FAST LEARNING SET THEORETIC ESTIMATION

Markus V.S. Lima and Paulo S.R. Diniz[1]

Universidade Federal do Rio de Janeiro
Email: markus.lima@ieee.org, diniz@smt.ufrj.br

ABSTRACT

This paper addresses set theoretic estimation used for online learning in an adaptive filtering context. The advantages of set theoretic estimation over the traditional point estimation are shown, among which we highlight the capability of reducing the computational burden leading to energy saving. The set-membership affine projection (SM-AP) algorithm is the main framework because it generalizes many of the set theoretic algorithms, besides having a popular point estimation counterpart for benchmarking, viz. the affine projection (AP) algorithm. In addition, we discuss the effects of the design of the involved sets in convergence speed and steady-state MSE. Each iteration of the SM-AP algorithm exploits the intersection of constraint sets and, although any point in this set is acceptable, some of its parts should be avoided during the update. Moreover, we propose a new configuration for the error constraints, which leads to low steady-state MSE, high convergence speed, and low probability of update.

***Index Terms*—** set theoretic estimation, set-membership

1. INTRODUCTION

For a long time, machine learning mechanisms have been based on the classical theory of point estimation. Nevertheless, the importance of set estimation theory is growing as the advantages of such a paradigm become clearer. Set theoretic optimization is the proper approach to tackle problems in which uncertainty is unavoidable, since instead of searching for a unique point within the feasible region that minimizes or maximizes some objective function, it searches for a set of points which are acceptable solutions given the inherent uncertainty of the problem [1]. This major feature of set theoretic optimization provides two main advantages over the classical approach: (i) *robustness against noise* provided some information about the nature of the problem is used, and (ii) *energy saving*, since the innovation in the observed data is checked before the data are used in the learning process.

This paper addresses set theoretic estimation used for online learning in adaptive filtering context. The set-membership filtering (SMF) concept is presented and the set-membership affine projection (SM-AP) algorithm is described. The SM-AP algorithm was chosen as the main family of algorithms because it generalizes many of the set theoretic adaptive filtering algorithms and due to the importance and existence of its point estimation counterpart, viz. affine projection (AP) algorithm. Indeed, the AP algorithm encompasses widely used algorithms of the least mean square (LMS) family such as the normalized LMS (NLMS) and binormalized LMS (BNLMS) and, therefore, is used for benchmarking.

The SM-AP algorithm can be seen as an iterative procedure, based on the intersection of constraint sets, to estimate the feasibility set, the set of acceptable solutions. The general form of the SM-AP algorithm, see [2, 3], requires a judicious choice of the *a posteriori* error constraints in order to control noise enhancement effects.

In this paper we consider some theoretic aspects of the learning stage of the SM-AP algorithm in order to introduce a general rule for setting the *a posteriori* error constraints. We also propose a smart way to preset them. This way it is possible to show how the constraint sets can be properly exploited so that the SM-AP algorithm yields high convergence speed and low steady-state mean square error (MSE).

This paper is organized as follows. Section 2 describes the set theoretic foundation employed in the SMF theory. Section 3 covers the SM-AP algorithm. In Section 4, we discuss the importance of an adequate choice for the parameter called constraint vector (CV). Indeed, we explain why a general choice of the CV usually leads to poor results (high MSE) in practical applications and we state a rule for properly choosing the CV. In Section 5, we compare the AP versus the SM-AP algorithm employing different CVs, including the one proposed in this paper, in terms of steady-state MSE level in stationary and nonstationary environments and convergence speed. The conclusions are drawn in Section 6.

2. SET-MEMBERSHIP FILTERING (SMF)

SMF is a set theoretic estimation paradigm suitable to adaptive filtering problems that are linear-in-parameters [4]. In the SMF theory, we are interested in a set of feasible/acceptable solutions, called *feasibility set*, rather than a single solution. The main advantage of the SMF over the standard point estimation theory is that the former can efficiently model the uncertainty inherited from the observed data leading to better estimates, in terms of MSE, as well as energy saving. This section briefly covers the SMF theory.

Let $e, d, y \in \mathbb{R}$ be the error, desired, and output signals of

This work was partially supported by CNPq, CAPES, and FAPERJ.

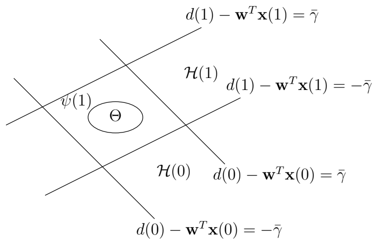


Fig. 1. Example of SMF concept: $\psi(1) = \mathcal{H}(0) \bigcap \mathcal{H}(1)$.

the adaptive filter, respectively. The output signal is defined as $y \triangleq \mathbf{w}^T\mathbf{x}$, where $\mathbf{w}, \mathbf{x} \in \mathbb{R}^{N+1}$ are the adaptive filter coefficient vector and input vector, respectively, and N is the filter order. The error signal is defined as $e \triangleq d - y$.

The SMF criterion is to estimate $\mathbf{w}$ that leads to an error signal whose magnitude is upper bounded by a constant $\overline{\gamma} \in \mathbb{R}_+$ for all possible pairs $(\mathbf{x}, d)$. That is, denoting by $\mathcal{S}$ the set comprised of all possible pairs $(\mathbf{x}, d)$, the SMF criterion aims at finding $\mathbf{w}$ satisfying $|e| = |d - \mathbf{w}^T\mathbf{x}| \leq \overline{\gamma}$, $\forall (\mathbf{x}, d) \in \mathcal{S}$. Hence, there exists a set Θ of solutions, called *feasibility set*, which is defined as $\Theta \triangleq \bigcap_{(\mathbf{x},d)\in\mathcal{S}} \{\mathbf{w} \in \mathbb{R}^{N+1} : |d - \mathbf{w}^T\mathbf{x}| \leq \overline{\gamma}\}$, and any $\mathbf{w} \in \Theta$ is an acceptable solution.

For online learning, an iterative procedure must be used to estimate Θ. At a given iteration k, we can use all previous data to estimate Θ via the *exact membership set* $\psi(k) \triangleq \bigcap_{i=0}^{k} \mathcal{H}(i)$, where $\mathcal{H}(i) \triangleq \{\mathbf{w} \in \mathbb{R}^{N+1} : |d(i) - \mathbf{w}^T\mathbf{x}(i)| \leq \overline{\gamma}\}$ is the *constraint set*, the set comprised of all $\mathbf{w}$ satisfying the error bound at iteration i. A geometrical interpretation of the involved sets is depicted in Fig. 1. For more details see [3, 5].

3. SET-MEMBERSHIP AFFINE PROJECTION

The set-membership affine projection (SM-AP) algorithm [2] is an iterative procedure, based on the SMF theory, whose updating rule exploits the intersection of constraint sets related to the last iterations. The use of more sets in the updating process leads to higher convergence speed. This section describes the SM-AP algorithm.

Assume we have available the last $L+1$ data-pairs of input vectors and desired signals, where L is a nonnegative integer called *data reuse factor*. The variables $\mathbf{e}(k), \mathbf{d}(k), \mathbf{y}(k) \in \mathbb{R}^{L+1}$ represent the error vector, desired vector, and output vector corresponding to the adaptive filter. The output vector is defined as $\mathbf{y}(k) \triangleq \mathbf{X}^T(k)\mathbf{w}(k)$, where $\mathbf{X}(k) \in \mathbb{R}^{(N+1)\times(L+1)}$ is the input matrix, and the error vector is given by $\mathbf{e}(k) \triangleq \mathbf{d}(k) - \mathbf{y}(k) = \mathbf{d}(k) - \mathbf{X}^T(k)\mathbf{w}(k)$. The *a posteriori* error vector is denoted as $\boldsymbol{\varepsilon}(k) \triangleq \mathbf{d}(k) - \mathbf{X}^T(k)\mathbf{w}(k+1) \in \mathbb{R}^{L+1}$ and $\boldsymbol{\gamma}(k) \in \mathbb{R}^{L+1}$ denotes the constraint vector (CV) of the SM-AP algorithm [2, 3, 5]. The inner structure of the following variables are helpful in further discussions:

$$
\begin{array}{rcllcl}
\mathbf{X}(k) & = & [\mathbf{x}(k) & \mathbf{x}(k-1) & \ldots & \mathbf{x}(k-L)] \\
\mathbf{x}(k) & = & [x(k) & x(k-1) & \ldots & x(k-N)]^T \\
\mathbf{d}(k) & = & [d(k) & d(k-1) & \ldots & d(k-L)]^T \\
\boldsymbol{\gamma}(k) & = & [\gamma_0(k) & \gamma_1(k) & \ldots & \gamma_L(k)]^T \\
\mathbf{e}(k) & = & [e_0(k) & e_1(k) & \ldots & e_L(k)]^T
\end{array}
\tag{1}
$$

For a given coefficient vector $\mathbf{w}(k)$, the SM-AP algorithm aims at obtaining an estimate $\mathbf{w}(k+1) \in \psi_{k-L}^{k}$ for all k, where $\psi_{k-L}^{k} \triangleq \bigcap_{i=k-L}^{k} \mathcal{H}(i)$ is the intersection of the last $L+1$ constraint sets. Hence, if $\mathbf{w}(k)$ already belongs to ψ_{k-L}^{k}, then no update is performed. Otherwise, $\mathbf{w}(k)$ is updated so that $\mathbf{w}(k+1)$ is some point in ψ_{k-L}^{k}. Indeed, when an update occurs, $\mathbf{w}(k+1)$ is generated as the solution to the following optimization problem (OP) [2]:

$$
\min \|\mathbf{w}(k+1) - \mathbf{w}(k)\|^2, \text{ s.t. } \boldsymbol{\varepsilon}(k) = \boldsymbol{\gamma}(k), \tag{2}
$$

where, in order to guarantee $\mathbf{w}(k+1) \in \psi_{k-L}^{k}$, the CV $\boldsymbol{\gamma}(k)$ must satisfy $|\gamma_i(k)| \leq \overline{\gamma}$ for $i = 0, 1, \ldots, L$. The solution to such OP is given by the following recursion [2, 3, 5]:

$$
\mathbf{w}(k+1) = \begin{cases} \mathbf{w}(k) + \mathbf{X}(k)\hat{\mathbf{S}}(k)\,[\mathbf{e}(k) - \boldsymbol{\gamma}(k)] & \text{if } |e_0(k)| > \overline{\gamma}, \\ \mathbf{w}(k) & \text{otherwise,} \end{cases} \tag{3}
$$

where $\hat{\mathbf{S}}(k) \triangleq \left[\mathbf{X}^T(k)\mathbf{X}(k) + \delta\mathbf{I}\right]^{-1} \in \mathbb{R}^{(L+1)\times(L+1)}$ and δ is the regularization factor used to avoid numerical issues.

4. CHOOSING THE CONSTRAINT VECTOR

In the adaptive filtering literature, it is well known that whenever $\mathbf{w}(k)$ is updated so that the *a posteriori* error becomes less than or equal to zero the noise enhancement is significant [3]. Thus, although any $\mathbf{w}(k+1) \in \psi_{k-L}^{k}$ is a valid solution, i.e., it takes the uncertainty caused by noise into consideration, some parts of ψ_{k-L}^{k} should not be used in order to control noise enhancement effects. This explains why the SM-AP algorithm using an arbitrary/random CV, i.e., any CV satisfying $|\gamma_i(k)| \leq \overline{\gamma}$, usually yields high steady-state MSE.

Fig. 2 depicts an example of updating process of the SM-AP algorithm with $L = 1$. In this case, $\psi_{k-1}^{k} = \mathcal{H}(k-1) \bigcap \mathcal{H}(k)$, but only the shaded region within ψ_{k-1}^{k} should be used to keep the noise enhancement controlled. Geometrically, it is easy to see that in order to generate $\mathbf{w}(k+1)$ in the shaded region we must design the CV as a function of the error components. In fact, we can use simply the sign of the error components, denoted by $\text{sign}\,[e_i(k)]$, since with this information we can determine which hyperplanes are closer to $\mathbf{w}(k)$. Definitions 1 to 3 present some choices for $\boldsymbol{\gamma}(k)$ that yield $\mathbf{w}(k+1)$ in the shaded region. These CVs clearly satisfy $|\gamma_i(k)| \leq \overline{\gamma}$, but for the ED-CV we must impose $\overline{\gamma} \leq 1$ (when $\overline{\gamma} = 1$ the ED-CV coincides with FMEB-CV).

Definition 1 (Simple choice (SC)) *The SC constraint vector (SC-CV), proposed in [2] and analyzed in [6], is defined as*

$$
\gamma_i(k) \triangleq \begin{cases} \overline{\gamma}\ \text{sign}\,[e_i(k)] & \textit{if } i = 0, \\ e_i(k) & \textit{for } i = 1, \ldots, L. \end{cases} \tag{4}
$$

Definition 2 (Fixed modulus error-based (FMEB)) *The FMEB constraint vector (FMEB-CV), proposed in [7] and analyzed in [5], is defined as*

$$
\gamma_i(k) \triangleq \overline{\gamma}\ \text{sign}\,[e_i(k)]\,, \textit{ for } i = 0, 1, \ldots, L. \tag{5}
$$

Definition 3 (Exponential decay (ED)) *The ED constraint vector (ED-CV), proposed here, assumes $\overline{\gamma} \leq 1$ and is given by*

$$\gamma_i(k) \triangleq \overline{\gamma}^{\ i+1} \ \operatorname{sign}[e_i(k)] \, , \textit{ for } i = 0, 1, \ldots, L. \tag{6}$$

4.1. Update: Geometrical Viewpoint

As stated in Eq. (2), when an update occurs $\mathbf{w}(k)$ is mapped to the closest point that yields *a posteriori* error vector equal to the CV. Following this reasoning, we can analyze geometrically the updating process related to each of the CVs.

In the SC-CV, $\mathbf{w}(k+1)$ is produced so that the error due to data-pair $(\mathbf{x}(k), d(k))$ is reduced in modulus to $\overline{\gamma}$, whereas the errors due to the past L data-pairs, i.e. $(\mathbf{x}(i), d(i))$ for $i = k-L, \ldots, k-1$, are kept unaltered. Hence, $\mathbf{w}(k+1)$ lies on the closest hyperplane of ψ_{k-L}^{k}. On the contrary, in the FMEB-CV all the errors due to the last $L+1$ data-pairs are modified so that their absolute values become $\overline{\gamma}$. As a result, $\mathbf{w}(k+1)$ is on a vertex belonging to the closest hyperplane of ψ_{k-L}^{k}. In the ED-CV case, $\mathbf{w}(k+1)$ is also on the closest hyperplane of ψ_{k-L}^{k}. Fig. 2 shows an example illustrating the updating process of the SM-AP algorithm for different CVs.

4.2. Update: Analytical Viewpoint

According to Eq. (3), when the SM-AP algorithm updates, $\mathbf{w}(k+1)$ is computed as

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \mathbf{A}(k)[\mathbf{e}(k) - \boldsymbol{\gamma}(k)], \tag{7}$$

where matrix $\mathbf{A}(k) \triangleq \mathbf{X}(k)\hat{\mathbf{S}}(k) \in \mathbb{R}^{(N+1)\times(L+1)}$ is a function of the data. In this subsection, we show that the aforementioned CVs can be interpreted as the weights applied to each of the columns $\mathbf{a}_i(k) \in \mathbb{R}^{N+1}, i = 0, \ldots, L$, of the matrix $\mathbf{A}(k)$. Indeed, for each CV, we can rewrite Eq. (7) as

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \mathbf{p}(k) \tag{8}$$

where $\mathbf{p}(k) \triangleq \mathbf{A}(k)\mathbf{D}(k)\mathbf{e}(k) \in \mathbb{R}^{N+1}$ is the perturbation applied to $\mathbf{w}(k)$ during the update and $\mathbf{D}(k) \in \mathbb{R}^{(L+1)\times(L+1)}$ is a diagonal matrix containing these weights. The equation above also states that the SM-AP algorithm with an adequate CV is nothing more than an AP algorithm featuring *data selection* (innovation check to decide if an update is necessary) and a *variable step-size matrix* $\mathbf{D}(k)$ (whose diagonal entries represent the step-size applied to each vector $\mathbf{a}_i(k)$).

The following equations describe matrix $\mathbf{D}(k)$ and vector $\mathbf{p}(k)$ for the different CVs. For the SC-CV, we have a rank-1 matrix $\mathbf{D}_{\rm SC}(k)$ implying that only the first column $\mathbf{a}_0(k)$ is used in the update, see Eq. (9). On the other hand, matrices $\mathbf{D}_{\rm FMEB}(k)$ and $\mathbf{D}_{\rm ED}(k)$ have full rank and, therefore, yield perturbation vectors as linear combinations of all columns of $\mathbf{A}(k)$, see Eqs. (10) and (11). As a consequence, it is expected the convergence speed of the FMEB-CV and ED-CV to be higher than the one of the SC-CV.

$$\begin{aligned}
\mathbf{D}_{\rm SC}(k) &= \operatorname{diag}\left\{\left(1 - \frac{\overline{\gamma}}{|e_0(k)|}, 0, \ldots, 0\right)\right\} \\
\mathbf{p}_{\rm SC}(k) &= \left(1 - \frac{\overline{\gamma}}{|e_0(k)|}\right) e_0(k)\mathbf{a}_0(k)
\end{aligned} \tag{9}$$

$$\begin{aligned}
\mathbf{D}_{\rm FMEB}(k) &= \operatorname{diag}\left\{\left(1 - \frac{\overline{\gamma}}{|e_0(k)|}, 1 - \frac{\overline{\gamma}}{|e_1(k)|}, \ldots, 1 - \frac{\overline{\gamma}}{|e_L(k)|}\right)\right\} \\
\mathbf{p}_{\rm FMEB}(k) &= \sum_{i=0}^{L}\left(1 - \frac{\overline{\gamma}}{|e_i(k)|}\right) e_i(k)\mathbf{a}_i(k)
\end{aligned} \tag{10}$$

$$\begin{aligned}
\mathbf{D}_{\rm ED}(k) &= \operatorname{diag}\left\{\left(1 - \frac{\overline{\gamma}}{|e_0(k)|}, 1 - \frac{\overline{\gamma}^2}{|e_1(k)|}, \ldots, 1 - \frac{\overline{\gamma}^{L+1}}{|e_L(k)|}\right)\right\} \\
\mathbf{p}_{\rm ED}(k) &= \sum_{i=0}^{L}\left(1 - \frac{\overline{\gamma}^{\ i+1}}{|e_i(k)|}\right) e_i(k)\mathbf{a}_i(k)
\end{aligned} \tag{11}$$

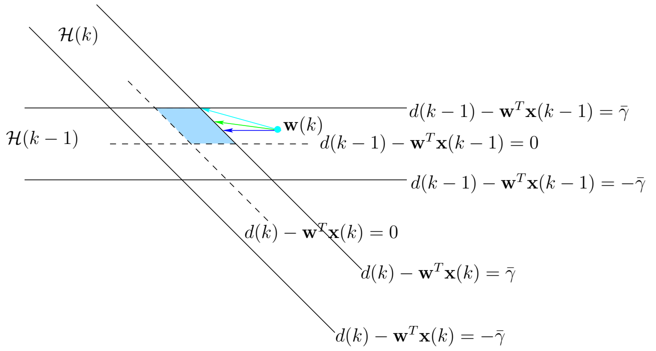


Fig. 2. SM-AP updating scheme in the parameter space for $L = 1$. The blue, cyan, and green arrows depict $\mathbf{w}(k+1)$ considering the SC-CV, FMEB-CV, and ED-CV, respectively. The shaded region represents the subregion of ψ_{k-L}^{k} where noise enhancement is controlled/reduced.

5. RESULTS

In this section, some aspects of the SM-AP algorithm using different choices for the CV are studied. These aspects are: (i) steady-state MSE level in stationary and nonstationary environments, (ii) convergence speed, and (iii) influence of $\overline{\gamma}$ in the steady-state MSE. The CVs considered are the SC-CV, FMEB-CV, ED-CV (see Definitions 1 to 3), and the general-choice CV (GC-CV), which is defined as $\gamma_i(k) \triangleq \overline{\gamma}$ for $i = 0, 1, \ldots, L$.

5.1. Simulation Scenario

We consider the problem of identifying an unknown system [3] whose impulse response is $\mathbf{h}(k) = \mathbf{w}_o$ for all k, where $\mathbf{w}_o \triangleq [0.1 \ \ 0.3 \ \ 0 \ \ -0.2 \ \ -0.4 \ \ -0.7 \ \ -0.4 \ \ -0.2]^T$. When evaluating the nonstationary behavior, which happens only in Fig. 4, the impulse response of the unknown system is given by $\mathbf{h}(k+1) = \lambda_{\mathbf{h}}\mathbf{h}(k) + \mathbf{n}_{\mathbf{h}}(k)$, where $\mathbf{h}(0) = \mathbf{w}_o$, $\lambda_{\mathbf{h}} = 0.99$, and $\mathbf{n}_{\mathbf{h}}(k)$ is white and Gaussian with variance 0.0015.

The input signal is drawn from a standard normal distribution and the noise variance is $\sigma_n^2 = 10^{-2}$. Most of the results were obtained by repeating the experiment 5×10^3 times except for the results in Fig. 6, in which we took an average of the last 10^4 samples from each of the 100 experiments, and then averaged over the experiments, as done in [5]. In addition, the adaptive filter order is $N = 7$, which is the same order of the unknown system, and is initialized with $\mathbf{w}(0) = \mathbf{0}$. The regularization factor is $\delta = 10^{-12}$.

5.2. Simulation Results

Figs. 3 to 5 depict the MSE learning curves for the SM-AP algorithm with different CVs. The standard AP algorithm is used as benchmark. Fig. 6 depicts steady-state excess MSE (EMSE) as a function of $\overline{\gamma}$.

In Figs. 3 and 4, the steady-state MSE in stationary and nonstationary environments is evaluated, respectively. Hence, the algorithms were set so that they have similar convergence speeds in the early iterations. In addition, the parameter $\overline{\gamma}$ is chosen as $\overline{\gamma} = \sqrt{\tau\sigma_n^2}$ with $\tau = 3$, which is a recommended value for both the SC-CV [6] and FMEB-CV [5] in order to achieve a balance between low steady-state MSE and low probability of update. Figs. 3(a) and 4(a) consider $L = 1$ (binormalized version of the algorithms), whereas $L = 4$ in Figs. 3(b) and 4(b).

Comparing Figs.3(a) and 3(b), despite the higher convergence speed and also a bit higher steady-state MSE exhibited by the algorithms using $L = 4$, these figures follow the same pattern. Indeed, the GC-CV led to the worst steady-state MSE, while the FMEB-CV led to a steady-state MSE level similar to the one achieved by the AP algorithm. The SC-CV led to the lowest steady-state MSE level, while the ED-CV reached an intermediate (i.e., between the SC-CV and the FMEB-CV) MSE level. In addition, using Fig.3(a) as example, the SM-AP employing the GC-CV, SC-CV, FMEB-CV, and ED-CV updated only about $45\%, 20\%, 30\%$, and 21% of the iterations, respectively. Fig. 3 illustrates the importance of a proper choice for the CV and also shows that the SM-AP can achieve similar to better results compared to the AP algorithm besides saving computational power.

In Fig. 4 the nonstationary behavior of the SM-AP is assessed. Once again, the GC-CV led to the worst results, while the results obtained using the SC-CV, FMEB-CV, and ED-CV are similar to the one of the AP algorithm. In fact, we observed that the ED-CV was a bit better than the others since it achieved the lowest steady-state MSE and also had the lowest probability of update. In Fig.4(b), e.g., the SM-AP employing the GC-CV, SC-CV, FMEB-CV, and ED-CV updated about $70\%, 62\%, 63\%$, and 60% of the iterations, respectively.

In Fig. 5, the convergence speed is studied. Hence, the algorithms were set so that they reach a similar steady-state MSE level. Fig. 5 shows that, for an arbitrarily given MSE level, the SM-AP with SC-CV was the slowest algorithm but also had the lowest probability of update. The highest convergence speeds were achieved by the SM-AP with FMEB-CV and the AP algorithms, with advantage for the SM-AP algorithm that does not update in all iterations. Interestingly, the convergence speed provided by the ED-CV was very close to the one of the FMEB-CV, but the ED-CV requires less updates. In this particular setup, the SM-AP employing the SC-CV, FMEB-CV, and ED-CV updated about $8\%, 65\%$, and 11% of the iterations, respectively.

Fig. 6 depicts the steady-state excess MSE (EMSE) as a function of τ, a parameter that determines $\overline{\gamma} = \sqrt{\tau\sigma_n^2}$, for different values of L. Observe that when $\tau = 0$, i.e., $\overline{\gamma} = 0$, the SM-AP employing the FMEB-CV and the ED-CV become the standard AP algorithm (i.e., AP with step size equal to 1). In [5], it was shown that by judiciously choosing τ the SM-AP with FMEB-CV could always achieve a steady-state MSE lower than the one of the standard AP algorithm in stationary environments, a fact that is corroborated by Fig. 6(b). Fig. 6(c) indicates that the same result should be valid for the SM-AP with ED-CV, with the advantage that in the ED-CV the range of values of τ leading to low EMSE is wider, which means that it is easier to set the SM-AP with ED-CV. Finally, Figs. 6(a) and 6(c) show that the SM-AP algorithm using the SC-CV and the ED-CV, respectively, can use high values of τ and still achieve low steady-state MSE. Recall that higher τ implies lower probability of update, i.e., more energy/computational saving.

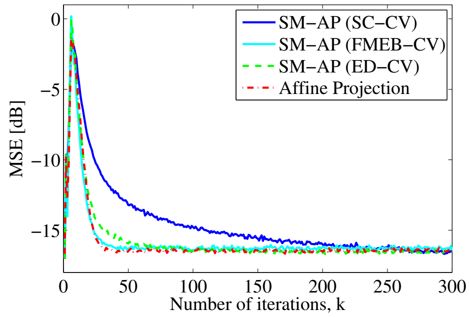
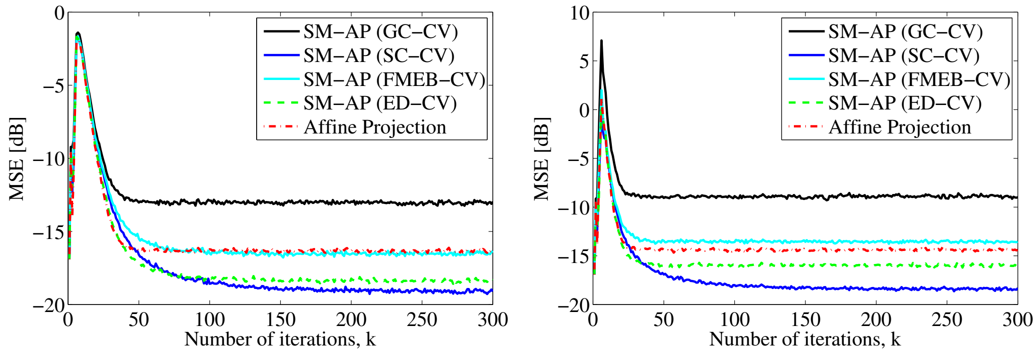


Fig. 5. Learning curve considering $L = 4$. Algorithms were set so that they reach a similar steady-state MSE.

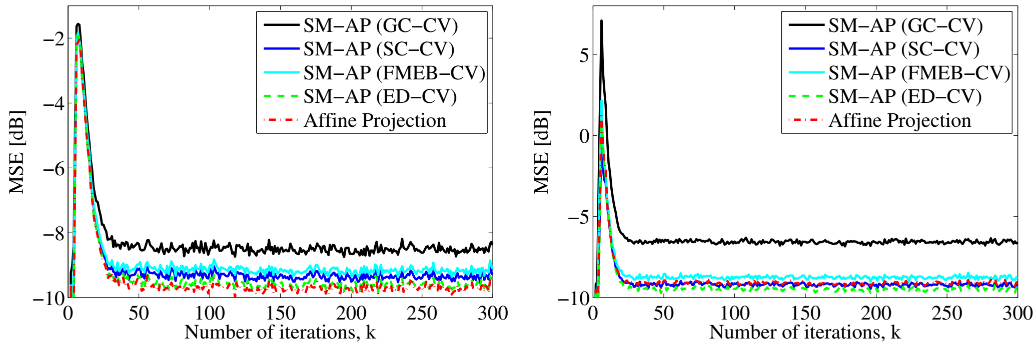
6. CONCLUSIONS

In this paper, we focused on the advantages of set theoretic estimation over the classical point estimation theory having the SMF concept as the underlying framework. The SM-AP algorithm was used since it is one of the most general algorithms for set theoretic estimation, whereas the classical point estimation was represented by the standard AP algorithm. In this context, we emphasized the importance of defining the constraint sets properly and the role played by the constraint vector (CV) of the SM-AP algorithm. Indeed, we explained why an arbitrary choice of CV may lead to poor results and showed a general rule for choosing the CV while controlling the noise enhancement. We discussed two widely used CVs (viz., SC-CV and FMEB-CV), proposed a new one (viz., ED-CV), and interpreted their updating processes from both geometrical and analytical viewpoints. Simulation results show that the convergence speed of the SM-AP with ED-CV is almost as high as the one of the SM-AP with FMEB-CV, since the ED-CV exploits all directions (columns) of matrix $\mathbf{A}(k)$, but its steady-state MSE level is much lower. In addition, the SM-AP with ED-CV was the most efficient in nonstationary environment and also presented a low probability of update, sometimes even lower than for the SM-AP using SC-CV.



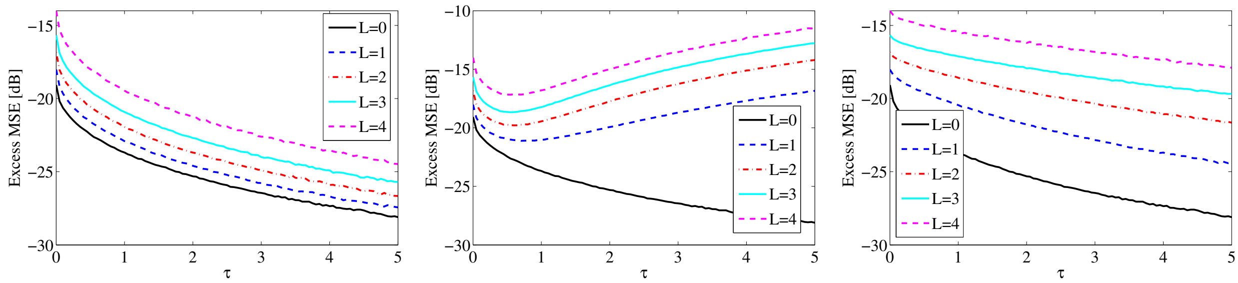
(a) $L = 1$ (b) $L = 4$

Fig. 3. Learning curve considering $\tau = 3$ in $\overline{\gamma} = \sqrt{\tau\sigma_n^2}$. Algorithms were set so that they have similar transient response.



(a) $L = 1$ (b) $L = 4$

Fig. 4. Learning curve in nonstationary environment using $\tau = 3$ in $\overline{\gamma} = \sqrt{\tau\sigma_n^2}$. Algorithms with similar transient response.



(a) SC-CV (b) FMEB-CV (c) ED-CV

Fig. 6. Steady-state Excess MSE vs. τ, where $\overline{\gamma} = \sqrt{\tau\sigma_n^2}$.

7. REFERENCES

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

[2] 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.

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

[4] 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.

[5] 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*, January 2013, DOI: 10.1007/s00034-012-9545-4.

[6] P.S.R. Diniz, "Convergence performance of the simplified set-membership affine projection algorithm," *Circuits, Systems, and Signal Processing*, vol. 30, no. 2, pp. 439–462, April 2011.

[7] M.V.S. Lima and P.S.R. Diniz, "Steady-state analysis of the set-membership affine projection algorithm," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2010)*, Dallas, TX, USA, March 2010, pp. 3802–3805.