Feature LMS Algorithm for Bandpass System Models

Paulo S. R. Diniz
Department of Electrical and Electronic Engineering
Federal University of Rio de Janeiro
Rio de Janeiro, Brazil
diniz@smt.ufrj.br

Hamed Yazdanpanah
Department of Computer Science
Institute of Mathematics and Statistics
University of São Paulo (USP)
São Paulo, Brazil
hamed.yazdanpanah@smt.ufrj.br

Markus V. S. Lima
Department of Electrical and Electronic Engineering
Federal University of Rio de Janeiro
Rio de Janeiro, Brazil
markus.lima@smt.ufrj.br

***Abstract*—Sparse representations of model parameters have been widely studied. In the adaptive filtering literature, most studies address the cases where the sparsity is directly observed, therefore, there is a growing interest in developing strategies to exploit hidden sparsity. Recently, the feature LMS (F-LMS) algorithm was proposed to expose the sparsity of models with low- and high-frequency contents. In this paper, the F-LMS algorithm is extended to expose hidden sparsity related to models with bandpass spectrum, including the cases of narrowband and broader passband sources. Some simulation results show that the proposed approaches lead to F-LMS algorithms with fast convergence, low misadjustment after convergence, and low computational cost.**

***Index Terms*—adaptive filtering, LMS algorithm, feature matrix, bandpass system, narrowband system**

I. Introduction

In classical adaptive filtering area the least-mean-square (LMS) algorithm has been by far the most widely used due to its simplicity and known properties [1], [2], [3]. Moreover, in the last decades, many variants of the LMS algorithm have been proposed [3]–[7]. The original LMS and many alternative algorithms do not exploit any sparsity inherent to problem at hand.

The importance of sparse representations has been recognized by the scientific community in the last two decades [8], [9]. The main benefit is the parsimonious use of parameters composing mathematical models of signal sources, in other words, one seeks the sparsest representation for a given signal. Sparsity is usually revealed by the number of zero entries in the set of parameters, see for example [10]–[15]. However, in many practical cases, the potential sparsity is not directly observed in the set of parameters. The feature LMS (F-LMS) algorithm has been proposed to exploit sparsity in models representing sources with low- or high-frequency spectrum [16], [17].

In this paper, the concept of feature adaptive filtering is extended to system models with bandpass spectrum. The simplest feature matrix reveals the sparsity of very narrowband sources, whereas a generalization of this approach can be applied to broader passband sources. Simulations demonstrate that the proposed F-LMS algorithms of bandpass sources lead to faster convergence and lower misadjustment after convergence, while reducing the computational cost. The ideas presented here can also be applied to more sophisticated algorithms exploiting sparsity such as those in [18] and [19].

This paper is organized as follows. Section II revisits the F-LMS family of algorithms. In Section III, the derivation of the feature matrix for narrow bandpass spectrum is presented. The following section introduces a strategy to generate feature matrices to exploit sparsity on a broader class of bandpass systems. Section V analyzes the convergence behavior of the coefficient vector of the proposed algorithm. Section VI presents some simulations to confirm the improved performance brought about by the proposed F-LMS algorithms. The conclusions are given in Section VII.

Notation: Scalars are represented by lower-case letters. Vectors (matrices) are denoted by lowercase (uppercase) boldface letters. For a given iteration k, the weight vector and the input vector are denoted by $\mathbf{w}(k), \mathbf{x}(k) \in \mathbb{R}^{N+1}$, respectively, where N is the adaptive filter order. The error signal at the k-th iteration is defined as $e(k) \triangleq d(k) - \mathbf{w}^T(k)\mathbf{x}(k)$, where $d(k) \in \mathbb{R}$ is the desired signal. The l_1-norm of a vector $\mathbf{w} \in \mathbb{R}^{N+1}$ is given by $\|\mathbf{w}\|_1 = \sum_{i=0}^{N} |w_i|$.

II. The Feature LMS algorithms

The feature LMS (F-LMS) algorithm had been previously proposed in [16] with the scope of exploiting some features of the unknown parameters. The F-LMS algorithm minimizes the objective function below

$$\xi_{\text{F-LMS}}(k) = \underbrace{\frac{1}{2}|e(k)|^2}_{\text{standard LMS term}} + \underbrace{\alpha\mathcal{P}\left(\mathbf{F}(k)\mathbf{w}(k)\right)}_{\text{feature-inducing term}}, \tag{1}$$

where $\alpha \in \mathbb{R}_+$ is the weight given to the *sparsity-promoting penalty function* $\mathcal{P}$, which maps a vector to the nonnegative reals $\mathbb{R}_+$. Moreover, $\mathbf{F}(k)$ is the so-called *feature matrix*, which exposes the hidden sparsity. Indeed, the outcome of

This study was financed in part by the São Paulo Research Foundation (FAPESP) grants #2015/22308-2, #2019/06280-1, the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) Finance Code 001, the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), and the Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ).

978-9-0827-9703-9/19/$31.00 ©2019 IEEE

multiplying $\mathbf{w}(k)$ by $\mathbf{F}(k)$ should be a sparse vector, i.e., most entries of the vector $\mathbf{F}(k)\mathbf{w}(k)$ must be close or equal to zero.

By utilizing the stochastic gradient descent method for the objective function (1), the recursion rule of the F-LMS algorithm can be characterized as [16]

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \mu e(k)\mathbf{x}(k) - \mu\alpha\mathbf{p}(k), \tag{2}$$

where μ is the step-size and should be adopted small enough to guarantee the convergence. Furthermore, $\mathbf{p}(k) \in \mathbb{R}^{N+1}$ is the gradient of function $\mathcal{P}\left(\mathbf{F}(k)\mathbf{w}(k)\right)$ with respect to $\mathbf{w}(k)$.

In general, $\mathbf{F}(k)$ can be any time-varying matrix so that $\mathbf{F}(k)\mathbf{w}(k)$ generates a sparse vector, and it is selected based on some *a priori* information about the unknown system $\mathbf{w}_*$. In [16], the authors have introduced the F-LMS algorithm to exploit hidden sparsity in systems containing lowpass and highpass frequency spectrum. However, the F-LMS algorithm can be used to expose sparsity in more general systems. The family of F-LMS algorithms, depending on the adopted $\mathbf{F}(k)$, can exploit different features in unknown systems. In the next section, we define a particular feature matrix $\mathbf{F}(k)$ for the F-LMS algorithm so that it can exploit the hidden sparsity in filters with narrow passband systems to improve the convergence rate and the mean-squared error (MSE) of the learning process.

III. The F-LMS algorithm for narrow passband systems

In this section, we propose the F-LMS algorithm for band-pass systems with narrow passband. Consider an unknown system $\mathbf{w}_*$ with an extremely narrow passband spectrum around the frequency $\omega_c \in [0, \pi)$ and small energy at all other frequencies. Our goal is to design the feature matrix $\mathbf{F}(k)$ such that the multiplication of $\mathbf{F}(k)$ and the impulse response of the unknown system results in a sparse vector.

To this end, we define the transfer function $F(z)$ as

$$\begin{aligned} F(z) \triangleq& (z - e^{-j\omega_c})(z - e^{j\omega_c}) = z^2 - 2\cos(\omega_c)z + 1 \\ =& [1 \;\; -2\cos(\omega_c) \;\; 1][z^2 \;\; z \;\; 1]^T, \end{aligned} \tag{3}$$

which has zeros at $e^{\pm j\omega_c}$. Then we can introduce $W_*^s(z)$ as

$$W_*^s(z) \triangleq F(z)W_*(z), \tag{4}$$

where $W_*(z)$ is the z transform of the unknown system $\mathbf{w}_*$. Since the transfer function $F(z)$ rejects the frequency ω_c and $W_*(z)$ attenuates all other frequencies different from ω_c, the impulse response of $W_*^s(z)$ will be a sparse vector. Therefore, inspired by the transfer function $F(z)$, the feature matrix can be adopted as the time-invariant matrix $\mathbf{F} \in \mathbb{R}^{(N-1)\times(N+1)}$,

$$\mathbf{F} \triangleq \begin{bmatrix} 1 & p_c & 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & p_c & 1 & 0 & \cdots & 0 \\ \vdots & \cdots & \ddots & \ddots & \ddots & \cdots & \vdots \\ 0 & \cdots & 0 & 1 & p_c & 1 & 0 \\ 0 & \cdots & \cdots & 0 & 1 & p_c & 1 \end{bmatrix}, \tag{5}$$

where $p_c = -2\cos(\omega_c)$.

Considering the system $\mathbf{w}_*$, the vector $\mathbf{F}\mathbf{w}_*$ would be a sparse vector. Thus, by utilizing $\mathbf{F}$ as the feature matrix in the objective function (1), we can exploit this sparsity with the help of the sparsity-promoting penalty function $\mathcal{P}$. Some candidates for $\mathcal{P}$ are the l_1-norm [20]–[23], the l_0-norm [17], [24], [25], the discard function [26], [27], etc. In this work, we consider the l_1-norm as the sparsity-promoting penalty function due to its simplicity. Hence, the optimization criterion (1) reduces to

$$\xi(k) = \frac{1}{2}|e(k)|^2 + \alpha\|\mathbf{F}\mathbf{w}(k)\|_1. \tag{6}$$

After using the stochastic gradient descent approach, the update equation of the F-LMS algorithm for bandpass systems with narrow passband is given by (2) in which $\mathbf{p}(k)$ is replaced by $\mathbf{p}_c(k)$ defined as

$$p_{c_i}(k) = \begin{cases} \mathrm{sgn}(w_0(k) + p_c w_1(k) + w_2(k)) \\ \qquad\qquad\qquad \text{if } i = 0, \\ p_c\mathrm{sgn}(w_0(k) + p_c w_1(k) + w_2(k)) \\ +\mathrm{sgn}(w_1(k) + p_c w_2(k) + w_3(k)) \\ \qquad\qquad\qquad \text{if } i = 1, \\ \mathrm{sgn}(w_{i-2}(k) + p_c w_{i-1}(k) + w_i(k)) \\ +p_c\mathrm{sgn}(w_{i-1}(k) + p_c w_i(k) + w_{i+1}(k)) \\ +\mathrm{sgn}(w_i(k) + p_c w_{i+1}(k) + w_{i+2}(k)) \\ \qquad\qquad\qquad \text{if } i = 2, \cdots, N-2, \\ \mathrm{sgn}(w_{N-3}(k) + p_c w_{N-2}(k) + w_{N-1}(k)) \\ +p_c\mathrm{sgn}(w_{N-2}(k) + p_c w_{N-1}(k) + w_N(k)) \\ \qquad\qquad\qquad \text{if } i = N-1, \\ \mathrm{sgn}(w_{N-2}(k) + p_c w_{N-1}(k) + w_N(k)) \\ \qquad\qquad\qquad \text{if } i = N, \end{cases} \tag{7}$$

where $\mathrm{sgn}(\cdot)$ stands for the sign function.

IV. The F-LMS algorithm for bandpass systems

In the previous section, we proposed the F-LMS algorithm for bandpass systems with narrow passband. In this section, we extend the idea for general bandpass systems. For a bandpass system, $\mathbf{w}_*$, with lower and upper cut-off frequencies at ω_l and ω_u, respectively, the main idea is to apply the transfer function $F(z)$ to $W_*(z)$ so that the impulse response of $F(z)W_*(z)$ becomes a sparse vector. In order to construct $F(z)$, we can cascade different transfer functions $F_1(z), \cdots, F_m(z)$, where $F_t(z)$ has zeros at $e^{\pm j\omega_t}$ for $t = 1, \cdots, m$ and $\omega_l < \omega_1 < \cdots < \omega_m < \omega_u$. Therefore, $F(z)$ can be given by

$$\begin{aligned} F(z) \triangleq& F_1(z)\cdots F_m(z) \\ =& (z - e^{-j\omega_1})(z - e^{j\omega_1})\cdots(z - e^{-j\omega_m})(z - e^{j\omega_m}) \\ =& \Big(z^2 - 2\cos(\omega_1)z + 1\Big)\cdots\Big(z^2 - 2\cos(\omega_m)z + 1\Big) \\ =& [1 \;\rho_1 \;\cdots\rho_{m-1} \;\rho_m \;\rho_{m-1} \;\cdots \;\rho_1 \;1][z^{2m} \;z^{2m-1} \;\cdots \;z \;1]^T, \end{aligned} \tag{8}$$

where the last line is the vectorial representation of its previous line.

The transfer function $F(z)$ has zeros at $e^{\pm j\omega_t}$, thus it rejects the frequencies ω_t in the passband region $[\omega_l, \omega_u]$, for $t = 1, \cdots, m$. Therefore, based on the transfer function $F(z)$, we can introduce the time-invariant feature matrix $\mathbf{F}$ as a toeplitz matrix of dimension $(N+1-2m) \times (N+1)$ whose first row and first column are given by $[1\ \rho_1\ \cdots\ \rho_{m-1}\ \rho_m\ \rho_{m-1}\ \cdots\ \rho_1\ 1\ 0\ \cdots\ 0]$ and $[1\ 0\ \cdots\ 0]^T$, respectively. Since the impulse response of $F(z)W_*(z)$ can be represented by a sparse vector, the application of the feature matrix $\mathbf{F}$ to the bandpass system model $\mathbf{w}_*$ would generate a sparse vector. As a result, this sparsity can be exploited by the l_1-norm, as in the previous section. Finally, the recursion rule of the F-LMS algorithm for bandpass systems is given by (2), where $\mathbf{p}(k)$ is the gradient of $\|\mathbf{F}\mathbf{w}(k)\|_1$ with respect to $\mathbf{w}(k)$.

Remark 1: For a bandpass system, $\mathbf{w}_*$, with the passband range of frequencies $[\omega_l, \omega_u] \subset [0, \pi)$, empirically, we require $[\frac{|\omega_u - \omega_l|}{0.07\pi}] + 1$ complex conjugate zeros for the transfer function $F(z)$ enabling that the feature matrix $\mathbf{F}$ reveals the hidden sparsity of $\mathbf{w}_*$, where $[x]$ denotes the integer part of x. In other words, for each subinterval of length 0.07π of $[\omega_l, \omega_u]$, we need to consider a complex conjugate zero for the transfer function $F(z)$.

V. Convergence Behavior of the Coefficient Vector

Assume that $\widetilde{\mathbf{w}}(k) \triangleq \mathbf{w}(k) - \mathbf{w}_*$ denotes the difference between the adaptive filter coefficients and the optimum solution $\mathbf{w}_*$. Also, we know that $e(k) = d(k) - y(k) = \mathbf{w}_*^T\mathbf{x}(k) + n(k) - \mathbf{w}^T(k)\mathbf{x}(k)$ thus, using the recursion rule (2), we get

$$\begin{aligned}\widetilde{\mathbf{w}}(k+1) =& \widetilde{\mathbf{w}}(k) + \mu e(k)\mathbf{x}(k) - \mu\alpha\mathbf{p}(k) \\ =& \widetilde{\mathbf{w}}(k) + \mu\mathbf{x}(k)(n(k) - \mathbf{x}^T(k)\widetilde{\mathbf{w}}(k)) - \mu\alpha\mathbf{p}(k) \\ =& (\mathbf{I} - \mu\mathbf{x}(k)\mathbf{x}^T(k))\widetilde{\mathbf{w}}(k) + \mu\mathbf{x}(k)n(k) - \mu\alpha\mathbf{p}(k),\end{aligned} \tag{9}$$

where $\mathbf{I}$ is the identity matrix. Suppose that $n(k)$ is a zero-mean random variable, and $\mathbf{x}(k)$, $n(k)$, $\widetilde{\mathbf{w}}(k)$ are independent, then by taking expectations on both sides of the above equation, we obtain

$$\mathbb{E}[\widetilde{\mathbf{w}}(k+1)] = (\mathbf{I} - \mu\mathbf{R})\mathbb{E}[\widetilde{\mathbf{w}}(k)] - \mu\alpha\mathbb{E}[\mathbf{p}(k)], \tag{10}$$

where $\mathbf{R} \triangleq \mathbb{E}[\mathbf{x}(k)\mathbf{x}^T(k)]$ is the autocorrelation matrix. Assuming the number of cascaded transfer functions to construct the feature matrix $\mathbf{F}$ is m, we conclude that the coefficients of $\mathbf{p}(k)$ are the sum of at most $2m+1$ sign functions, thus the vector $\mu\alpha\mathbb{E}[\mathbf{p}(k)]$ is bounded. Indeed, we have

$$-\mu\alpha\rho_{\text{max}}(2m+1)\mathbf{1} \leq \mu\alpha\mathbb{E}[\mathbf{p}(k)] \leq \mu\alpha\rho_{\text{max}}(2m+1)\mathbf{1}, \tag{11}$$

where $\mathbf{1} = [1\ 1\ \cdots\ 1]^T$ and $\rho_{\text{max}} = \max\{1, |\rho_1|, \cdots, |\rho_m|\}$. Hence, $\mathbb{E}[\widetilde{\mathbf{w}}(k+1)]$ converges if $0 < \mu < \frac{2}{\lambda_{\max}}$, where $\lambda_{\max}$ is the largest eigenvalue of $\mathbf{R}$. Therefore, we conclude that, for $0 < \mu < \frac{2}{\lambda_{\max}}$,

$$\mathbb{E}[\mathbf{w}(\infty)] = \mathbf{w}_* - \alpha\mathbf{R}^{-1}\mathbb{E}[\mathbf{p}(\infty)]. \tag{12}$$

TABLE I
The specifications of the tested bandpass systems

System	LCF	UCF	LTF	UTF
$\mathbf{w}_*$	$\frac{\pi}{4} - 0.05$	$\frac{\pi}{4} + 0.05$	$\frac{\pi}{4} - 0.15$	$\frac{\pi}{4} + 0.15$
$\mathbf{w}'_*$	$\frac{\pi}{4} - 0.2$	$\frac{\pi}{4} + 0.2$	$\frac{\pi}{4} - 0.3$	$\frac{\pi}{4} + 0.3$
$\mathbf{w}''_*$	$\frac{\pi}{3} - 0.1\pi$	$\frac{\pi}{3} + 0.1\pi$	$\frac{\pi}{3} - 0.45$	$\frac{\pi}{3} + 0.45$

VI. Simulations

In this section, we utilize the LMS and the F-LMS algorithms to identify some unknown bandpass filters. All unknown systems have order 199, i.e., they contain 200 coefficients. The first bandpass system, $\mathbf{w}_*$, has a narrow passband frequency. The second bandpass system, $\mathbf{w}'_*$, has a wider passband frequency, and the third bandpass system, $\mathbf{w}''_*$, has the widest passband frequency among the tested systems. The lower cut-off frequency (LCF), the upper cut-off frequency (UCF), the lower transition frequency (LTF), and the upper transition frequency (UTF) of the tested systems are listed in Table I.

The adopted input signal is a zero-mean white Gaussian noise with unit variance. The signal-to-noise ratio (SNR) is 20 dB, and the initial vector is chosen as $\mathbf{w}(0) = [0, \cdots, 0]^T$. The weight given to the sparsity-promoting penalty function is selected as $\alpha = 0.01$. The values of the step-size μ are reported later for each simulated scenario. The MSE learning curves of the LMS and the F-LMS algorithms are computed by averaging the outcomes of 1000 independent trials.

The central frequency of $\mathbf{w}_*$ is $\frac{\pi}{4}$, and the length of the passband frequency is less than 0.07π; thus, we consider the complex conjugate zeros $e^{\pm j\frac{\pi}{4}}$ for the transfer function $F(z)$. Therefore, the feature matrix $\mathbf{F}$ is given by (5) in which $p_c = -2\cos(\frac{\pi}{4})$. The magnitude response of the filter represented by $\mathbf{w}_*$ is depicted in Figure 1(a). Furthermore, Figure 1(b) shows the impulse response of $\mathbf{F}\mathbf{w}_*$ in which $\mathbf{F}$ is defined by (5); we can observe that $\mathbf{F}\mathbf{w}_*$ is a sparse vector. In other words, the feature matrix $\mathbf{F}$ reveals the hidden sparsity of the bandpass system $\mathbf{w}_*$. Figure 1(c) illustrates the MSE learning curves of the LMS and the F-LMS algorithms considering $\mathbf{w}_*$. As can be seen, the F-LMS algorithm, the solid red curve, has lower MSE and higher convergence rate when it utilizes the same step-size as the LMS algorithm, the dash-dotted black curve. Moreover, for the dashed blue curve, when the LMS algorithm uses a smaller step-size, 0.002, in order to attain the same MSE as that of the F-LMS algorithm, the convergence speed of the LMS algorithm degrades significantly. Hence, the F-LMS algorithm outperforms the LMS algorithm by obtaining lower MSE and higher convergence rate.

The passband frequency of $\mathbf{w}'_*$ is $[\frac{\pi}{4} - 0.2, \frac{\pi}{4} + 0.2]$, and its length, 0.4, is between 0.07π and 0.14π, thus we consider the complex conjugate zeros $e^{\pm j(\frac{\pi}{4} - 0.05)}$ and $e^{\pm j(\frac{\pi}{4} + 0.05)}$ for the transfer function $F(z)$. By utilizing the feature matrix generated by this transfer function as explained in Section IV, we can apply the F-LMS algorithm to identify $\mathbf{w}'_*$. The

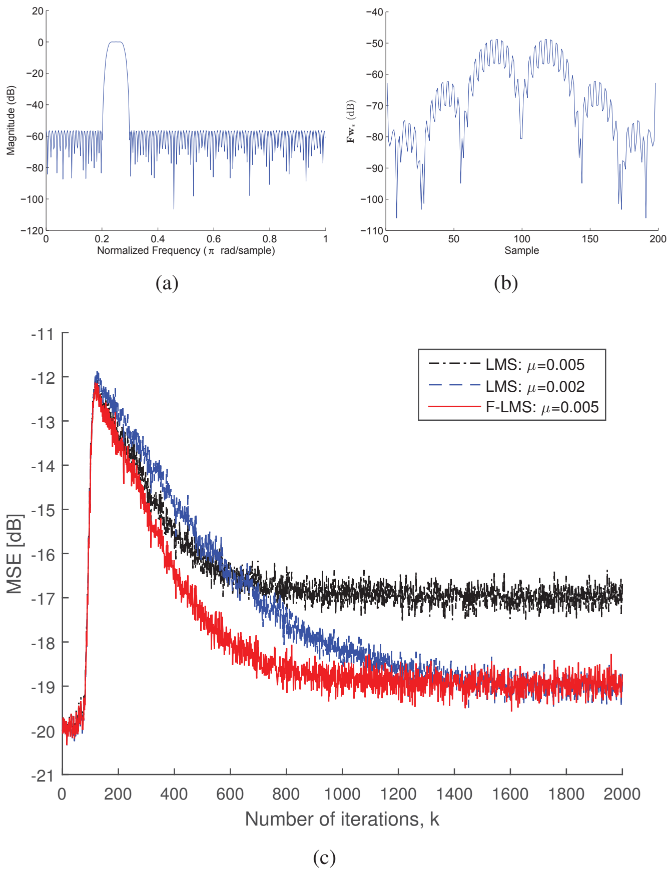

Fig. 1. (a) The magnitude response of the filter represented by $\mathbf{w}_*$ expressed in decibels; (b) $\mathbf{F}\mathbf{w}_*$ in decibels; (c) the MSE learning curves of the LMS and the F-LMS algorithms considering $\mathbf{w}_*$.

MSE learning curves of the LMS and the F-LMS algorithms, considering the bandpass system $\mathbf{w}'_*$, are represented in Figure 2(a). We can observe that, by using the step-size 0.003, the F-LMS algorithm, the solid red curve, has lower MSE and slightly higher convergence rate in comparison with the LMS algorithm, the dash-dotted black curve. In the dashed blue curve, we have used a small step-size, 0.001, for the LMS algorithm so that it attains the same MSE as the F-LMS algorithm. We can observe a notable decrease in the convergence rate of the LMS algorithm.

The passband frequency of $\mathbf{w}''_*$ is $[\frac{\pi}{3} - 0.1\pi, \frac{\pi}{3} + 0.1\pi]$, and its length is 0.2π, between 0.14π and 0.21π. Thus, by Remark 1, the transfer function $F(z)$ should contain three complex conjugate zeros. For this purpose, we have cascaded three transfer functions with zeros at $e^{\pm j(\frac{\pi}{3}-0.05\pi)}$, $e^{\pm j\frac{\pi}{3}}$, and $e^{\pm j(\frac{\pi}{3}+0.05\pi)}$ to form the transfer function $F(z)$. Figure 2(b) depicts the MSE learning curves of the LMS and the F-LMS algorithms considering the bandpass system $\mathbf{w}''_*$. As can be seen, the F-LMS algorithm with the step-size 0.003, the solid red curve, has lower MSE and slightly higher convergence speed as compared to the LMS algorithm with the same step-size. However, when we utilized the step-size 0.001 for the LMS algorithm in the dashed blue curve to achieve the same MSE as that of the F-LMS algorithm, we observed a significant

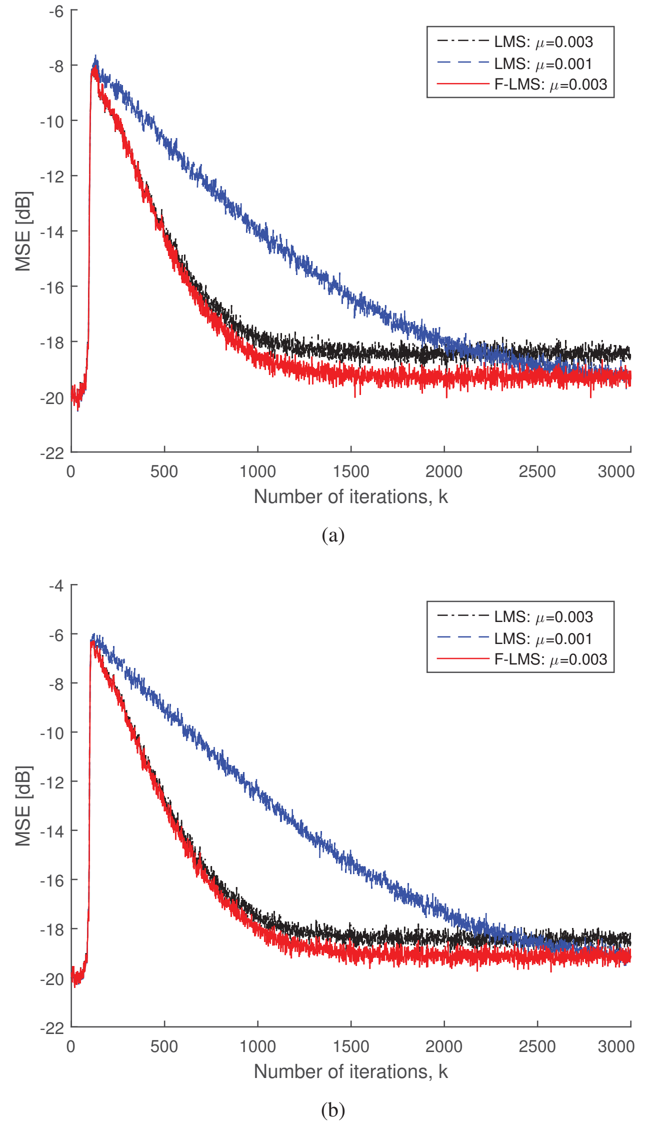

Fig. 2. The MSE learning curves of the LMS and the F-LMS algorithms considering: (a) $\mathbf{w}'_*$; (b) $\mathbf{w}''_*$.

reduction in the convergence rate of the LMS algorithm.

VII. Conclusions

In this paper, we advanced how to exploit sparsity in passband system models utilizing the feature LMS (F-LMS) algorithms. The F-LMS algorithms for bandpass models expose their inherent hidden features leading to higher convergence speed and reduced steady-state MSE. The feature matrices related to passband models are the key ingredient to achieve the improved performance, and can be applied to many engineering problems where these kinds of hidden sparsity are known to exist. Some simulations show the effectiveness of the proposed F-LMS algorithms when exposing the hidden sparsity feature.

In forthcoming publications, we will discuss computationally efficient versions of the proposed algorithms along with their MSE analysis. The derivation of feature adaptive filtering utilizing alternative algorithm will also be subject of future investigation.

References

[1] B. Widrow and S.D. Stearns, *Adaptive Signal Processing*, Prentice Hall, Englewood Cliffs, NJ, 1985.

[2] S. Haykin, *Adaptive Filter Theory*, Prentice Hall, Englewood Cliffs, NJ, 4th edition, 2002.

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

[4] S.C. Douglas, "A family of normalized LMS algorithms," *IEEE Signal Processing Letters*, vol. 1, no. 3, pp. 49–51, Mar. 1994.

[5] H.-C. Shin, A.H. Sayed, and W.-J. Song, "Variable step-size NLMS and affine projection algorithms," *IEEE Signal Processing Letters*, vol. 11, no. 2, pp. 132–135, Feb. 2004.

[6] 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.

[7] H. Yazdanpanah, M.V.S. Lima, and P.S.R. Diniz, "On the robustness of set-membership adaptive filtering algorithms," *EURASIP Journal on Advances in Signal Processing*, vol. 72, 2017.

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

[9] C. Paleologu, J. Benesty, and S. Ciochină, *Sparse Adaptive Filters for Echo Cancellation*, Synthesis Lectures on Speech and Audio Processing. Morgan & Claypool Publishers, 2010.

[10] 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., Sept. 2011, pp. 1–5.

[11] 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.

[12] J.F. de Andrade, M.L.R. de Campos, and J.A. Apolinário, "An l_1-norm linearly constrained affine projection algorithm," in *2016 IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM 2016)*, July 2016, pp. 1–5.

[13] J.F. de Andrade, M.L.R. de Campos, and J.A. Apolinário, "L_1-constrained normalized LMS algorithms for adaptive beamforming," *IEEE Transactions on Signal Processing*, vol. 63, no. 24, pp. 6524–6539, Dec. 2015.

[14] D. Angelosante, J.A. Bazerque, and G.B. Giannakis, "Online adaptive estimation of sparse signals: Where RLS meets theℓ_1-norm," *IEEE Transactions on Signal Processing*, vol. 58, no. 7, pp. 3436–3447, July 2010.

[15] Y.V. Zakharov and V.H. Nascimento, "DCD-RLS adaptive filters with penalties for sparse identification," *IEEE Transactions on Signal Processing*, vol. 61, no. 12, pp. 3198–3213, June 2013.

[16] P.S.R. Diniz, H. Yazdanpanah, and M.V.S. Lima, "Feature LMS algorithms," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2018)*, Calgary, Canada, Apr. 2018, pp. 4144–4148.

[17] H. Yazdanpanah, J.A. Apolinário, P.S.R. Diniz, and M.V.S. Lima, "l_0-norm feature LMS algorithms," in *2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP 2018)*, Anaheim, CA, USA, Nov. 2018, pp. 311–315.

[18] B. Babadi, N. Kalouptsidis, and V. Tarokh, "Sparls: The sparse rls algorithm," *IEEE Transactions on Signal Processing*, vol. 58, no. 8, pp. 4013–4025, Aug. 2010.

[19] 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.

[20] Y. Chen, Y. Gu, and A.O. Hero, "Sparse LMS for system identification," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2009)*, Taipei, Taiwan, Apr. 2009, pp. 3125–3128.

[21] 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.

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

[23] E.J. Candes, M.B. Wakin, and S.P. Boyd, "Enhancing sparsity by reweighted l_1 minimization," *Journal of Fourier Analysis and Applications*, vol. 14, no. 5, pp. 877–905, Dec. 2008.

[24] 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.

[25] 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.

[26] H. Yazdanpanah, P.S.R. Diniz, and M.V.S. Lima, "A simple set-membership affine projection algorithm for sparse system modeling," in *24th European Signal Processing Conference (EUSIPCO 2016)*, Budapest, Hungary, Sept. 2016, pp. 1798–1802.

[27] H. Yazdanpanah and P.S.R. Diniz, "Recursive least-squares algorithms for sparse system modeling," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2017)*, New Orleans, LA, USA, Mar. 2017, pp. 3879–3883.