**	-1e-9	**2e-2**	1e-7	-5e-5	12e-6	1e-8	-5e-6	4e-6	-1e-5	**-2e-3**
$\mathbf{w}'_o$	2e-7	-21e-10	17e-8	21e-6	-3e-7	**24e-4**	**5e-1**	**2e-2**	**33e-4**	**-2e-3**	-5e-5	18e-9	-5e-6	28e-7	-19e-6
$\mathbf{w}''_o$	2e-8	-1e-9	1e-7	-3e-7	**-64e-3**	**2e-1**	**5e-1**	**2e-1**	**-64e-3**	-5e-5	12e-6	1e-8	-5e-6	4e-6	-1e-5

threshold bound vector $\bar{\gamma}(k)$ is selected as the *simple-choice constraint vector* [8] which is defined as $\bar{\gamma}_0(k) = \frac{\bar{\gamma} e_0(k)}{|e_0(k)|}$ and $\bar{\gamma}_i(k) = d(k-i) - \mathbf{w}^T(k)\mathbf{x}(k-i)$, for $i = 1, \cdots, L$. The convergence factor of the NLMS algorithm is $\mu = 0.9$. The initial vector $\mathbf{w}(0)$ and the regularization factor are $10^{-3} \times [1, \cdots, 1]^T$ and 10^{-12}, respectively. The constant ϵ in the IS-SM-AP algorithm is chosen as 2×10^{-4}; that is, on average, 5 out of 15 coefficients (boldface coefficients in $\mathbf{w}_o$, $\mathbf{w}'_o$, and $\mathbf{w}''_o$ shown in Table II) are updated at each iteration. We have selected $\alpha = 5 \times 10^{-3}$, $\beta = 5$, and $\varepsilon = 100$ for the SM-PAPA and the SSM-AP algorithms. The learning curves are the results of averaging of the outcomes of 500 trials. Figures 2, 3, and 4 depict the learning curves for the IS-SM-AP, the SM-PAPA, the SSM-AP, and the NLMS algorithms to identify the unknown systems $\mathbf{w}_o$, $\mathbf{w}'_o$, and $\mathbf{w}''_o$, respectively. In the case of $\mathbf{w}_o$ and $\mathbf{w}'_o$, the average number of updates implemented by the IS-SM-AP, the SM-PAPA, and the SSM-AP algorithms are 6.3%, 5.3%, and 8.9%, respectively. In the case of $\mathbf{w}''_o$, the average number of updates performed by the IS-SM-AP, the SM-PAPA, and the SSM-AP algorithms are 7.6%, 5.9%, and 20.5%, respectively.

In addition, we have applied all the aforementioned algorithms, using the parameters that were already defined in the previous paragraph, but changing the input signal model to an autoregressive (AR) process in order to identify the unknown system $\mathbf{w}_o$. The new input signal is generated as a first-order AR process defined as $x(k) = 0.95x(k-1) + n(k)$. In this case, the learning curves of the algorithms are shown in Figure 5, and the average number of updates performed by the IS-SM-AP, the SM-PAPA, and the SSM-AP algorithms are 8.4%, 7.7%, and 5.6%, respectively. Also, the overall number of arithmetic operations required by the IS-SM-AP, the SM-PAPA, and the SSM-AP algorithms are 41635, 110835, and 84396, respectively. That is, even in a scenario where the IS-SM-AP realizes more updates than the competing algorithms, the overall number of arithmetic operations it performs is inferior due to its reduced complexity.

Regarding the MSE performance, in every scenario we tested the IS-SM-AP algorithm performed as good as the other state-of-the-art sparsity-aware algorithms, but our proposal has the advantage of requiring fewer computations, since at each iteration in which an update occurs only a subset (on average, one third) of the coefficients was actually updated. In addition, we observed that the MSE of the IS-SM-AP algorithm was always very similar to the MSE obtained using the SM-PAPA. Another interesting observation is that both the IS-SM-AP and SM-PAPA algorithms worked better with BPSK input signal, whereas the SSM-AP algorithm was better when a correlated input signal was used.

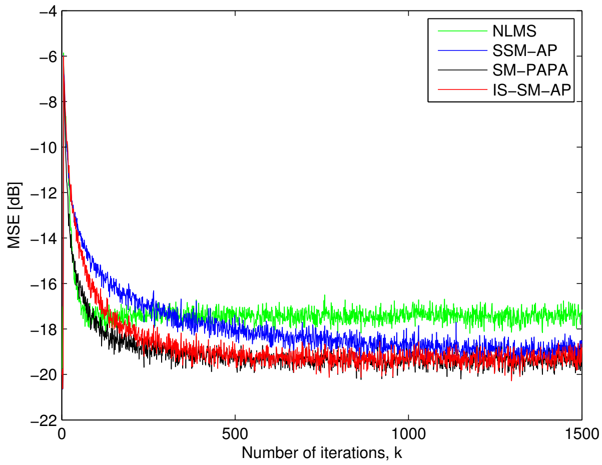

Figure 2. The learning curves of the SM-PAPA, the SSM-AP, the IS-SM-AP, and NLMS algorithms applied on $\mathbf{w}_o$.

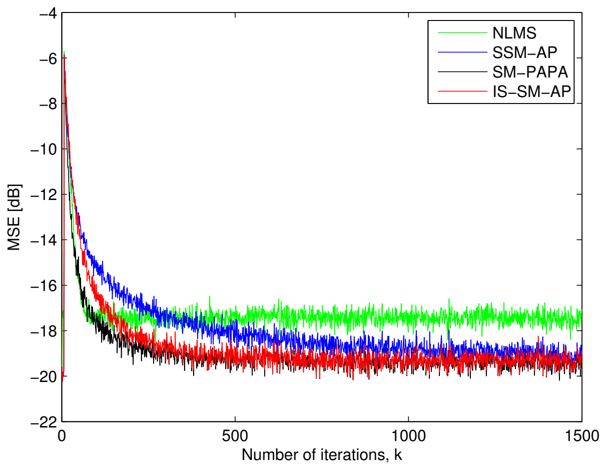

Figure 3. The learning curves of the SM-PAPA, the SSM-AP, the IS-SM-AP, and NLMS algorithms applied on $\mathbf{w}'_o$.

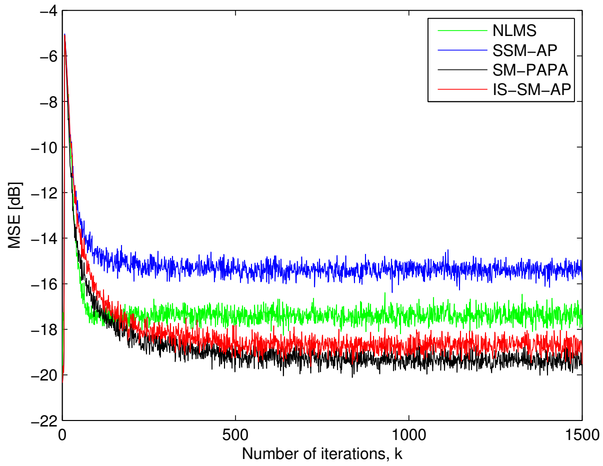

Figure 4. The learning curves of the SM-PAPA, the SSM-AP, the IS-SM-AP, and NLMS algorithms applied on $\mathbf{w}_o''$.

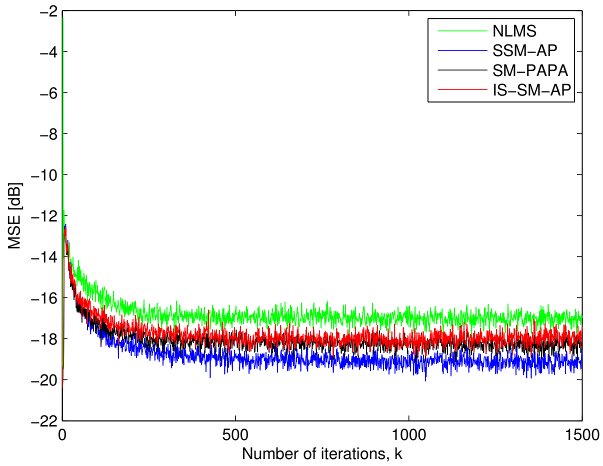

Figure 5. The learning curves of the SM-PAPA, the SSM-AP, the IS-SM-AP, and NLMS algorithms applied on $\mathbf{w}_o$ using an auto-regressive input signal.

V. Conclusions

In this paper, we have proposed the S-SM-AP and the IS-SM-AP algorithms to take advantage of sparsity in the signal models while attaining low computational complexity. To reach this target, we have derived a simple update equation which only updates the filter coefficients whose magnitudes are larger than a pre-determined value. Also, this method is jointly applied with the well-known set-membership approach aiming at obtaining even lower computational complexity and better convergence rate. The simulation results have shown the excellent performance of the algorithm and lower computational complexity compared to some other sparsity-aware data-selective adaptive filters. Indeed, the proposed algorithm performed as well as the SM-PAPA algorithm while requiring fewer arithmetic operations (for the scenarios described in Section IV, it entailed about 38% of the operations spent by the SM-PAPA).

As for the future works, we intend to analyze the proposed algorithms, derive their complex-valued version, adapt these algorithms to tackle the problem of identifying time-varying sparse systems, and compare the proposed algorithms with algorithms following the Bayesian approach, such as the ones introduced in [20]–[22].

References

[1] 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, Sept. 2000.

[2] 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.

[3] 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.

[4] S. Werner, J.A. Apolinario 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. 1–10, Jan. 2007.

[5] C. Paleologu, S. Ciochina, and J. Benesty, "An efficient proportionate affine projection algorithm for echo cancellation," *IEEE Signal Processing Letters*, vol. 17, no. 2, pp. 165–168, Feb 2010.

[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*, London, U.K., Sept. 2011, pp. 1–5.

[7] 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.

[8] M.V.S. Lima, T.N. Ferreira, W.A. Martins, and P.S.R. Diniz, "Sparsity-aware data-selective adaptive filters," *IEEE Transactions on Signal Processing*, vol. 62, no. 17, pp. 4557–4572, Sept. 2014.

[9] M.V.S. Lima, I. Sobron, W.A. Martins, and P.S.R. Diniz, "Stability and MSE analyses of affine projection algorithms for sparse system identification," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2014)*, Florence, Italy, May 2014, pp. 6399–6403.

[10] 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.

[11] 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, Sept. 2009.

[12] K. Pelekanakis and M. Chitre, "New sparse adaptive algorithms based on the natural gradient and the l0-norm," *IEEE Journal of Oceanic Engineeering*, vol. 38, no. 2, pp. 323–332, 2013.

[13] T.N. Ferreira, M.V.S. Lima, P.S.R. Diniz, and W.A. Martins, "Low-complexity proportionate algorithms with sparsity-promoting penalties," in *IEEE International Symposium on Circuits and Systems (ISCAS 2016)*, Canada, May 2016 (accepted).

[14] 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, no. 5, pp. 111–114, May 1998.

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

[16] S. Werner and P.S.R. Diniz, "Set-membership affine projection algorithm," *IEEE Signal Processing Letters*, vol. 8, no. 8, pp. 231–235, Aug. 2001.

[17] X. Wang and H.V. Poor, *Wireless Communication Systems: Advanced Techniques for Signal Reception*, Prentice Hall, Upper Saddle River, NJ, 2004.

[18] E. Hansler and G. Schmidt, *Acoustic Echo and Noise Control: A Practical Approach*, Wiley, Hoboken, NJ, USA, 2004.

[19] J. Benesty, T. Gansler, D. R. Morgan, M. M. Sondhi, and S. L. Gay, *Advances in Network and Acoustic Echo Cancellation*, Springer, Berlin Heidelberg, Germany, 2010.

[20] K.E. Themelis, A.A. Rontogiannis, and K.D. Koutroumbas, "A variational bayes framework for sparse adaptive estimation," *IEEE Transactions on Signal Processing*, vol. 62, no. 18, pp. 4723–4736, Sept. 2014.

[21] P.V. Giampouras, A.A. Rontogiannis, K.E. Themelis, and K.D. Koutroumbas, "Online bayesian low-rank subspace learning from partial observations," in *European Signal Processing Conference (EUSIPCO 2015)*, Nice, France, Sept. 2015, pp. 2526–2530.

[22] K.E. Themelis, A.A. Rontogiannis, , and K.D. Koutroumbas, "Online bayesian group sparse parameter estimation using a generalized inverse gaussian markov chain," in *European Signal Processing Conference (EUSIPCO 2015)*, Nice, France, Sept. 2015, pp. 1686–1690.