Sparsity-Aware Adaptive Algorithms Based on Alternating Optimization and Shrinkage

Rodrigo C. de Lamare, *Senior Member, IEEE*, and Raimundo Sampaio-Neto

***Abstract*—This letter proposes a novel sparsity-aware adaptive filtering scheme and algorithms based on an alternating optimization strategy with shrinkage. The proposed scheme employs a two-stage structure that consists of an alternating optimization of a diagonally-structured matrix that speeds up the convergence and an adaptive filter with a shrinkage function that forces the coefficients with small magnitudes to zero. We devise alternating optimization least-mean square (LMS) algorithms for the proposed scheme and analyze its mean-square error. Simulations for a system identification application show that the proposed scheme and algorithms outperform in convergence and tracking existing sparsity-aware algorithms.**

***Index Terms*—Adaptive filters, iterative methods, sparse signal processing.**

I. INTRODUCTION

IN THE last few years, there has been a growing interest in adaptive algorithms that can exploit the sparsity present in various signals and systems that arise in applications of adaptive signal processing [1]–[10]. The basic idea is to exploit prior knowledge about the sparsity present in the data that need to be processed for applications in system identification, communications and array signal processing. Several algorithms based on the least-mean square (LMS) [1], [2] and the recursive least-squares (RLS) [3], [4], [5], [6] techniques have been reported in the literature along with different penalty or shrinkage functions. These penalty functions perform a regularization that attracts to zero the coefficients of the adaptive filter that are not associated with the weights of interest. With this objective in mind, several penalty functions that account for the sparisty of data signal have been considered, namely: an approximation of the l_0-norm [1], [6], the l_1- norm penalty [2], [5], and the log-sum penalty [2], [5], [8]. These algorithms solve problems with sparse features without relying on the computationally complex oracle algorithm, which requires an exhaustive search for the location of the non-zero coefficients of the system. However, the available algorithms in the literature also exhibit a performance degradation as compared to the oracle algorithm, which might affect the performance of some applications of adaptive algorithms.

Manuscript received November 19, 2013; accepted December 30, 2013. Date of publication January 09, 2014; date of current version January 14, 2014. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Eric Moreau.

R. C. de Lamare is with CETUC-PUC-Rio, 22453-900 Rio de Janeiro, Brazil, and also with the Communications Research Group, Department of Electronics, University of York, York Y010 5DD, U.K.

R. Sampaio-Neto is with CETUC/PUC-RIO, 22453-900 Rio de Janeiro, Brazil (e-mail: rcdl500@york.ac.uk, raimundo@cetuc.puc-rio.br)

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LSP.2014.2298116

Motivated by the limitation of existing sparse adaptive techniques, we propose a novel sparsity-aware adaptive filtering scheme and algorithms based on an alternating optimization strategy with shrinkage. The proposed scheme employs a two-stage structure that consists of an alternating optimization of a diagonally-structured matrix that accelerates the convergence and an adaptive filter with a shrinkage function that attracts the coefficients with small magnitudes to zero. The diagonally-structure matrix aims to perform what the oracle algorithm does and helps to accelerate the convergence of the scheme and improve its steady-state performance. We devise sparsity-aware alternating optimization least-mean square (SA-ALT-LMS) algorithms for the proposed scheme and derive analytical formulas to predict their mean-square error (MSE) upon convergence. Simulations for a system identification application show that the proposed scheme and algorithms outperform in convergence and tracking the state-of-the-art sparsity-aware algorithms.

II. PROBLEM STATEMENT AND THE ORACLE ALGORITHM

In this section, we state the sparse system identification problem and describe the optimal strategy known as as the oracle algorithm, which knows the positions of the non-zero coefficients of the sparse system.

A. Sparse System Identification Problem

In the sparse system identification problem of interest, the system observes a complex-valued signal represented by an $M \times 1$ vector $\boldsymbol{x}[i]$ at time instant i, performs filtering and obtains the output $d[i] = \boldsymbol{w}_o^H \boldsymbol{x}[i]$, where $\boldsymbol{w}_o$ is an M-length finite-impulse-response (FIR) filter that represents the actual system. For system identification, an adaptive filter with M coefficients $\boldsymbol{w}[i]$ is employed in such a way that it observes $\boldsymbol{x}[i]$ and produces an estimate $\hat{d}[i] = \boldsymbol{w}^H[i]\boldsymbol{x}[i]$. The system identification scheme then compares the output of the actual system $d[i]$ and the adaptive filter $\hat{d}[i]$, resulting in an error signal $e[i] = d[i] + n[i] - \hat{d}[i]$, where $n[i]$ is the measurement noise. In this context, the goal of an adaptive algorithm is to identify the system by minimizing the MSE defined by

$$\boldsymbol{w}_o = \arg\min_{\boldsymbol{w}} E[|d[i] + n[i] - \boldsymbol{w}^H[i]\boldsymbol{x}[i]|^2] \quad (1)$$

A key problem in electronic measurement systems which are modeled by sparse adaptive filters, where the number of non-zero coefficients $K \ll M$, is that most adaptive algorithms do not exploit their sparse structure to obtain performance benefits and/or a computational complexity reduction. If an adaptive algorithm can identify and exploit the non-zero coefficients of the system to be identified, then it can obtain performance improvements and a reduction in the computational complexity.

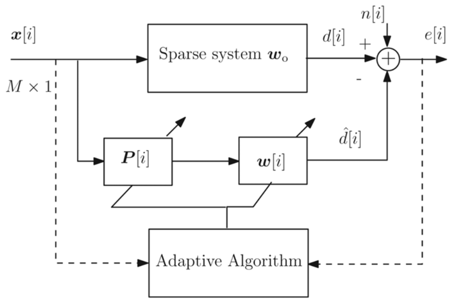

Fig. 1. Proposed adaptive filtering scheme.

B. The Oracle Algorithm

The optimal algorithm for processing sparse signals and systems is known as the oracle algorithm. It can identify the positions of the non-zero coefficients and fully exploit the sparsity of the system under consideration. In the context of sparse system identification and other linear filtering problems, we can state the oracle algorithm as

$$\{\boldsymbol{P}_{\rm or}, \boldsymbol{w}_{\rm or}\} = \arg\min_{\boldsymbol{P},\boldsymbol{w}} E[|d[i] + n[i] - \boldsymbol{w}^H \boldsymbol{P}\boldsymbol{x}[i]|^2] \qquad (2)$$

where $\boldsymbol{P}_{\rm or}$ is an $M \times M$ diagonal matrix with the actual K positions of the non-zero coefficients. It turns out that the oracle algorithm requires an exhaustive search over all the possible K positions over M possibilities, which is an NP-hard problem with extremely high complexity if M is large. Moreover, the oracle algorithm also requires the computation of the optimal filter, which is a continuous optimization problem. For these reasons, it is fundamental to devise low-complexity algorithms that can cost-effectively process sparse signals.

III. Proposed Alternating Optimization with Shrinkage Scheme

In this section, we present an adaptive filtering scheme that employs an alternating optimization strategy with shrinkage that exploits the sparsity in the identification of linear systems. Unlike existing methods, the proposed technique introduces two adaptive filters that are optimized in an alternating fashion, as illustrated in Fig. 1. The first adaptive filter $\boldsymbol{p}[i]$ with M coefficients is applied as a diagonal matrix $\boldsymbol{P}[i] = {\rm diag}(\boldsymbol{p}[i])$ to $\boldsymbol{x}[i]$ and performs the role of the oracle algorithm, which was defined as $\boldsymbol{P}_{\rm or}$ in the previous section. The second adaptive filter $\boldsymbol{w}[i]$ with M coefficients is responsible for the system identification. Both $\boldsymbol{p}[i]$ and $\boldsymbol{w}[i]$ employ l_1-norm shrinkage techniques to attract to zero the coefficients that have small magnitudes. The output of the proposed adaptive filtering scheme is given by

$$\begin{aligned}\hat{d}[i] &= \boldsymbol{w}^H[i] \underbrace{\boldsymbol{P}[i]}_{{\rm diag}(\boldsymbol{p}[i])} \boldsymbol{x}[i] = \boldsymbol{p}^T[i] \underbrace{\boldsymbol{W}^*[i]}_{{\rm diag}(\boldsymbol{w}^*[i])} \boldsymbol{x}[i] \\ &= \boldsymbol{x}^T[i]\boldsymbol{P}[i]\boldsymbol{w}^*[i] = \boldsymbol{x}^T[i]\boldsymbol{W}^*[i]\boldsymbol{p}[i] \qquad (3)\end{aligned}$$

A. Adaptive Algorithms

In order to devise adaptive algorithms for this scheme, we need to cast an optimization problem with a cost function that depends on $\boldsymbol{p}[i]$, $\boldsymbol{w}[i]$ and a shrinkage function $f(\cdot)$, where $f(\boldsymbol{a})$ represents this function applied to a generic parameter vector $\boldsymbol{a}$ with M coefficients. Let us consider the following cost function

$$C(\boldsymbol{p}[i], \boldsymbol{w}[i]) = E[|d[i] - \hat{d}[i]|^2] + \lambda f(\boldsymbol{p}[i])) + \tau f(\boldsymbol{w}[i]), \quad (4)$$

where λ and τ are the regularization terms. In order to derive an adaptive algorithm to minimize the cost function in (4) and perform system identification, we employ an alternating optimization strategy. We compute the instantaneous gradient of (4) with respect to $\boldsymbol{p}[i]$ and $\boldsymbol{w}[i]$ and devise LMS-type algorithms:

$$\begin{aligned}\boldsymbol{p}[i+1] &= \boldsymbol{p}[i] - \eta\lambda \frac{\partial C(\boldsymbol{p}[i], \boldsymbol{w}[i])}{\partial \boldsymbol{p}^*[i]} \\ &= \boldsymbol{p}[i] + \eta e(\boldsymbol{w}[i], \boldsymbol{p}[i])\boldsymbol{W}[i]\boldsymbol{x}^*[i] - \underbrace{\eta\lambda}_{\alpha} \frac{\partial f(\boldsymbol{p}[i])}{\partial \boldsymbol{p}^*[i]},\end{aligned} \qquad (5)$$

$$\begin{aligned}\boldsymbol{w}[i+1] &= \boldsymbol{w}[i] - \mu \frac{\partial C(\boldsymbol{p}[i], \boldsymbol{w}[i])}{\partial \boldsymbol{w}^*[i]} \\ &= \boldsymbol{w}[i] + \mu e(\boldsymbol{w}[i], \boldsymbol{p}[i])^*\boldsymbol{P}[i]\boldsymbol{x}[i] - \underbrace{\mu\tau}_{\gamma} \frac{\partial f(\boldsymbol{w}[i])}{\partial \boldsymbol{w}^*[i]},\end{aligned} \qquad (6)$$

where $e(\boldsymbol{w}[i], \boldsymbol{p}[i]) = d[i] - \boldsymbol{w}^H[i]\boldsymbol{P}[i]\boldsymbol{x}[i]$ is the error signal and μ and η are the step sizes of the LMS recursions, which are used in an alternating way. In Table I, different shrinkage functions are shown with their partial derivatives and other features. A key requirement of the proposed scheme is the initialization which results in the adjustment of $\boldsymbol{p}[i]$ to shrink the coefficients corresponding to zero elements of the system and $\boldsymbol{w}[i]$ to estimate the non-zero coefficients. Specifically, $\boldsymbol{p}[i]$ is initialized as an all-one vector ($\boldsymbol{p}[0] = \boldsymbol{1}$ or $\boldsymbol{P}[0] = \boldsymbol{I}$) and $\boldsymbol{w}[i]$ is initialized as an all-zero vector ($\boldsymbol{w}[0] = \boldsymbol{0}$). When $\boldsymbol{p}[i]$ is fixed, the scheme is equivalent to a standard shrinkage algorithm. The two-step approach outperforms the single-step method since $\mathbf{P}[i]$ strives to perform the role of the Oracle algorithm ($\mathbf{P}_{\rm or}$) by decreasing the values of its entries in the positions of the zero coefficients. This helps the recursion that adapts $\mathbf{w}[i]$ to perform the estimation of the non-zero coefficients. This process is then alternated over the iterations, resulting in better performance. When $\mathbf{P}_{\rm or}$ is employed, $\mathbf{w}[i]$ has the information about the actual positions of the zero coefficients.

B. Computational Complexity

We detail the computational complexity in terms of arithmetic operations of the proposed and some existing algorithms. Specifically, we consider the conventional LMS algorithm, sparsity-aware LMS (SA-LMS) algorithms, and the proposed SA-ALT-LMS algorithm. The details are shown in Table II.

IV. Mean-Square Error Analysis

In this section, we develop an MSE analysis of the proposed SA-ALT-LMS algorithm and devise analytical expressions to describe the transient and steady-state performances. By defining $\boldsymbol{w}_{\rm o}$ as the optimal filter and $\boldsymbol{p}_{\rm o}$ as the oracle vector ($\boldsymbol{P}_{\rm o} = {\rm diag}(\boldsymbol{p}_{\rm o})$) with the non-zero coefficients, we can write

$$\boldsymbol{e}_w = \boldsymbol{w}[i] - \boldsymbol{w}_{\rm o} \text{ and } \boldsymbol{e}_p = \boldsymbol{p}[i] - \boldsymbol{p}_{\rm o}. \qquad (7)$$

The error signal can then be rewritten as

$$\begin{aligned}e(\boldsymbol{w}[i], \boldsymbol{p}[i]) = e_{\rm o} - \boldsymbol{x}^T[i]({\rm diag}(\boldsymbol{e}_{\rm p}^T[i])\boldsymbol{e}_w^*[i] \\ + {\rm diag}(\boldsymbol{e}_p^T[i])\boldsymbol{w}_{\rm o}^* + {\rm diag}(\boldsymbol{p}_{\rm o}^T)\boldsymbol{e}_w^*[i]),\end{aligned} \qquad (8)$$

TABLE I
SHRINKAGE FUNCTIONS

Function	Partial Derivative	$\boldsymbol{L}_{\boldsymbol{a}}$	Cost of Shrinkage (C_s)
$f(\boldsymbol{a}) = \lVert\boldsymbol{a}\rVert_1$	$\frac{\partial f(\boldsymbol{a}[i])}{\partial \boldsymbol{a}^*[i]} = \text{sgn}(\boldsymbol{a}) = \text{sgn}(\Re[\boldsymbol{a}]) + j\text{sgn}(\Im[\boldsymbol{a}])$	$\approx \text{sgn}[\boldsymbol{a}_{\text{opt}}]\text{sgn}[\boldsymbol{a}_{\text{opt}}^H]$	$2M\text{ad} + 4M\text{mult} + 2M\text{div}$
$f(\boldsymbol{a}) = \sum_{m=1}^{M} \log(1 + \vert a_m\vert/\epsilon)$	$\frac{\partial f(\boldsymbol{a}[i])}{\partial \boldsymbol{a}^*[i]} = \frac{\text{sgn}(\Re[\boldsymbol{a}]) + j\text{sgn}(\Im[\boldsymbol{a}])}{1+\epsilon\lVert\boldsymbol{a}\rVert_1}$	$\approx \frac{\text{sgn}[\boldsymbol{a}_{\text{opt}}]}{1+\epsilon\vert\boldsymbol{a}_{\text{opt}}\vert} \frac{\text{sgn}[\boldsymbol{a}_{\text{opt}}^H]}{1+\epsilon\vert\boldsymbol{a}_{\text{opt}}\vert}$	$4M\text{ad} + 7M\text{mult} + 3M\text{div}$
$f(\boldsymbol{a}) = \lVert a\rVert_0$ $\approx \sum_{m=1}^{M}(1 - e^{-\beta\vert a_m\vert})$	$\frac{\partial f(a_m[i])}{\partial a_m^*[i]} = \begin{cases} \beta\big(\text{sgn}(\Re[a_m]) + j\text{sgn}(\Im[a_m])\big) - \beta^2 a_m & \text{if } \vert a_m\vert \leq 1/\beta \\ 0 & \text{elsewhere} \end{cases}$	$\approx \beta^2\text{sgn}[\boldsymbol{a}_{\text{opt}}]\text{sgn}[\boldsymbol{a}_{\text{opt}}^H]$ $-\beta^3\text{sgn}[\boldsymbol{a}_{\text{opt}}]\boldsymbol{a}_{\text{opt}}^H$ $-\beta^3\boldsymbol{a}_{\text{opt}}\text{sgn}[\boldsymbol{a}_{\text{opt}}^H]$ $+\beta^4\boldsymbol{a}_{\text{opt}}\boldsymbol{a}_{\text{opt}}^H$	$3M\text{ad} + 6M\text{mult} + 2M\text{div}$

TABLE II
COMPUTATIONAL COMPLEXITY OF ALGORITHMS

Algorithm	Computational Complexity
LMS	$2M\text{ad} + 2M\text{mult}$
SA-LMS	$2M\text{ad} + 2M\text{mult} + 2C_s$
SA-ALT-LMS	$5M\text{ad} + 7M\text{mult} + 2C_s$

where $e_{\text{o}} = e(\boldsymbol{w}_{\text{o}}, \boldsymbol{p}_o) = d[i] - \boldsymbol{x}^T[i]\text{diag}(\boldsymbol{p}_{\text{o}})^T\boldsymbol{w}_{\text{o}}^*$ is the error signal of the optimal sparse filter. The MSE is written as

$$\begin{aligned} \text{MSE} &= E[\vert e(\boldsymbol{w}[i], \boldsymbol{p}[i])\vert^2] \\ &= E[\vert e_{\text{o}} - \boldsymbol{x}^T[i](\text{diag}(\boldsymbol{e}_{\text{p}}^T[i])\boldsymbol{e}_w^*[i] \\ &\quad + \text{diag}(\boldsymbol{e}_p^T[i])\boldsymbol{w}_{\text{o}}^* + \text{diag}(\boldsymbol{p}_{\text{o}}^T)\boldsymbol{e}_w^*[i])\vert^2] \end{aligned} \tag{9}$$

Using the independence assumption between $\boldsymbol{e}_p[i]$, $\boldsymbol{e}_w[i]$ and $\boldsymbol{x}[i]$, we have:

$$\begin{aligned} \text{MSE} = J_{\min} &+ E[\boldsymbol{x}^H[i]\text{diag}(\boldsymbol{e}_p^H)\boldsymbol{e}_w[i]\boldsymbol{e}_w^H[i]\text{diag}(\boldsymbol{e}_p[i])\boldsymbol{x}[i]] \\ &+ E[\boldsymbol{x}^H[i]\text{diag}(\boldsymbol{e}_p^H)\boldsymbol{w}_{\text{o}}[i]\boldsymbol{w}_{\text{o}}^H[i]\text{diag}(\boldsymbol{e}^p[i])\boldsymbol{x}[i]] \\ &+ E[\boldsymbol{x}^H[i]\text{diag}(\boldsymbol{p}_{\text{o}}^H)\boldsymbol{e}_w[i]\boldsymbol{e}_w^H[i]\text{diag}(\boldsymbol{p}_{\text{o}}[i])\boldsymbol{x}[i]], \end{aligned} \tag{10}$$

where $J_{\min} = E[\vert e(\boldsymbol{w}_o, \boldsymbol{p}_o)\vert^2]$. The expectation of the scalar values that are functions of triple vector products can be rewritten [11] and the MSE expressed by

$$\begin{aligned} \text{MSE} = J_{\min} &+ \text{tr}[\boldsymbol{R}_x(\boldsymbol{K}_w \odot \boldsymbol{K}_p)] \\ &+ \text{tr}[\boldsymbol{R}_x(\boldsymbol{R}_{w_{\text{o}}} \odot \boldsymbol{K}_p)] + \text{tr}[\boldsymbol{R}_x(\boldsymbol{R}_{\text{or}} \odot \boldsymbol{K}_w)], \end{aligned} \tag{11}$$

where $\odot$ is the Hadamard product, $\boldsymbol{R}_x = E[\boldsymbol{x}[i]\boldsymbol{x}^H[i]]$, $\boldsymbol{K}_w = E[\boldsymbol{e}_w[i]\boldsymbol{e}_w^H[i]]$, $\boldsymbol{K}_p = E[\boldsymbol{e}_p[i]\boldsymbol{e}_p^H[i]]$, $\boldsymbol{R}_{w_{\text{o}}} = E[\boldsymbol{w}_{\text{o}}\boldsymbol{w}_{\text{o}}^H]$, and $\boldsymbol{R}_{\text{or}} = E[\boldsymbol{p}_{\text{o}}\boldsymbol{p}_{\text{o}}^H]$. Using (5) and (6) into $\boldsymbol{K}_p$ and $\boldsymbol{K}_w$, we obtain

$$\begin{aligned} \boldsymbol{K}_w[i+1] &= (\boldsymbol{I} - \mu\boldsymbol{R}_{px})\boldsymbol{K}_w[i](\boldsymbol{I} - \mu\boldsymbol{R}_{px}) \\ &\quad + \mu^2\boldsymbol{R}_{px}J_{\text{MSE}}^{(i)}(\boldsymbol{w}_{\text{o}}) + \gamma^2\boldsymbol{L}_w, \end{aligned} \tag{12}$$

$$\begin{aligned} \boldsymbol{K}_p[i+1] &= (\boldsymbol{I} - \eta\boldsymbol{R}_{wx})\boldsymbol{K}_p[i](\boldsymbol{I} - \eta\boldsymbol{R}_{wx}) \\ &\quad + \eta^2\boldsymbol{R}_{wx}J_{\text{MSE}}^{(i)}(\boldsymbol{p}_{\text{o}}) + \alpha^2\boldsymbol{L}_p, \end{aligned} \tag{13}$$

where $J_{\text{MSE}}^{(i)}(\boldsymbol{w}_o) \triangleq E[\vert e(\boldsymbol{w}_o, \boldsymbol{p}[i])\vert^2]$ and $J_{\text{MSE}}^{(i)}(\boldsymbol{p}_o) \triangleq E[\vert e(\boldsymbol{w}[i], \boldsymbol{p}_o)\vert^2]$ appear in (12) and (13). The other quantities are $\boldsymbol{R}_{wx} = E[\boldsymbol{W}[i]\boldsymbol{x}[i]\boldsymbol{x}^H[i]\boldsymbol{W}^H[i]]$, $\boldsymbol{L}_w = E[f'[\boldsymbol{w}[i]]f'^H[\boldsymbol{w}[i]]$, $\boldsymbol{R}_{px} = E[\boldsymbol{P}[i]\boldsymbol{x}[i]\boldsymbol{x}^H[i]\boldsymbol{P}^H[i]]$, $\boldsymbol{L}_p = E[f'[\boldsymbol{p}[i]]f'^H[\boldsymbol{p}[i]]$ and $f'(\cdot)$ is the partial derivative with respect to the variable of the argument. In Table I, we use the variable $\boldsymbol{a}$ that plays the role of $\boldsymbol{p}[i]$ or $\boldsymbol{w}[i]$. We obtained approximations for $\boldsymbol{L}_a = E[f'[\boldsymbol{a}[i]]f'^H[\boldsymbol{a}[i]]$, where $\boldsymbol{a}$ is a generic function, to compute the matrices $\boldsymbol{L}_p$ and $\boldsymbol{L}_w$ for a given shrinkage function as shown in the 3rd column of Table I.

We can express $\boldsymbol{R}_{wx}$ and $\boldsymbol{R}_{px}$ as $\boldsymbol{R}_{wx} = \boldsymbol{R}_x \odot \boldsymbol{R}_w[i]]$ and $\boldsymbol{R}_{px} = \boldsymbol{R}_x \odot \boldsymbol{R}_p[i]$, where $\boldsymbol{R}_w = E[\boldsymbol{w}[i]\boldsymbol{w}^H[i]]$ and $\boldsymbol{R}_p = E[\boldsymbol{p}[i]\boldsymbol{p}^H[i]]$. To simplify the analysis, we assume that the samples of the signal $\boldsymbol{x}[i]$ are uncorrelated, i.e., $\boldsymbol{R}_x = \sigma_x^2\boldsymbol{I}$ with σ_x^2 being the variance. Using the diagonal matrices $\boldsymbol{R}_x = \boldsymbol{\Lambda}_x = \sigma_x^2\boldsymbol{I}$, $\boldsymbol{R}_{px} = \boldsymbol{\Lambda}_{px}[i] = \sigma_x^2\boldsymbol{I} \odot \boldsymbol{R}_p[i]$ and $\boldsymbol{R}_{wx} = \Lambda_{wx}[i] = \sigma_x^2\boldsymbol{I} \odot \boldsymbol{R}_w[i]$, we can write

$$\begin{aligned} \boldsymbol{K}_w[i+1] &= (\boldsymbol{I} - \mu\boldsymbol{\Lambda}_{px}[i])\boldsymbol{K}_w[i](\boldsymbol{I} - \mu\boldsymbol{\Lambda}_{px}[i]) \\ &\quad + \mu^2 J_{\text{MSE}}^{(i)}(\boldsymbol{w}_o)\boldsymbol{\Lambda}_{px}[i] + \gamma^2\boldsymbol{L}_w[i] \end{aligned} \tag{14}$$

$$\begin{aligned} \boldsymbol{K}_p[i+1] &= (\boldsymbol{I} - \eta\boldsymbol{\Lambda}_{wx}[i])\boldsymbol{K}_p[i](\boldsymbol{I} - \eta\boldsymbol{\Lambda}_{wx}[i]) \\ &\quad + \eta^2 J_{\text{MSE}}^{(i)}(\boldsymbol{p}_o)\boldsymbol{\Lambda}_{wx}[i] + \alpha^2\boldsymbol{L}_p[i] \end{aligned} \tag{15}$$

Due to the structure of the above equations, the approximations and the quantities involved, we can decouple them into

$$\begin{aligned} K_w^n[i+1] &= (1 - \mu\lambda_{px}^n[i])K_w^n[i](1 - \mu\lambda_{px}^n[i]) \\ &\quad + \mu^2 J_{\text{MSE}}^{(i)}(\boldsymbol{w}_o)\lambda_{px}^n[i] + \gamma^2 L_w^n[i] \end{aligned} \tag{16}$$

$$\begin{aligned} K_p^n[i+1] &= (1 - \eta\lambda_{wx}^n[i])K_p^n[i](1 - \eta\lambda_{wx}^n[i]) \\ &\quad + \eta^2 J_{\text{MSE}}^{(i)}(\boldsymbol{p}_o)\lambda_{wx}[i] + \alpha^2 L_p^n[i] \end{aligned} \tag{17}$$

where $K_w^n[i]$ and $K_p^n[i]$ are the nth elements of the main diagonals of $\boldsymbol{K}_w[i]$ and $\boldsymbol{K}_p[i]$, respectively. By taking $\lim_{i\to\infty} K_w^n[i+1]$ and $\lim_{i\to\infty} K_p^n[i+1]$, we obtain

$$K_w^n = \frac{J(\boldsymbol{w}_o)}{(2/\mu - \lambda_{px}^n)} + \frac{\gamma^2 L_w^n}{\mu^2\lambda_{px}^n(2/\mu - \lambda_{px}^n)} \tag{18}$$

$$K_p^n = \frac{J(\boldsymbol{p}_o)}{(2/\eta - \lambda_{wx}^n)} + \frac{\alpha^2 L_p^n}{\eta^2\lambda_{wx}^n(2/\eta - \lambda_{wx}^n)}, \tag{19}$$

where $J(\boldsymbol{w}_o) = \lim_{i\to\infty} J_{\text{MSE}}^{(i)}(\boldsymbol{w}_o)$ and $J(\boldsymbol{p}_o) = \lim_{i\to\infty} J_{\text{MSE}}^{(i)}(\boldsymbol{p}_o)$. For stability, we must have $\vert 1 - \mu\lambda_x^n\vert < 1$ and $\vert 1 - \eta d_x^n\vert < 1$, which results in

$$0 < \mu < 2/\max_n[\lambda_{px}^n] \text{ and } 0 < \eta < 2/\max_n[\lambda_{wx}^n], \tag{20}$$

where $\lambda_{px}^n = \lim_{i\to\infty}\sigma_x^2 E[\vert p^n[i]\vert^2]$, $\lambda_{wx}^n = \lim_{i\to\infty}\sigma_x^2 E[\vert w^n[i]\vert^2]$, with $p^n[i]$ and $w^n[i]$ being the nth elements of $\boldsymbol{p}[i]$ and $\boldsymbol{w}[i]$, respectively. The MSE is then given by

$$\begin{aligned} \text{MSE} = J_{\min} &+ \sigma_x^2\sum_{n=1}^{M} K_p^n M_w^n \\ &+ \sigma_x^2\sum_{n=1}^{M} p_o^n\vert w_o^n\vert^2 K_p^n + \sigma_x^2\sum_{n=1}^{M} p_o^n K_w^n, \end{aligned} \tag{21}$$

where w_o^n and p_o^n are the elements of $\boldsymbol{w}_{\rm o}$ and $\boldsymbol{p}_{\rm o}$, respectively. This MSE analysis is valid for uncorrelated input data, whereas a model for correlated input data remains an open problem which is highly involved due to the triple products in (11). However, the SA-ALT-LMS algorithms work very well for both correlated and uncorrelated input data.

V. Simulations

In this section, we assess the performance of the existing LMS, SA-LMS, and the proposed SA-ALT-LMS algorithms with different shrinkage functions. The shrinkage functions considered are the ones shown in Table II, which give rise to the SA-LMS with the l_1-norm [2], the SA-LMS with the log-sum penalty [2], [5], [8] and the l_0-norm [1], [6]. We consider system identification examples with both time-invariant and time-varying parameters in which there is a sparse system with a significant number of zeros to be identified. The input signal $\boldsymbol{x}[i]$ and the noise $\boldsymbol{n}[i]$ are drawn from independent and identically distributed complex Gaussian random variables with zero mean and variances σ_x^2 and σ_n^2, respectively, resulting in a signal-to-noise ratio (SNR) given by SNR $= \sigma_x^2/\sigma_n^2$. The filters are initialized as $\boldsymbol{p}[0] = \mathbf{1}$ and $\boldsymbol{w}[0] = \mathbf{0}$. In the first experiment, there are $N = 16$ coefficients in a time-invariant system, only $K = 2$ coefficients are non-zero when the algorithms start and the input signal is applied to a first-order auto-regressive filter which results in correlated samples obtained by $x_c[i] = 0.8x_c[i-1] + x[i]$ that are normalized. After 1000 iterations, the sparse system is suddenly changed to a system with $N = 16$ coefficients but in which $K = 4$ coefficients are non-zero. The positions of the non-zero coefficients are chosen randomly for each independent simulation trial. The curves are averaged over 200 independent trials and the parameters are optimized for each example. We consider the log-sum penalty [2], [5], [8] and the l_0-norm [1], [6] because they have shown the best performances.

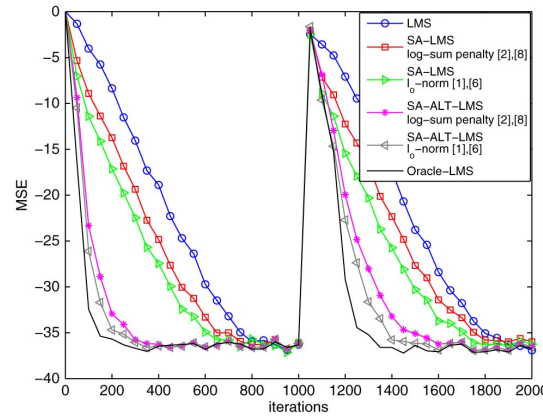

Fig. 2. MSE performance against number of iterations for correlated input data. Parameters: SNR $= 40$ dB, $\sigma_x^2 = 1$, $\mu = 0.015$, $\eta = 0.012$, $\tau = 0.02$, $\lambda = 0.02$, $\epsilon = 10$, and $\beta = 10$.

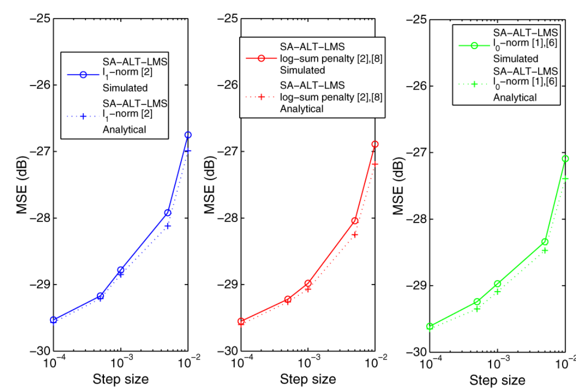

Fig. 3. MSE performance against step size for $\mu = \eta$. Parameters: SNR $= 30$ dB, $\sigma_x^2 = 1$, $\tau = 0.02$, $\lambda = 0.02$, $\epsilon = 10$, and $\beta = 10$.

The results of the first experiment are shown in Fig. 2, where the existing LMS and SA-LMS algorithms are compared with the proposed SA-ALT-LMS algorithm. The curves show that that MSE performance of the proposed SA-ALT-LMS algorithms is significantly superior to the existing LMS and SA-LMS algorithms for the identification of sparse system. The SA-ALT-LMS algorithms can approach the performance of the Oracle-LMS algorithm, which has full knowledge about the positions of the non-zero coefficients. A performance close to the Oracle-LMS algorithm was verified for various situations of interest including different values of SNR, degrees of sparsity (K) and for both small and large sparse systems $(10 \le N \le 200)$.

In a second experiment, we have assessed the validity of the MSE analysis and the formulas obtained to predict the MSE as indicated in (21) and in Table II for uncorrelated input data. In the evaluation of (18) and (19), we made the following approximations $J(\boldsymbol{w}_o) \approx J(\boldsymbol{p}_o) \approx J_{\min}$, $\lambda_{px}^n \approx \sigma_x^2 p_o^n$ and $\lambda_{wx}^n \approx \sigma_x^2 w_o^n$. We have considered a scenario where the input signal and the observed noise are white Gaussian random sequences with variance of 1 and 10^{-3}, respectively, i.e., SNR $= 30$ dB. There are $N = 32$ coefficients in a time-invariant system that are randomly generated and only $K = 4$ coefficients are non-zero. The positions of the non-zero coefficients are again chosen randomly for each independent simulation trial. The curves are averaged over 200 independent trials and the algorithms operate for 1000 iterations in order to ensure their convergence. We have compared the simulated curves obtained with the SA-ALT-LMS strategy using the l_1-norm [2], the SA-LMS with the log-sum penalty [2], [5], [8] and the l_0-norm [1], [6]. The results in Fig. 3 indicate that there is a close match between the simulated and the analytical curves for the shrinkage functions employed, suggesting that the formulas obtained and the simplifications made are valid and resulted in accurate methods to predict the MSE performance of the proposed SA-ALT-LMS algorithms.

VI. Conclusion

We have proposed a novel sparsity-aware adaptive filtering scheme and algorithms based on an alternating optimization strategy that is general and can operate with different shrinkage functions. We have devised alternating optimization LMS algorithms, termed as SA-ALT-LMS for the proposed scheme and developed an MSE analysis, which resulted in analytical formulas that can predict the performance of the SA-ALT-LMS algorithms. Simulations for a system identification application show that the proposed scheme and SA-ALT-LMS algorithms outperform existing sparsity-aware algorithms.

REFERENCES

[1] Y. Gu, J. Jin, and S. Mei, "L_0 norm constraint LMS algorithm for sparse system identification," *IEEE Signal Process. Lett.*, vol. 16, pp. 774–777, 2009.

[2] Y. Chen, Y. Gu, and A. O. Hero, "Sparse LMS for system identification," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing*, Apr. 19-24, 2009, pp. 3125–3128.

[3] B. Babadi, N. Kalouptsidis, and V. Tarokh, "SPARLS: The sparse RLS algorithm," *IEEE Trans. Signal Process.*, vol. 58, no. 8, pp. 4013–4025, 2010.

[4] D. Angelosante, J. A. Bazerque, and G. B. Giannakis, "Online adaptive estimation of sparse signals: Where RLS meets the l_1-norm," *IEEE Trans. Signal Process.*, vol. 58, no. 7, pp. 3436–3447, 2010.

[5] E. M. Eksioglu, "Sparsity regularized RLS adaptive filtering," *IET Signal Process.*, vol. 5, no. 5, pp. 480–487, Aug. 2011.

[6] E. M. Eksioglu and A. L. Tanc, "RLS algorithm with convex regularization," *IEEE Signal Process. Lett.*, vol. 18, no. 8, pp. 470–473, Aug. 2011.

[7] N. Kalouptsidis, G. Mileounis, B. Babadi, and V. Tarokh, "Adaptive algorithms for sparse system identification," *Signal Process.*, vol. 91, no. 8, pp. 1910–1919, Aug. 2011.

[8] E. J. Candes, M. Wakin, and S. Boyd, "Enhancing sparsity by reweighted l1 minimization," *J. Fourier Anal. Applicat.*, 2008.

[9] R. C. de Lamare and R. Sampaio-Neto, "Adaptive reduced-rank MMSE filtering with interpolated FIR filters and adaptive interpolators," *IEEE Signal Process. Lett.*, vol. 12, no. 3, Mar. 2005.

[10] R. C. de Lamare and R. Sampaio-Neto, "Adaptive reduced-rank processing based on joint and iterative interpolation, decimation, and filtering," *IEEE Trans. Signal Process.*, vol. 57, no. 7, pp. 2503–2514, Jul. 2009.

[11] S. Haykin, *Adaptive Filter Theory*, 4th ed. ed. Upper Saddle River, NJ, USA: Prentice-Hall, 2002.