Received: 29 May 2020 | Revised: 29 September 2022 | Accepted: 7 November 2022

DOI: 10.1002/acs.3528

RESEARCH ARTICLE

WILEY

Improved feature least mean square algorithm

Hamed Yazdanpanah

Department of Computer Science, University of São Paulo (USP), São Paulo, Brazil

Correspondence
Hamed Yazdanpanah, Department of Computer Science, University of São Paulo (USP), São Paulo, SP 05508-090, Brazil.
Email: hamed.yazdanpanah@smt.ufrj.br

Summary

In this paper, we propose the improved feature least-mean-square (IF-LMS) algorithm to exploit hidden sparsity in unknown systems. Recently, the feature least-mean-square (F-LMS) algorithm has been introduced, but its application is limited to particular systems since it uses predetermined feature matrices. However, the proposed IF-LMS algorithm utilizes the stochastic gradient descent (SGD) method to learn feature matrices; thus, it can be used in any system that the classical LMS algorithm is applicable. Hence, by employing a learnable feature matrix, the IF-LMS algorithm has a vast application area as compared to the F-LMS algorithm. Moreover, mathematically, we discuss some parameters of the IF-LMS algorithm. Simulation results, in synthetic and real-life scenarios, demonstrate that the IF-LMS algorithm has superior filtering accuracy to the well-known LMS algorithm.

KEYWORDS

adaptive filtering, feature, hidden sparsity, LMS, stochastic gradient descent

1 | INTRODUCTION

Stochastic gradient descent (SGD) is one of the most famous approaches in adaptive learning algorithms. Since 1960, the least-mean-square (LMS) algorithm is a prominent member of the family of SGD algorithms due to its simplicity, computational efficiency, and modest hardware requirements.[1,2] The LMS algorithm and its variants have many applications in real-life problems, such as active noise control,[3] adaptive beamforming,[4,5] system identification,[6] signal prediction,[7] etc. The traditional LMS algorithm is unable to take advantage of the structure of adaptive coefficients, whereas, in many cases, we can improve the performance of the algorithm by exploiting the adaptive filter structure.

Recently, the feature LMS (F-LMS) algorithm is proposed to outperform the classical LMS algorithm by exploiting hidden sparsity in some systems, such as lowpass, highpass, and bandpass systems.[8-10] However, the proposed F-LMS algorithm has two drawbacks: (i) its application is restricted to some particular systems, such as lowpass, highpass, and bandpass systems; (ii) we do require some a priori knowledge about the spectral characteristics of unknown system, otherwise its performance can be inferior to the conventional LMS algorithm. Therefore, we should avoid using the F-LMS algorithm for an arbitrary system or when we do not have a priori information about the spectral characteristics of the system.

To remove the restrictions of the F-LMS algorithm, in this work, the improved F-LMS (IF-LMS) algorithm is introduced. Indeed, the IF-LMS algorithm can be applied to any unknown system that the classical LMS algorithm is applicable, and we need no a priori information about the spectral characteristics of the systems. To this end, we propose a time-varying feature matrix so that all nonzero entries of the feature matrix are adapted at each iteration. In other words, instead of adopting a predetermined feature matrix as in the conventional F-LMS algorithm, we utilize the SGD approach to learn feature matrices. In simulation results, we will observe that the learned feature matrices can expose hidden sparsity in any system without requiring a priori knowledge about the spectral characteristics of the system.

Finally, it is worthwhile to mention that although the IF-LMS algorithm uses some sparsity-promoting tools, this algorithm is fundamentally different from conventional sparsity-aware adaptive filtering algorithms, such as those in References 11-13. The conventional sparsity-aware adaptive filtering algorithms are designed to exploit systems/coefficients which are sparse, whereas the IF-LMS algorithm is designed to exploit a different feature. That is, the IF-LMS algorithm applies a transformation that maps any system (which does not need to be sparse in its original domain) to a different domain where the transformed coefficients are sparse. Then, the IF-LMS algorithm exploits the sparsity in the transformed domain in order to improve the learning process. However, in future works, the proposed idea in this paper can be applied to other sparsity-aware adaptive filtering algorithms, such as proportionate filters, to obtain additional gains.

This paper is organized as follows. Section 2 introduces the IF-LMS algorithm. In Section 3, we discuss some properties of the IF-LMS algorithm, such as the stability and the computational complexity. Experimental results, including numerical and real-life examples, are presented in Section 4. Finally, conclusions are drawn in Section 5.

1.1 | Notation

Scalars are denoted by lowercase letters. Vectors (matrices) are represented by lowercase (uppercase) boldface letters. For a given iteration k, the weight vector, the optimum solution, and the input vector are presented by $\mathbf{w}(k), \mathbf{w}_*, \mathbf{x}(k) \in \mathbb{R}^{N+1}$, respectively, where N is the adaptive filter order. Also, the desired signal is denoted by $d(k) \in \mathbb{R}$, and the error signal is described by $e(k) = d(k) - \mathbf{w}^T(k)\mathbf{x}(k)$. The l_1-norm and the Euclidean norm of a vector $\mathbf{w} \in \mathbb{R}^{N+1}$ are defined by $\|\mathbf{w}\|_1 = \sum_{i=0}^{N}|w_i|$ and $\|\mathbf{w}\|^2 = \mathbf{w}^T\mathbf{w} = \sum_{i=0}^{N} w_i^2$, respectively. Furthermore, $(\cdot)^{\mathrm{T}}$, $\odot$, $\nabla_{\mathbf{w}}(\cdot)$ stand for the vector transpose operator, the Hadamard product of two vectors, and the gradient with respect to $\mathbf{w}$, respectively. Also, $E[\cdot]$ and $\mathcal{U}(-1,1)$ denote the expected value operator and the continuous uniform distribution with the distribution's support $(-1,1)$, respectively. Moreover, $\mathrm{sgn}(\cdot)$ shows the sign function, and it is defined by

$$\mathrm{sgn}(z) = \begin{cases} 1, & \text{if } z > 0 \\ 0, & \text{if } z = 0 \\ -1, & \text{if } z < 0 \end{cases} \tag{1}$$

2 | THE IF-LMS ALGORITHM

In this section, we propose the IF-LMS algorithm. To this end, first, we introduce the objective criterion of the algorithm in its general form. Then, by adopting the l_1-norm as the sparsity-promoting penalty function, we use the stochastic gradient descent approach to learn the feature matrix and the adaptive coefficients of the algorithm.

The objective function of the IF-LMS algorithm is given by

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

where $\alpha \in \mathbb{R}_+$ is the weight dedicated to the sparsity-promoting penalty function $\mathcal{P} : \mathbb{R}^{N+1} \to \mathbb{R}_+$. Moreover, the function of the feature matrix $\mathbf{F}(k)$ is to expose the hidden sparsity. Indeed, the feature matrix should transform the adaptive filter $\mathbf{w}(k)$ to a sparse vector, that is, $\mathbf{F}(k)\mathbf{w}(k)$ should be a vector with many entries close or equal to zero. It is worth mentioning that $\mathbf{w}(k)$ does not require to be a sparse vector; however, $\mathbf{F}(k)$ must be chosen such that $\mathbf{F}(k)\mathbf{w}(k)$ produces a sparse vector. Once the hidden sparsity is revealed by $\mathbf{F}(k)$, the function $\mathcal{P}$ should exploit the exposed sparsity.

There are various candidates for $\mathcal{P}$ to exploit revealed sparsity, such as the l_0-norm,[13-15] the l_1-norm,[16-19] the thresholding approaches,[20-23] etc. In this work, we adopt the l_1-norm due to its simplicity and computational efficiency. Therefore, the minimization problem (2) is reduced to

$$\xi_{\text{IF-LMS}}(k) = \frac{1}{2}|e(k)|^2 + \alpha\|\mathbf{F}(k)\mathbf{w}(k)\|_1. \tag{3}$$

The feature matrix $\mathbf{F}(k)$ represents any linear combination applied to $\mathbf{w}(k)$ in order to generate a sparse vector. For the sake of computational efficiency, we construct $\mathbf{F}(k)$ so that at each row exactly two adjacent entries are different from zero. Therefore, for adjacent coefficients $w_i(k)$ and $w_{i+1}(k)$, for some i, we want to design parameters $m, u \in \mathbb{R}$ such that $mw_i(k) + uw_{i+1}(k)$ is equal or close to zero. After learning the desired m and u, by using the l_1-norm penalty function, we should learn the adaptive filter coefficients $w_i(k)$ and $w_{i+1}(k)$ such that the relation $mw_i(k) + uw_{i+1}(k) \approx 0$ is satisfied. Hence, we can assume $\mathbf{F}(k) \in \mathbb{R}^{N\times(N+1)}$ of the following form

$$\mathbf{F}(k) = \begin{bmatrix} m_1(k) & u_1(k) & 0 & \cdots & 0 \\ 0 & m_2(k) & u_2(k) & \cdots & 0 \\ \vdots & & \ddots & \ddots & \\ 0 & 0 & \cdots & m_N(k) & u_N(k) \end{bmatrix}, \tag{4}$$

where all entries of $\mathbf{F}(k)$ are zero except the main diagonal $\mathbf{m}(k) = [m_1(k) \ \cdots \ m_N(k)]^T$ and the upper diagonal $\mathbf{u}(k) = [u_1(k) \ \cdots \ u_N(k)]^T$.

To expose the hidden sparsity of $\mathbf{w}(k)$ by $\mathbf{F}(k)$, we should learn the nonzero entries of $\mathbf{F}(k)$ such that $\|\mathbf{F}(k)\mathbf{w}(k)\|_1$ is minimized. For this purpose, we can use the stochastic gradient descent technique and propose the recursions

$$\mathbf{m}(k+1) = \mathbf{m}(k) - \mu_F \nabla_{\mathbf{m}(k)}\|\mathbf{F}(k)\mathbf{w}(k)\|_1, \tag{5}$$

$$\mathbf{u}(k+1) = \mathbf{u}(k) - \mu_F \nabla_{\mathbf{u}(k)}\|\mathbf{F}(k)\mathbf{w}(k)\|_1, \tag{6}$$

where μ_F denotes the learning rate. After computing the gradients, we get

$$\mathbf{m}(k+1) = \mathbf{m}(k) - \mu_F \mathbf{w}_f(k) \odot \text{sgn}(\mathbf{g}(k)), \tag{7}$$

$$\mathbf{u}(k+1) = \mathbf{u}(k) - \mu_F \mathbf{w}_l(k) \odot \text{sgn}(\mathbf{g}(k)), \tag{8}$$

where

$$\mathbf{w}_f(k) = [w_0(k) \cdots \ w_{N-1}(k)]^T, \tag{9}$$

$$\mathbf{w}_l(k) = [w_1(k) \ \cdots \ w_N(k)]^T. \tag{10}$$

In other words, $\mathbf{w}_f(k)$ and $\mathbf{w}_l(k)$ are the first and the last N components of $\mathbf{w}(k)$, respectively. Also, $\mathbf{g}(k)$ is defined by

$$\mathbf{g}(k) = \mathbf{m}(k) \odot \mathbf{w}_f(k) + \mathbf{u}(k) \odot \mathbf{w}_l(k). \tag{11}$$

We also utilize the gradient descent strategy to minimize the objective function (3). Thus, we will have

$$\mathbf{w}(k+1) = \mathbf{w}(k) - \mu \nabla_{\mathbf{w}(k)} \left(\frac{1}{2}|d(k) - \mathbf{w}^T(k)\mathbf{x}(k)|^2 + \alpha\|\mathbf{F}(k)\mathbf{w}(k)\|_1 \right), \tag{12}$$

where μ is the step-size parameter. After computing the gradient, the recursion for $\mathbf{w}(k)$ can be given by

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \mu e(k)\mathbf{x}(k) - \mu\alpha\mathbf{p}(k), \tag{13}$$

where $\mathbf{p}(k) \in \mathbb{R}^{N+1}$ is the gradient of $\|\mathbf{F}(k)\mathbf{w}(k)\|_1$ with respect to $\mathbf{w}(k)$. Hence, the ith entry of $\mathbf{p}(k)$ is given by

Algorithm 1. Pseudocode for the IF-LMS algorithm

Initialization
$\mathbf{w}(0) = [0 \cdots 0]^{\mathrm{T}}$, $\mathbf{m}(0) \sim \mathcal{U}(-1, 1)$, $\mathbf{u}(0) \sim \mathcal{U}(-1, 1)$
choose a positive α and the learning rates μ and μ_F
for $k \geq 0$ **do**
  $e(k) = d(k) - \mathbf{w}^T(k)\mathbf{x}(k)$
  $\mathbf{g}(k) = \mathbf{m}(k) \odot \mathbf{w}_f(k) + \mathbf{u}(k) \odot \mathbf{w}_l(k)$
  $\mathbf{m}(k+1) = \mathbf{m}(k) - \mu_F \mathbf{w}_f(k) \odot \mathrm{sgn}(\mathbf{g}(k))$
  $\mathbf{u}(k+1) = \mathbf{u}(k) - \mu_F \mathbf{w}_l(k) \odot \mathrm{sgn}(\mathbf{g}(k))$
  for $0 \leq i \leq N$ **do**
    $$p_i(k) = \begin{cases} m_1(k)\mathrm{sgn}\big(m_1(k)w_0(k) + u_1(k)w_1(k)\big), & \text{if } i = 0 \\ u_i(k)\mathrm{sgn}\big(m_i(k)w_{i-1}(k) + u_i(k)w_i(k)\big) + m_{i+1}(k)\mathrm{sgn}\big(m_{i+1}(k)w_i(k) + u_{i+1}(k)w_{i+1}(k)\big), & \text{if } i = 1, \cdots, N-1 \\ u_N(k)\mathrm{sgn}\big(m_N(k)w_{N-1}(k) + u_N(k)w_N(k)\big), & \text{if } i = N. \end{cases}$$
  $\mathbf{w}(k+1) = \mathbf{w}(k) + \mu e(k)\mathbf{x}(k) - \mu\alpha\mathbf{p}(k)$
end for

$$p_i(k) = \begin{cases} m_1(k)\mathrm{sgn}\left(m_1(k)w_0(k) + u_1(k)w_1(k)\right), & \text{if } i = 0 \\ u_i(k)\mathrm{sgn}\left(m_i(k)w_{i-1}(k) + u_i(k)w_i(k)\right) + m_{i+1}(k)\mathrm{sgn}\left(m_{i+1}(k)w_i(k) + u_{i+1}(k)w_{i+1}(k)\right), & \text{if } i = 1, \cdots, N-1. \\ u_N(k)\mathrm{sgn}\left(m_N(k)w_{N-1}(k) + u_N(k)w_N(k)\right), & \text{if } i = N. \end{cases} \tag{14}$$

Therefore, to implement the IF-LMS algorithm, at each iteration, we firstly learn the nonzero entries of the feature matrix by updating Equations (7) and (8), then we learn the adaptive filter coefficients by updating Equation (13). To summarize the IF-LMS algorithm, all steps are described in Algorithm 1.

Remark 1. In theory, $\mathbf{F}(k)$ can be zero for some k, but in practice, it is very rare since for happening this event both $\mathbf{m}(k)$ and $\mathbf{u}(k)$ should be zero. When at least one of $\mathbf{m}(k)$ or $\mathbf{u}(k)$ is different from zero, by (7) and (8), for the next update, the nonzero vector will push the zero one to a nonzero vector too. Moreover, when $\mathbf{F}(k)$ is equal to zero for some k, it will not generate any risk for the convergence of the IF-LMS algorithm. Since for $\mathbf{F}(k) = \mathbf{0}$, the IF-LMS algorithm will be reduced to the conventional LMS algorithm.

3 | SOME PROPERTIES OF THE IF-LMS ALGORITHM

In this section, we analyze the step-size parameter μ to introduce the valid range for this parameter so that the convergence is guaranteed. Also, we study the weight given to the l_1-norm penalty, that is, α. To this end, denote the difference between the adaptive filter and the optimum solution by $\tilde{\mathbf{w}}(k) \triangleq \mathbf{w}_* - \mathbf{w}(k)$. If we subtract from $\mathbf{w}_*$ the both sides of (13), and then take the Euclidean norm of both sides, we get

$$\begin{aligned} \|\tilde{\mathbf{w}}(k+1)\|^2 - \|\tilde{\mathbf{w}}(k)\|^2 = {} & \mu^2\alpha^2\|\mathbf{p}(k)\|^2 + \alpha\left(2\mu\tilde{\mathbf{w}}^T(k)\mathbf{p}(k) - 2\mu^2 e(k)\mathbf{x}^T(k)\mathbf{p}(k)\right) \\ & + \left(\mu^2 e^2(k)\|\mathbf{x}(k)\|^2 - 2\mu e(k)\tilde{\mathbf{w}}^T(k)\mathbf{x}(k)\right). \end{aligned} \tag{15}$$

By taking the expected values of both sides of the equation above, we get

$$\begin{aligned} E\left[\|\tilde{\mathbf{w}}(k+1)\|^2 - \|\tilde{\mathbf{w}}(k)\|^2\right] = {} & \mu^2\alpha^2 E\left[\|\mathbf{p}(k)\|^2\right] + \alpha\left(2\mu E\left[\tilde{\mathbf{w}}^T(k)\mathbf{p}(k)\right] - 2\mu^2 E\left[e(k)\mathbf{x}^T(k)\mathbf{p}(k)\right]\right) \\ & + \left(\mu^2 E\left[e^2(k)\|\mathbf{x}(k)\|^2\right] - 2\mu E\left[e(k)\tilde{\mathbf{w}}^T(k)\mathbf{x}(k)\right]\right). \end{aligned} \tag{16}$$

Note that the convergence of the IF-LMS algorithm must be guaranteed even when hidden sparsity is not exploited (when $\alpha = 0$).[13,24,25] Therefore, assuming $\alpha = 0$, Equation (16) reduces to

$$E\left[\|\tilde{\mathbf{w}}(k+1)\|^2 - \|\tilde{\mathbf{w}}(k)\|^2\right] = \mu^2 E\left[e^2(k)\|\mathbf{x}(k)\|^2\right] - 2\mu E\left[e(k)\tilde{\mathbf{w}}^T(k)\mathbf{x}(k)\right]. \tag{17}$$

TABLE 1 Computational cost of the least mean square (LMS) and the improved feature LMS (IF-LMS) algorithms in terms of real multiplications and real additions per iteration

Algorithm	Number of additions	Number of multiplications
LMS	$2N + 2$	$2N + 3$
IF-LMS	$8N + 2$	$8N + 6$

Then using the same argument as in References 2, pp. 85–87, it can be concluded that a necessary condition for the convergence of the IF-LMS algorithm is

$$0 < \mu < \frac{2}{\mathrm{tr}(\mathbf{R})}, \tag{18}$$

where $\mathrm{tr}(\cdot)$ is the trace operator, and $\mathbf{R} = E\left[\mathbf{x}(k)\mathbf{x}^T(k)\right]$.

To study the weight given to the l_1-norm, α, note that the right-hand side of (16) can be considered as a quadratic function of α such as

$$f(\alpha) = a(k)\alpha^2 + b(k)\alpha + c(k), \tag{19}$$

where

$$a(k) = \mu^2 E\left[\|\mathbf{p}(k)\|^2\right] > 0, \tag{20}$$

$$b(k) = 2\mu E\left[\tilde{\mathbf{w}}^T(k)\mathbf{p}(k)\right] - 2\mu^2 E\left[e(k)\mathbf{x}^T(k)\mathbf{p}(k)\right], \tag{21}$$

$$c(k) = \mu^2 E\left[e^2(k)\|\mathbf{x}(k)\|^2\right] - 2\mu E\left[e(k)\tilde{\mathbf{w}}^T(k)\mathbf{x}(k)\right]. \tag{22}$$

If we choose μ as in (18), we have the necessary condition for the convergence of the IF-LMS algorithm. Thus, for μ as in (18), we have $E\left[\|\tilde{\mathbf{w}}(k+1)\|^2 - \|\tilde{\mathbf{w}}(k)\|^2\right] < 0$. However, by Equation (17), $E\left[\|\tilde{\mathbf{w}}(k+1)\|^2 - \|\tilde{\mathbf{w}}(k)\|^2\right] = c(k)$. Therefore, if we select μ as in (18), we get $c(k) < 0$. Thus, the discriminant of $f(\alpha)$ is positive, that is

$$b^2(k) - 4a(k)c(k) > 0. \tag{23}$$

As a result, the quadratic function (19) has two real and distinct roots. If we denote these roots by $\alpha_{\min}(k)$ and $\alpha_{\max}(k)$, since $a(k) > 0$, we conclude that $f(\alpha) < 0$, for $\alpha \in [\alpha_{\min}(k), \alpha_{\max}(k)]$. Thus, the necessary conditions for the convergence of the IF-LMS algorithm are $\alpha \in [\alpha_{\min}(k), \alpha_{\max}(k)]$ and $0 < \mu < \frac{2}{\mathrm{tr}(\mathbf{R})}$.

Remark 2. Note that $\alpha_{\min}(k)\alpha_{\max}(k) = \frac{c(k)}{a(k)} < 0$; thus, $\alpha_{\min}(k) < 0 < \alpha_{\max}(k)$. As a consequence, we choose $\alpha \in [0, \alpha_{\max})$. In practice, by adopting the grid search approach, we select α in a similar manner of choosing μ. Indeed, we adopt μ and α as small positive real numbers; if the algorithm diverges, we reduce them.

Remark 3. The sufficient condition for the convergence of the IF-LMS algorithm may require a significantly smaller μ than $\frac{2}{\mathrm{tr}(\mathbf{R})}$, especially when the measurement noise is strong. Therefore, in practice, we choose a very small μ.

Remark 4. The computational complexity of the IF-LMS algorithm is not much higher than that of the LMS algorithm, and both algorithms have computational burden $\mathcal{O}(N)$. The required number of real additions and real multiplications per iteration for the LMS and the IF-LMS algorithms are described in Table 1.

4 | SIMULATIONS

In this section, we utilize the LMS and the IF-LMS algorithms in some system identification problems. In all cases, the adaptive filter coefficients are initialized with the null vector. Also, the nonzero entries of the feature matrix are initialized randomly using $\mathcal{U}(-1, 1)$. The input signal has zero-mean white Gaussian distribution with unit variance, except one experiment where will be mentioned. The mean squared error (MSE) learning curves and the normalized misalignment (MIS) curves, $\|\mathbf{w}_* - \mathbf{w}(k)\|/\|\mathbf{w}_*\|$, are computed by averaging the outcomes of 100 independent trials, and they are smoothed by a box filter of length 100. Moreover, the grid search strategy is adopted to choose the values of hyperparameters, such as μ, α, and μ_F.

4.1 | Synthetic examples

As the first synthetic example, the LMS and the IF-LMS algorithms are employed to identify an unknown random system $\mathbf{w}_*$ of order 79 whose coefficients are drawn from zero-mean white Gaussian distribution with unit variance. The signal-to-noise ratio (SNR) is equal to 0 dB in this scenario, where $\text{SNR} = 10\log_{10}\left(\frac{\sigma^2}{\sigma_n^2}\right)$ in which σ^2 and σ_n^2 are the output and noise signal variances, respectively. In the IF-LMS algorithm, μ_F and α are chosen as 0.002 and 0.2, respectively. Figure 1A shows the MSE learning curves of the LMS and the IF-LMS algorithms using three different step-size values. As can be seen, for each step-size value, the IF-LMS algorithm can attain lower steady-state MSE compared to the LMS algorithm. Also, when the step-size value is equal to 0.02 and 0.018, the IF-LMS algorithm has a higher convergence rate compared to the LMS algorithm, whereas for the step-size equal to 0.015, they have the same convergence speed. Figure 1B depicts the impulse response of $\mathbf{w}_*$. Moreover, Figure 1C, D, and E present $\mathbf{F}(k)\mathbf{w}(k)$, $\mathbf{m}(k)$, and $\mathbf{u}(k)$ after the convergence of the IF-LMS algorithm, respectively. As can be observed, $\mathbf{w}_*$ is not a sparse system. But $\mathbf{F}(k)$ reveals the hidden sparsity in $\mathbf{w}_*$ and transforms the adaptive filter to the sparse domain.

Furthermore, Figure 2A shows the MSE learning curves of the IF-LMS algorithm identifying $\mathbf{w}_*$, when the SNR is 0 dB, and the values of α vary from 0.01 to 1. As can be seen, for all tested values of α, the IF-LMS algorithm

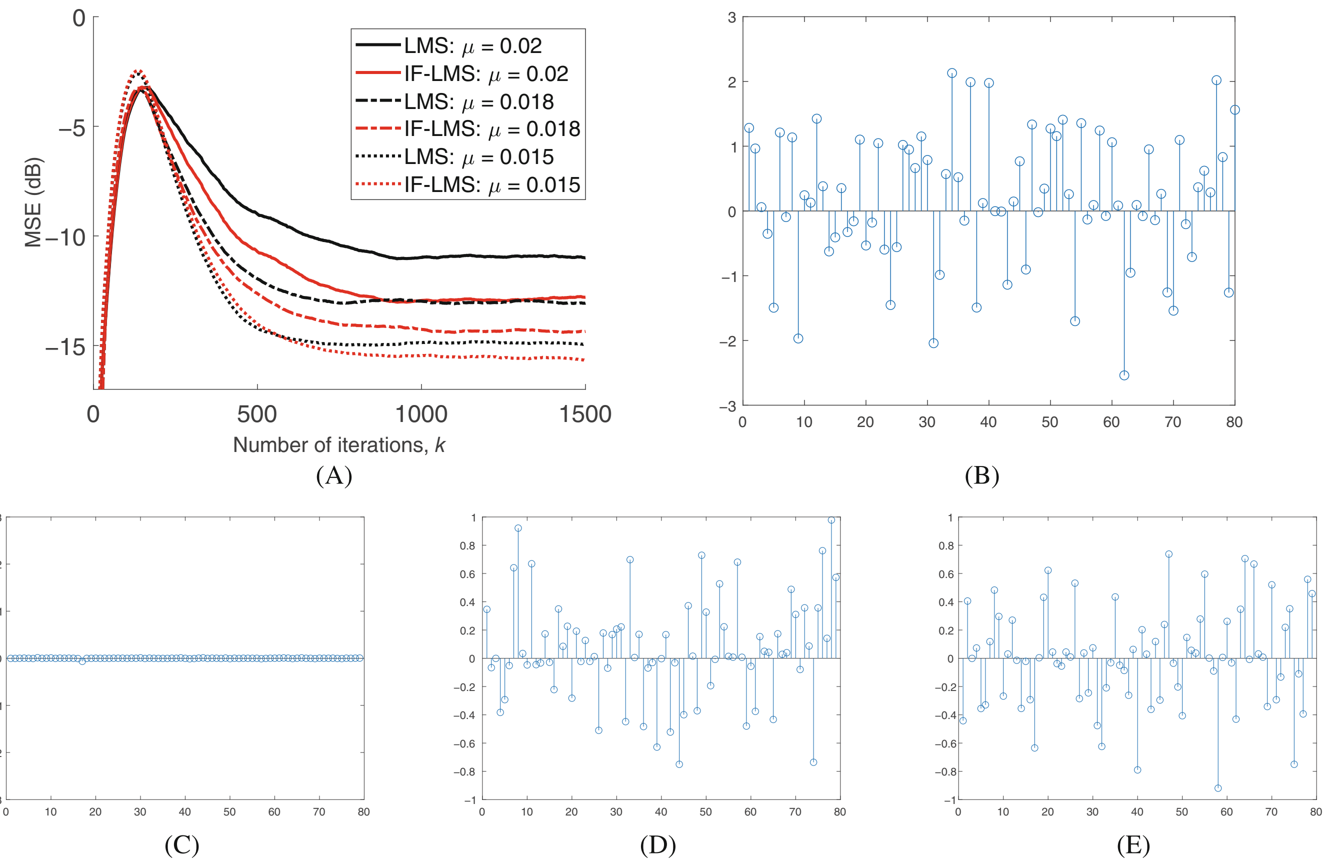

FIGURE 1 (A) The mean squared error learning curves of the least-mean-square and the improved feature least-mean-square (IF-LMS) algorithms identifying the unknown random system $\mathbf{w}_*$ using various step-size values; (B) the impulse response of $\mathbf{w}_*$; (C) $\mathbf{F}(k)\mathbf{w}(k)$; (D) $\mathbf{m}(k)$; (E) $\mathbf{u}(k)$ after the convergence of the IF-LMS algorithm

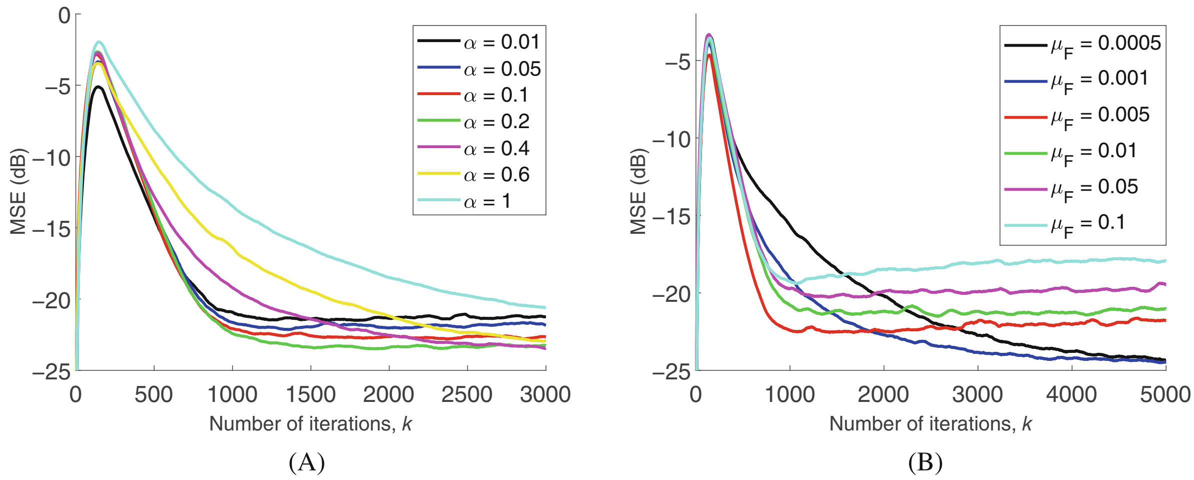

FIGURE 2 The mean squared error learning curves of the improved feature least-mean-square algorithm identifying $\mathbf{w}_*$ for: (A) different values of α; (B) different values of μ_F

converged, whereas its steady-state MSE and convergence rate are affected significantly by changing the value of α. Also, Figure 2B depicts the MSE learning curves of the IF-LMS algorithm identifying $\mathbf{w}_*$, when the SNR is 0 dB and the values of μ_F change from 0.0005 to 0.1. As can be observed, the IF-LMS algorithm can obtain a good agreement between the steady-state MSE and the convergence velocity by choosing μ_F around 0.001. Moreover, we can observe that even for a large μ_F, such as 1, the IF-LMS algorithm did not diverge, whereas it attained a high steady-state MSE.

As the second synthetic example, we use the LMS, the variable step-size LMS (VSS-LMS),[26] and the IF-LMS algorithms to identify an unknown random system $\mathbf{w}'_*$ of order 79 whose coefficients are drawn from zero-mean white Gaussian distribution with unit variance. Figure 3A, shows the MSE learning curves of the LMS, the VSS-LMS, and the IF-LMS algorithms identifying $\mathbf{w}'_*$. For the first 2000 iterations, the additive noise on the output of the system is a zero-mean white Gaussian noise (WGN), where the SNR is equal to 10 dB. However, at the iteration 2001, the SNR is changed to 14 dB, and the impulse response of the unknown system changed to another random system of the same order whose coefficients are drawn from zero-mean white Gaussian distribution with unit variance. For the LMS and IF-LMS algorithms, the step-size parameter, μ, is adopted as 0.02. In the IF-LMS algorithm, μ_F and α are chosen as 0.002 and 0.2, respectively. For the VSS-LMS algorithm, $\mu(0)$, α, and γ are chosen as 0.02, 0.995, and 5×10^{-7}, respectively. As can be seen, the IF-LMS algorithm has lower steady-state MSE and slightly higher convergence rate as compared to the LMS algorithm. Initially, compared with the IF-LMS algorithm, the VSS-LMS algorithm attains a higher convergence speed and the same MSE. However, after an abrupt change in the unknown system, the VSS-LMS algorithm has the worst MSE compared to the LMS and IF-LMS algorithms. Thus, the IF-LMS algorithm has the lowest MSE and a competitive tracking capability. Moreover, Figure 3B,C, shows the impulse response of $\mathbf{w}'_*$ and $\mathbf{F}(k)\mathbf{w}(k)$ after the convergence of the IF-LMS algorithm, respectively. As can be seen, $\mathbf{w}'_*$ is not a sparse system; however, $\mathbf{F}(k)$ reveals the hidden sparsity and transforms the adaptive filter to the sparse domain.

As the third synthetic example, we utilize the LMS and the IF-LMS algorithms to identify a bandpass system $\mathbf{w}''_*$. This bandpass system is of order 199, and the lower transition frequency, the lower cut-off frequency, the upper cut-off frequency, and the upper transition frequency of $\mathbf{w}''_*$ are given by $\frac{\pi}{3} - 0.45$, $\frac{\pi}{3} - 0.1\pi$, $\frac{\pi}{3} + 0.1\pi$, and $\frac{\pi}{3} + 0.45$, respectively. Figure 4A illustrates the MSE learning curves of the tested algorithms identifying $\mathbf{w}''_*$. In this case, the SNR is 10 dB, and the step-sizes of both algorithms are 0.005. Moreover, for the IF-LMS algorithm, the parameters μ_F and α are set to 0.05 and 0.1, respectively. We can observe that the IF-LMS algorithm outperformed the LMS algorithm by obtaining remarkable lower steady-state MSE. Moreover, Figure 4B depicts the MSE learning curves of the IF-LMS and the LMS algorithms identifying $\mathbf{w}''_*$ when the input signal is a first-order autoregressive process. In this case, the input signal is generated by $x(k) = 0.95x(k-1) + m(k)$, where $m(k)$ is a zero-mean WGN with unit variance. For this experiment, the step-sizes of both algorithms are chosen as 0.002. Also, for the IF-LMS algorithm, the values of μ_F and α are selected as 0.005 and 0.9, respectively. As can be seen in Figure 4B, the convergence rate of the IF-LMS algorithm is significantly higher than that of the LMS algorithm.

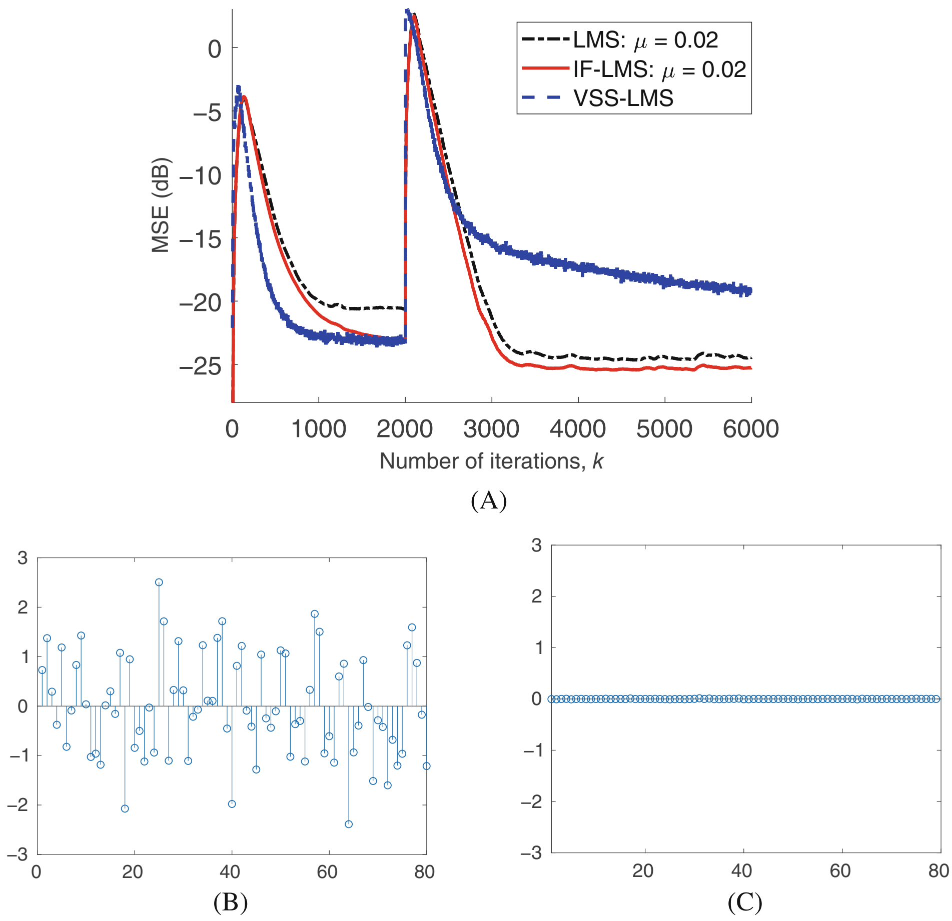

FIGURE 3 (A) The mean squared error learning curves of the least-mean-square (LMS), the variable step-size LMS, and the improved feature least-mean-square (IF-LMS) algorithms identifying the unknown random system $\mathbf{w}'_*$; (B) the impulse response of $\mathbf{w}'_*$; (C) $\mathbf{F}(k)\mathbf{w}(k)$ after the convergence of the IF-LMS algorithm

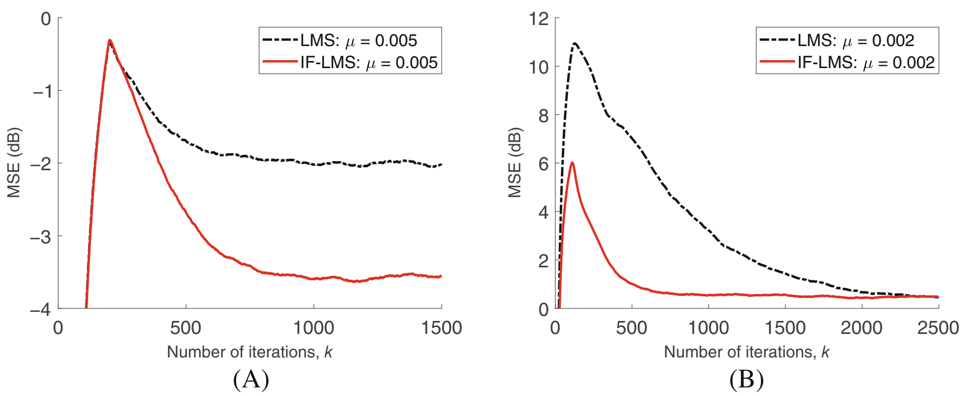

FIGURE 4 The mean squared error learning curves of the least-mean-square (LMS) and the improved feature least-mean-square algorithms identifying the unknown bandpass system $\mathbf{w}''_*$, when the input signal is: (A) zero-mean white Gaussian noise with unit variance; (B) a first-order autoregressive process

4.2 | Underwater channel estimation

In this scenario, as a real-life example, we utilize the LMS and the IF-LMS algorithms to estimate the impulse response of an underwater communication channel. The underwater channel acquisition details are presented in Reference 27. In this real-life example, the step-sizes of both algorithms are 0.0006, and the SNR is 10 dB. Also, the adopted adaptive filter is of order 2880. For the IF-LMS algorithm, μ_F and α are selected as 0.002 and 0.02, respectively. Figure 5A depicts the MSE learning curves of the LMS and the IF-LMS algorithms. Also, the MIS curves of the IF-LMS and the LMS algorithms are shown in Figure 5B. As can be observed, the IF-LMS algorithm attained notable lower steady-state MSE and MIS in comparison with the LMS algorithm.

4.3 | Wireless channel estimation

In this scenario, as another real-life test, we employ the LMS and the IF-LMS algorithms to estimate the impulse response of a wireless channel between 240 and 300 GHz for the transmitter-receiver distance of 80 cm. More details regarding this wireless channel acquisition are provided in Reference 28. For this scenario, the adaptive filter order is 4096. Also, the step-size parameters of both algorithms are 0.0004, and the SNR is selected as 14 dB. For the IF-LMS algorithm, μ_F and α are chosen as 0.5 and 0.15, respectively. Figure 6A,B presents the MSE learning curves and the MIS curves of the employed algorithms. We can observe that, as compared to the LMS algorithm, the IF-LMS algorithm obtained considerably lower steady-state MSE. Also, note that although the convergence speed of the IF-LMS algorithm is lower than that of the LMS algorithm at the beginning of the transient period, the IF-LMS algorithm can reach the LMS algorithm before entering the steady-state.

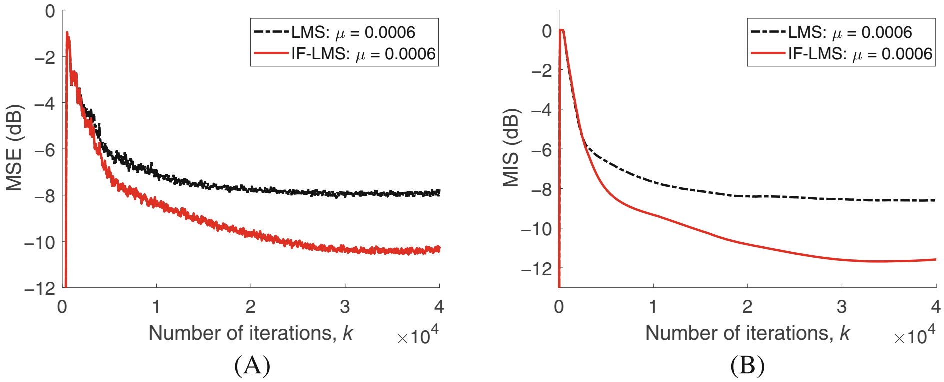

FIGURE 5 Underwater channel estimation: (A) the mean squared error learning curves; (B) the misalignment curves

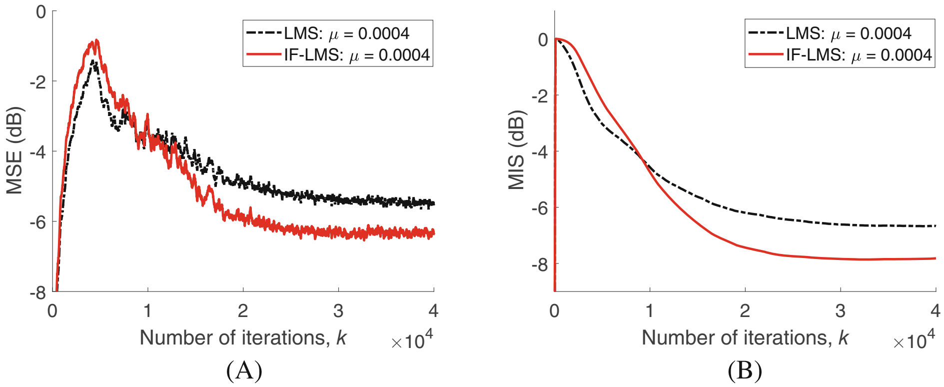

FIGURE 6 Wireless channel estimation: (A) the mean squared error learning curves; (B) the misalignment curves

5 | CONCLUSIONS

In this paper, by introducing a learnable feature matrix, the IF-LMS algorithm has been proposed. This algorithm can exploit hidden sparsity in unknown systems without requiring a priori information about the spectral characteristics of the systems. Moreover, this algorithm can be utilized in all systems that the traditional LMS algorithm is applicable. Also, we have analyzed the step-size parameter of the IF-LMS algorithm and the weight given to the sparsity-promoting penalty function. Furthermore, in simulation results, we have employed the IF-LMS algorithm to identify some synthetic and real-life systems. Finally, we should mention that the proposed idea is not limited to the LMS algorithm and can be extended to other algorithms, such as the normalized LMS, the affine projection, and the recursive least squares. Also, for the sake of computational efficiency, we assumed that only two entries at each row of the feature matrix are nonzero; however, we can suppose all entries to be nonzero and learn them using the SGD approach.

ACKNOWLEDGMENTS

The author would like to thank the São Paulo Research Foundation (FAPESP) grants #2015/22308-2 and #2019/06280-1.

CONFLICT OF INTEREST

The author declares no potential conflict of interests.

ORCID

Hamed Yazdanpanah https://orcid.org/0000-0002-7108-7866

REFERENCES

1. Widrow B, Hoff ME. Adaptive switching circuits. *IRE WESCOM Conv Rec*. 1960;4(TR-1553-1):96-104.
2. Diniz PSR. *Adaptive Filtering: Algorithms and Practical Implementation*. 4th ed. Springer; 2013.
3. Roopa S, Narasimhan SV, Babloo B. Steiglitz–McBride adaptive notch filter based on a variable-step-size LMS algorithm and its application to active noise control. *Int J Adapt Control Signal Process*. 2016;30(1):16-30.
4. Steinhardt AO, Van Veen BD. Adaptive beamforming. *Int J Adapt Control Signal Process*. 1989;3(3):253-281.
5. Duppala VR, Misra IS, Sanyal SK. Design of digital signal processor based adaptive beamformer using hybrid residual least mean square algorithm for improved performance. *Int J RF Microw Comput-Aid Eng*. 2020;30(3):e22073.
6. Salman MS. Sparse leaky-LMS algorithm for system identification and its convergence analysis. *Int J Adapt Control Signal Process*. 2014;28(10):1065-1072.
7. Yazdanpanah H, Diniz PSR. New trinion and quaternion set-membership affine projection algorithms. *IEEE Trans Circuits Syst II Express Briefs*. 2017;64(2):216-220.
8. Diniz PSR, Yazdanpanah H, Lima MVS. Feature LMS algorithms. Proceedings of the IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP 2018); 2018:4144-4148; Calgary, Alberta, Canada.
9. Yazdanpanah H, Diniz PSR, Lima MVS. Feature adaptive filtering: exploiting hidden sparsity. *IEEE Trans Circuits Syst I Regul Pap*. 2020;67(7):2358-2371.
10. Yazdanpanah H, Apolinário J. The extended feature LMS algorithm: exploiting hidden sparsity for systems with unknown spectrum. *Circuits Syst Signal Process*. 2021;40:174-192.
11. Duttweiler D. Proportionate normalized least-mean-squares adaptation in echo cancelers. *IEEE Trans Speech Audio Process*. 2000;8(5):508-518.
12. Benesty J, Gay SL. An improved PNLMS algorithm. Proceedings of the IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP 2002); Vol. 2, 2002:II-1881-II-1884; Orlando, FL.
13. Lima MVS, Ferreira TN, Martins WA, Diniz PSR. Sparsity-aware data-selective adaptive filters. *IEEE Trans Signal Process*. 2014;62(17):4557-4572.
14. Gu Y, Jin J, Mei S. l_0 norm constraint LMS algorithm for sparse system identification. *IEEE Signal Process Lett*. 2009;16(9):774-777.
15. Yazdanpanah H, Carini A, Lima MVS. L_0-norm adaptive Volterra filters. Proceedings of the 27th European Signal Processing Conference (EUSIPCO 2019); 2019:1-5; A Coruña, Spain.
16. Kopsinis Y, Slavakis K, Theodoridis S. Online sparse system identification and signal reconstruction using projections onto weighted l_1 balls. *IEEE Trans Signal Process*. 2011;59(3):936-952.
17. Li Y, Hamamura M. Zero-attracting variable-step-size least mean square algorithms for adaptive sparse channel estimation. *Int J Adapt Control Signal Process*. 2015;29(9):1189-1206.
18. Pan CD, Yu L, Liu HL, Chen ZP, Luo WF. Moving force identification based on redundant concatenated dictionary and weighted l_1-norm regularization. *Mech Syst Signal Process*. 2018;98:32-49.
19. Das RL, Chakraborty M. Improving the performance of the PNLMS algorithm using l_1 norm regularization. *IEEE/ACM Trans Audio Speech Lang Process*. 2016;24(7):1280-1290.

20. Hu T, Chklovskii DB. Sparse LMS via online linearized Bregman iteration. Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2014); 2014:7213-7217; Florence, Italy.
21. Yazdanpanah H, Diniz PSR, Lima MVS. Improved simple set-membership affine projection algorithm for sparse system modeling: analysis and implementation. *IET Signal Process*. 2020;14(2):81-88.
22. Azghani M, Ghorbani A, Marvasti F. Blind iterative nonlinear distortion compensation based on thresholding. *IEEE Trans Circuits Syst II Express Briefs*. 2017;64(7):852-856.
23. Yazdanpanah H, Diniz PSR, Lima MVS. Low-complexity feature stochastic gradient algorithm for block-lowpass systems. *IEEE Access*. 2019;7:141587-141593.
24. Haykin SS. *Adaptive Filter Theory*. 2nd ed. Prentice Hall; 1991.
25. Sayed AH. *Fundamentals of Adaptive Filtering*. John Wiley & Sons; 2003.
26. Kwong RH, Johnston EW. A variable step size LMS algorithm. *IEEE Trans Signal Process*. 1992;40(7):1633-1642.
27. Huang J, Diamant R. Channel impulse responses from March 2019 long range experiment (Mediterranean Sea). *IEEE Dataport*. 2019;2019. doi:10.21227/nzgr-ds72
28. Tekbıyık K, Ekti AR, Kurt GK, Görçin A. Modeling and analysis of sub-terahertz communication channel via mixture of gamma distribution. arXiv preprint arXiv:1912.10420; 2019.

How to cite this article: Yazdanpanah H. Improved feature least mean square algorithm. *Int J Adapt Control Signal Process*. 2022;1-11. doi: 10.1002/acs.3528