A Family of Optimized LMS-Based Algorithms for System Identification

Silviu Ciochină*, Constantin Paleologu*, Jacob Benesty†, Steven L. Grant‡, and Andrei Anghel*
*University Politehnica of Bucharest, Romania
Email: {silviu,pale}@comm.pub.ro, andrei.anghel@munde.pub.ro
†INRS-EMT, University of Quebec, Montreal, Canada, Email: benesty@emt.inrs.ca
‡Missouri University of Science and Technology, Rolla, USA

This paper is dedicated to the memory of Steven L. Grant for his exceptional contributions to the echo cancellation problem.

***Abstract*—The performance of the least-mean-square (LMS) algorithm is governed by its step-size parameter. In this paper, we present a family of optimized LMS-based algorithms (in terms of the step-size control), in the context of system identification. A time-variant system model is considered and the optimization criterion is based on the minimization of the system misalignment. Simulations performed in the context of acoustic echo cancellation indicate that these algorithms achieve a proper compromise in terms of fast convergence/tracking and low misadjustment.**

I. Introduction

The least-mean-square (LMS) algorithm and its variants [such as the normalized LMS (NLMS)] are frequently used for system identification problems [1], [2]. The main parameter that controls the behavior of these algorithms is the step-size. In this context, the variable step-size (VSS) algorithms were designed to achieve a proper compromise between fast convergence rate and low misadjustment, e.g., see [3]– [7] and the references therein.

In the framework of system identification, a natural optimization criterion to follow is the minimization of the system misalignment. However, most of the VSS algorithms were developed assuming that the unknown system is time invariant, which is rarely the case in real-world applications. Consequently, these algorithms could require some additional features to control their behavior, like system change detectors or extra parameters, which are not easy tasks in practice.

In this paper, we present a family of optimized LMS-based algorithms, which act like VSS adaptive filters. First, we consider that the unknown system is variable in time (based on a first-order Markov model) and the minimization of the system misalignment is carried out under this circumstance. Second, we extend the approach to the more general case of an LMS algorithm with individual control factors, which is somehow similar to the idea behind sparse adaptive filters [8]. Simulations performed in the context of acoustic echo cancellation (which is a very challenging system identification problem) indicate the good behavior of these algorithms.

II. System Model

Let us consider a system identification problem, where an adaptive filter is used to model an unknown system, both driven by the same zero-mean input signal $x(n)$. In this context, the reference signal at the discrete-time index n is

$$d(n) = \mathbf{h}^T(n)\mathbf{x}(n) + v(n), \tag{1}$$

where $\mathbf{h}(n) = \begin{bmatrix} h_0(n) & h_1(n) & \cdots & h_{L-1}(n) \end{bmatrix}^T$ is the impulse response (of length L) of the system that we need to identify, superscript ^{T} is the transpose operator, $\mathbf{x}(n) = \begin{bmatrix} x(n) & x(n-1) & \cdots & x(n-L+1) \end{bmatrix}^T$ is a vector containing the L most recent samples of the input signal, and $v(n)$ is the system noise, usually considered as a zero-mean white Gaussian noise signal.

In the following, let us consider that $\mathbf{h}(n)$ follows a simplified first-order Markov model, i.e.,

$$\mathbf{h}(n) = \mathbf{h}(n-1) + \mathbf{w}(n), \tag{2}$$

where $\mathbf{w}(n)$ is a zero-mean white Gaussian noise signal vector, which is uncorrelated with $\mathbf{h}(n-1)$. The correlation matrix of $\mathbf{w}(n)$ is assumed to be $\mathbf{R_w} = \sigma_w^2\mathbf{I}_L$, where $\mathbf{I}_L$ is the $L \times L$ identity matrix and the variance, σ_w^2, captures the uncertainties in $\mathbf{h}(n)$. This model can be valid in many system identification problems, like in acoustic echo cancellation (since the acoustic echo paths are time-variant systems) [9], [10].

The objective is to estimate $\mathbf{h}(n)$ with an adaptive filter, defined by $\widehat{\mathbf{h}}(n) = \begin{bmatrix} \widehat{h}_0(n) & \widehat{h}_1(n) & \cdots & \widehat{h}_{L-1}(n) \end{bmatrix}^T$. It is interesting to notice that equations (1) and (2) define a state variable model, similar to Kalman filtering [11]– [13].

III. An Optimized LMS-Based Algorithm

The LMS algorithm [1], [2] is defined by the update:

$$\widehat{\mathbf{h}}(n) = \widehat{\mathbf{h}}(n-1) + \mu\mathbf{x}(n)e(n), \tag{3}$$

where μ is the step-size parameter and

$$e(n) = d(n) - \widehat{\mathbf{h}}^T(n-1)\mathbf{x}(n) \tag{4}$$

is the error signal of the adaptive filter. Also, let us define the a posteriori misalignment as $\mathbf{c}(n) = \mathbf{h}(n) - \widehat{\mathbf{h}}(n)$, so that, developing (3), the update results in

$$\mathbf{c}(n) = \mathbf{c}(n-1) + \mathbf{w}(n) - \mu\mathbf{x}(n)e(n). \tag{5}$$

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

Moreover, based on (1) and (2), the error signal from (4) can be rewritten as

$$e(n) = \mathbf{x}^T(n)\mathbf{c}(n-1) + \mathbf{x}^T(n)\mathbf{w}(n) + v(n). \tag{6}$$

Since we deal with a system identification problem, a reasonable optimization criterion to follow is the minimization of the system misalignment. Therefore, taking the ℓ_2 norm in (5), then mathematical expectation on both sides, we obtain

$$\begin{aligned} &E\left[\|\mathbf{c}(n)\|_2^2\right] = E\left[\|\mathbf{c}(n-1)\|_2^2\right] + L\sigma_w^2 \\ &- 2\mu E\left[\mathbf{x}^T(n)\mathbf{c}(n-1)e(n)\right] - 2\mu E\left[\mathbf{x}^T(n)\mathbf{w}(n)e(n)\right] \\ &+ \mu^2 E\left[e^2(n)\mathbf{x}^T(n)\mathbf{x}(n)\right]. \end{aligned} \tag{7}$$

Next, taking (6) into account, assuming that the input signal is a zero-mean Gaussian process with the variance σ_x^2, and removing the uncorrelated products, the first cross-correlation term from the second line of (7) results in

$$\begin{aligned} &E\left[\mathbf{c}^T(n-1)\mathbf{x}(n)e(n)\right] \\ &\approx \mathrm{tr}\left\{E\left[\mathbf{c}(n-1)\mathbf{c}^T(n-1)\right] E\left[\mathbf{x}(n)\mathbf{x}^T(n)\right]\right\} \\ &= \sigma_x^2 E\left[\|\mathbf{c}(n-1)\|_2^2\right], \end{aligned} \tag{8}$$

where $\mathrm{tr}\{\cdot\}$ denotes the trace of a matrix. Similarly, the last cross-correlation term from the second line of (7) can be developed as

$$\begin{aligned} &E\left[\mathbf{w}^T(n)\mathbf{x}(n)e(n)\right] \\ &\approx \mathrm{tr}\left\{E\left[\mathbf{w}(n)\mathbf{w}^T(n)\right] E\left[\mathbf{x}(n)\mathbf{x}^T(n)\right]\right\} = L\sigma_x^2\sigma_w^2 \end{aligned} \tag{9}$$

and, finally, the last term of (7) results in

$$\begin{aligned} &E\left[e^2(n)\mathbf{x}^T(n)\mathbf{x}(n)\right] = \mathrm{tr}\left\{E\left[v^2(n)\mathbf{x}(n)\mathbf{x}^T(n)\right]\right\} \\ &+ \mathrm{tr}\left\{E\left\{\left[\mathbf{c}^T(n-1)\mathbf{x}(n) + \mathbf{w}^T(n)\mathbf{x}(n)\right]^2 \mathbf{x}(n)\mathbf{x}^T(n)\right\}\right\}. \end{aligned} \tag{10}$$

The last term in (10) can be developed based on the Gaussian moment factoring theorem [1] (i.e., the Isserlis' theorem). Thus, after several computations, (10) becomes

$$\begin{aligned} &E\left[e^2(n)\mathbf{x}^T(n)\mathbf{x}(n)\right] \\ &= L\sigma_x^2\sigma_v^2 + (L+2)\sigma_x^4\left\{E\left[\|\mathbf{c}(n-1)\|_2^2\right] + L\sigma_w^2\right\}, \end{aligned} \tag{11}$$

where $\sigma_v^2 = E\left[v^2(n)\right]$ is the variance of the additive noise.

Denoting $m(n) = E\left[\|\mathbf{c}(n)\|_2^2\right]$ and introducing (8), (9), and (11) in (7), we obtain

$$\begin{aligned} m(n) &= \left[1 - 2\mu\sigma_x^2 + (L+2)\mu^2\sigma_x^4\right] m(n-1) \\ &+ L\mu^2\sigma_x^2\left[\sigma_v^2 + (L+2)\sigma_x^2\sigma_w^2\right] - 2L\mu\sigma_x^2\sigma_w^2 + L\sigma_w^2. \end{aligned} \tag{12}$$

At this point, let us consider that the step-size μ is time dependent, in order to evaluate $\partial m(n)/\partial\mu(n) = 0$. This leads to an optimal step-size:

$$\mu(n) = \frac{1}{(L+2)\sigma_x^2 + \xi}, \tag{13}$$

where $\xi = L\sigma_v^2/\left[m(n-1) + L\sigma_w^2\right]$. So, the filter update is

$$\widehat{\mathbf{h}}(n) = \widehat{\mathbf{h}}(n-1) + \mu(n)\mathbf{x}(n)e(n). \tag{14}$$

To update the parameter $m(n)$, the step-size from (13) is introduced in (12) and, after some computations, it results in

$$m(n) = \left[1 - \mu(n)\sigma_x^2\right]\left[m(n-1) + L\sigma_w^2\right]. \tag{15}$$

Summarizing, the resulting optimized LMS-based algorithm is defined by the relations (4), (14), and (15), using the initialization $\widehat{\mathbf{h}}(0) = \mathbf{0}$ and $m(0) = \varepsilon$ (where ε is a positive constant). This algorithm is very similar to the simplified Kalman filter from [13], which was obtained as an approximation of the general Kalman filter; similar works can be found in [14] and [15]. Also, the resulting algorithm can be obtained following the line of the NLMS algorithm and a joint-optimization problem on both the normalized step-size and regularization parameters [7]. For the sake of consistency, we will refer this algorithm as the joint-optimized NLMS (JO-NLMS).

IV. An Optimized LMS-Based Algorithm with Individual Control Factors

In this section, we extend the previous approach to an LMS-based algorithm with individual control factors. This algorithm is defined by the update [16]:

$$\widehat{\mathbf{h}}(n) = \widehat{\mathbf{h}}(n-1) + \mu\mathbf{K}(n-1)\mathbf{x}(n)e(n), \tag{16}$$

where $\mathbf{K}(n)$ is a diagonal matrix ($L\times L$) containing the control factors at time index n. The filter update (16) can be also rewritten in terms of the a posteriori misalignment as

$$\mathbf{c}(n) = \mathbf{c}(n-1) + \mathbf{w}(n) - \mu\mathbf{K}(n-1)\mathbf{x}(n)e(n). \tag{17}$$

Next, taking the ℓ_2 norm in (17), then mathematical expectation on both sides (also removing the uncorrelated products), we obtain

$$\begin{aligned} E\left[\|\mathbf{c}(n)\|_2^2\right] &= E\left[\|\mathbf{c}(n-1)\|_2^2\right] + L\sigma_w^2 \\ &- 2\mu E\left[\mathbf{x}^T(n)\mathbf{K}(n-1)\mathbf{w}(n)e(n)\right] \\ &- 2\mu E\left[\mathbf{x}^T(n)\mathbf{K}(n-1)\mathbf{c}(n-1)e(n)\right] \\ &+ \mu^2 E\left[e^2(n)\mathbf{x}^T(n)\mathbf{K}^2(n-1)\mathbf{x}(n)\right]. \end{aligned} \tag{18}$$

At this point, we need to process the last three terms in (18). Based on (6) and assuming that the input signal is white, the first cross-correlation term becomes

$$\begin{aligned} &E\left[\mathbf{x}^T(n)\mathbf{K}(n-1)\mathbf{w}(n)e(n)\right] \\ &= E\left\{\mathrm{tr}\left[\mathbf{w}(n)\mathbf{w}^T(n)\mathbf{K}(n-1)\mathbf{x}(n)\mathbf{x}^T(n)\right]\right\} \\ &= \sigma_w^2\sigma_x^2\mathrm{tr}\left[\mathbf{K}(n-1)\right] \end{aligned} \tag{19}$$

and the second cross-correlation term results in

$$\begin{aligned} &E\left[\mathbf{x}^T(n)\mathbf{K}(n-1)\mathbf{c}(n-1)e(n)\right] \\ &= \mathrm{tr}\left\{E\left[\mathbf{c}(n-1)\mathbf{c}^T(n-1)\right] E\left[\mathbf{x}(n)\mathbf{x}^T(n)\right]\mathbf{K}(n-1)\right\} \\ &= \sigma_x^2\mathrm{tr}\left[\mathbf{R}_{\mathbf{c}}(n-1)\mathbf{K}(n-1)\right], \end{aligned} \tag{20}$$

where $\mathbf{R}_{\mathbf{c}}(n) = E\left[\mathbf{c}(n)\mathbf{c}^T(n)\right]$.

The last term in (18) is more difficult to process. However, for a certain stationarity degree of the input signal and for large values of L ($L \gg 1$), we can treat the term $\|\mathbf{K}(n-1)\mathbf{x}(n)\|_2^2$ as a deterministic quantity. Also, we can use the orthogonality principle, i.e., assuming that the adaptive filter has converged to a certain degree. Using these assumptions, the last term in (18) can be developed as

$$\begin{aligned} &E\left[e^2(n)\mathbf{x}^T(n)\mathbf{K}^2(n-1)\mathbf{x}(n)\right] \\ &= \mathrm{tr}\left\{E\left[\mathbf{x}(n)\mathbf{x}^T(n)\right]\mathbf{K}^2(n-1)\right\}E\left[e^2(n)\right] \\ &= \sigma_x^2 \mathrm{tr}\left[\mathbf{K}^2(n-1)\right]\left[\sigma_v^2 + L\sigma_w^2\sigma_x^2 + \sigma_x^2 m(n-1)\right]. \quad (21)\end{aligned}$$

Finally, using (19)–(21) in (18), it results in

$$\begin{aligned} m(n) &= m(n-1) + L\sigma_w^2 \quad (22) \\ &- 2\mu\sigma_x^2 \mathrm{tr}\left\{\left[\mathbf{R_c}(n-1) + \sigma_w^2\mathbf{I}_L\right]\mathbf{K}(n-1)\right\} \\ &+ \mu^2\sigma_x^2 \mathrm{tr}\left[\mathbf{K}^2(n-1)\right]\left[\sigma_v^2 + L\sigma_w^2\sigma_x^2 + \sigma_x^2 m(n-1)\right].\end{aligned}$$

As in Section III, we can impose $\partial m(n)/\partial\mu(n) = 0$ (considering that the step-size parameter is time dependent), to obtain an optimal step-size:

$$\mu(n) = \frac{\mathrm{tr}\left\{\left[\mathbf{R_c}(n-1) + \sigma_w^2\mathbf{I}_L\right]\mathbf{K}(n-1)\right\}}{\mathrm{tr}\left[\mathbf{K}^2(n-1)\right]\left\{\sigma_v^2 + \sigma_x^2\left[m(n-1) + L\sigma_w^2\right]\right\}}. \quad (23)$$

Also, using (23) in (22), the update of $m(n)$ becomes

$$\begin{aligned} m(n) &= m(n-1) + L\sigma_w^2 \\ &- \frac{\sigma_x^2\left\{\mathrm{tr}\left\{\left[\mathbf{R_c}(n-1) + \sigma_w^2\mathbf{I}_L\right]\mathbf{K}(n-1)\right\}\right\}^2}{\mathrm{tr}\left[\mathbf{K}^2(n-1)\right]\left\{\sigma_v^2 + \sigma_x^2\left[m(n-1) + L\sigma_w^2\right]\right\}}. \quad (24)\end{aligned}$$

Let us focus on the main term of the numerator in (24). First, we introduce the notation:

$$\begin{aligned} \mathbf{\Gamma}(n) &= \mathbf{R_c}(n) + \sigma_w^2\mathbf{I}_L, \\ \boldsymbol{\gamma}(n) &= \left[\begin{array}{cccc} \gamma_0(n) & \gamma_1(n) & \cdots & \gamma_{L-1}(n) \end{array}\right]^T, \\ \mathbf{k}(n) &= \left[\begin{array}{cccc} k_0(n) & k_1(n) & \cdots & k_{L-1}(n) \end{array}\right]^T,\end{aligned}$$

where $\boldsymbol{\gamma}(n)$ and $\mathbf{k}(n)$ are two vectors containing the diagonal elements of $\mathbf{\Gamma}(n)$ and $\mathbf{K}(n)$, respectively. Since $\mathbf{K}(n)$ is a diagonal matrix, we can use the Cauchy-Schwarz inequality:

$$\left\{\mathrm{tr}\left[\mathbf{\Gamma}(n-1)\mathbf{K}(n-1)\right]\right\}^2 \le \|\boldsymbol{\gamma}(n-1)\|_2^2\,\|\mathbf{k}(n-1)\|_2^2. \quad (25)$$

It is known that the equality in (25) is obtained if and only if $\boldsymbol{\gamma}(n-1)$ and $\mathbf{k}(n-1)$ are linearly dependent, i.e., $\mathbf{k}(n-1) = q\boldsymbol{\gamma}(n-1)$, with $q > 0$. Hence,

$$\begin{aligned} \mathrm{tr}\left[\mathbf{K}(n-1)\right] &= q\,\mathrm{tr}\left[\mathbf{\Gamma}(n-1)\right] = q\left[m(n-1) + L\sigma_w^2\right], \\ \mathrm{tr}\left[\mathbf{K}^2(n-1)\right] &= \|\mathbf{k}(n-1)\|_2^2 = q^2\,\|\boldsymbol{\gamma}(n-1)\|_2^2. \quad (26)\end{aligned}$$

Consequently, the optimal step-size from (23) results in

$$\mu(n) = \frac{1}{q\left\{\sigma_v^2 + \sigma_x^2\left[m(n-1) + L\sigma_w^2\right]\right\}} \quad (27)$$

and the update of the parameter $m(n)$ from (24) becomes

$$m(n) = m(n-1) + L\sigma_w^2 - q\mu(n)\sigma_x^2\,\|\boldsymbol{\gamma}(n-1)\|_2^2. \quad (28)$$

Finally, the filter update from (16) is evaluated as

$$\widehat{\mathbf{h}}(n) = \widehat{\mathbf{h}}(n-1) + q\mu(n)\boldsymbol{\gamma}(n-1)\odot\mathbf{x}(n)\mathbf{e}(n), \quad (29)$$

where $\odot$ denotes the Hadamard product.

However, we still need to find the value of q and an update relation for $\boldsymbol{\gamma}(n)$. The parameter q is related to the value of $\mathrm{tr}\left[\mathbf{K}(n-1)\right]$, as shown in (26). In the classical algorithm, the trace of this matrix is equal to L [since $\mathbf{K}(n-1)$ is replaced by $\mathbf{I}_L$]. In our case, we could impose $\mathrm{tr}\left[\mathbf{K}(n-1)\right] = L$, but the gains will be individually distributed through the elements of $\boldsymbol{\gamma}(n-1)$. Consequently, we could evaluate

$$q = \frac{L}{m(n-1) + L\sigma_w^2}. \quad (30)$$

The vector $\boldsymbol{\gamma}(n)$ contains the diagonal elements of the matrix $\mathbf{\Gamma}(n)$. But this matrix depends on $\mathbf{R_c}(n)$. It is interesting to notice that the elements of $\boldsymbol{\gamma}(n)$, i.e., the individual control factors, depend on the coefficients' misalignment (as compared to the classical proportionate-type algorithms, where the adaptation is controlled depending on the coefficients' magnitude). Next, based on (17), we can write

$$\begin{aligned} \mathbf{R_c}(n) &= \mathbf{R_c}(n-1) + \sigma_w^2\mathbf{I}_L \\ &- \mu E\left[\mathbf{w}(n)\mathbf{x}^T(n)\mathbf{K}(n-1)e(n)\right] \\ &- \mu E\left[\mathbf{K}(n-1)\mathbf{x}(n)\mathbf{w}^T(n)e(n)\right] \\ &- \mu E\left[\mathbf{c}(n-1)\mathbf{x}^T(n)\mathbf{K}(n-1)e(n)\right] \\ &- \mu E\left[\mathbf{K}(n-1)\mathbf{x}(n)\mathbf{c}^T(n-1)e(n)\right] \\ &+ \mu^2 E\left[e^2(n)\mathbf{K}(n-1)\mathbf{x}(n)\mathbf{x}^T(n)\mathbf{K}(n-1)\right]. \quad (31)\end{aligned}$$

Following a similar approach we used to process (18), we could further develop (31), in order to finally obtain [also using the optimal step-size from (27)]:

$$\begin{aligned} \boldsymbol{\gamma}(n) &= \boldsymbol{\gamma}(n-1) + \sigma_w^2\mathbf{1}_{L\times 1} \\ &+ q\mu(n)\sigma_x^2\left\{\sigma_v^2 + \sigma_x^2\left[m(n-1) + L\sigma_w^2\right] - 2q\mu(n)\right\} \\ &\times \boldsymbol{\gamma}(n-1)\odot\boldsymbol{\gamma}(n-1), \quad (32)\end{aligned}$$

where $\mathbf{1}_{L\times 1}$ denotes a column vector with all its L elements equal to one. Due to stability reasons, a normalization should be also performed in this step, i.e., $\overline{\gamma}_i(n) = \gamma_i(n)/\max\left[q\boldsymbol{\gamma}(n)\right]$, with $0 \le i \le L-1$. Summarizing, the resulting optimized LMS algorithm with individual control factors (OLMS-ICF) is defined by the relations (27)–(30) and (32) (including the normalization).

V. Practical Considerations

There are three main parameters that should be set or estimated within both JO-NLMS and OLMS-ICF algorithms. The first one is the variance of the input signal, which could be easily evaluated as in the NLMS algorithm, i.e., $\widehat{\sigma}_x^2(n) = \mathbf{x}^T(n)\mathbf{x}(n)/L$. The second parameter is the noise power, σ_v^2. This parameter can be estimated in different ways; for example, in echo cancellation, it can be estimated during silences of the near-end talker [4]. Also, other practical methods to estimate σ_v^2 can be found in [17], [18].

The third parameter to be found is σ_w^2, which is a specific one. In practice, we propose to estimate σ_w^2 as in [13], by taking the ℓ_2 norm on both sides of (2) and replacing $\mathbf{h}(n)$ by its estimate $\widehat{\mathbf{h}}(n)$, thus resulting

$$\widehat{\sigma}_w^2(n) = \frac{1}{L}\left\|\widehat{\mathbf{h}}(n) - \widehat{\mathbf{h}}(n-1)\right\|_2^2. \quad (33)$$

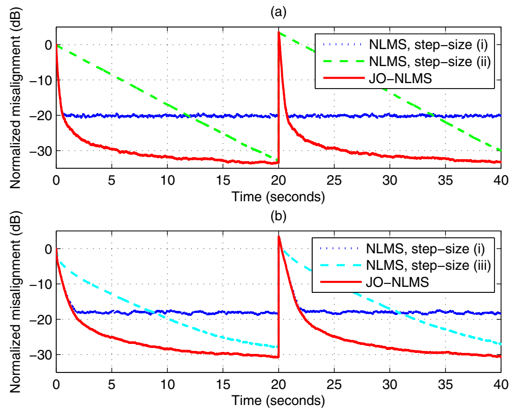

Fig. 1. Misalignment of the NLMS algorithm using different step-sizes, i.e., (i) $1/[\delta + \mathbf{x}^T(n)\mathbf{x}(n)]$, (ii) $0.025/[\delta + \mathbf{x}^T(n)\mathbf{x}(n)]$, and (iii) $0.15/[\delta + \mathbf{x}^T(n)\mathbf{x}(n)]$, and misalignment of the JO-NLMS algorithm. Echo path changes after 20 seconds, $L = 1000$, and SNR = 20 dB. The input signal is (a) white Gaussian and (b) an AR(1) process.

The difference from the right-hand side of (33) represents the update term of the algorithm, which can be used to save L subtractions associated with the direct evaluation of (33).

VI. Simulation Results

Simulations are performed in an acoustic echo cancellation scenario [9], [10]. In this context, an adaptive filter is used to estimate the impulse response (i.e., the acoustic echo path) between the loudspeaker and the microphone of a hands-free communication device. This is a very challenging system identification problem, due to the high length and time-varying nature of the acoustic echo paths.

The length of the acoustic impulse response used in the experiments is $L = 1000$ (the sampling rate is 8 kHz); the same length is set for the adaptive filter. The input signal, $x(n)$, is either a white Gaussian noise, an AR(1) process generated by filtering a white Gaussian noise through a first-order system $1/\left(1 - 0.8z^{-1}\right)$, or a speech sequence. The output of the echo path is corrupted by an independent white Gaussian noise $v(n)$ and the signal-to-noise ratio (SNR) is set to 20 dB. In our simulations, we assume that σ_v^2 is known; in practice, it can be estimated like in [4], [17], [18].

In the first set of experiments, we evaluate the VSS features of the JO-NLMS algorithm. An echo path change scenario is simulated, by shifting the impulse response to the right by 12 samples, in the middle of the simulation. The results are presented in Figs. 1-3, using a white Gaussian noise or an AR(1) process as the input. The NLMS algorithm (with different step-sizes and including a regularization parameter $\delta = 20\sigma_x^2$) is used for comparison.

In Fig. 1, the performance measure is the normalized misalignment (in dB), defined as $20\log_{10} \left\| \mathbf{h}(n) - \widehat{\mathbf{h}}(n) \right\|_2 / \left\| \mathbf{h}(n) \right\|_2$. As we can notice from this figure, the JO-NLMS algorithm has a fast convergence rate and tracking, similar to the NLMS algorithm using the largest step-size, while achieving a lower misalignment level.

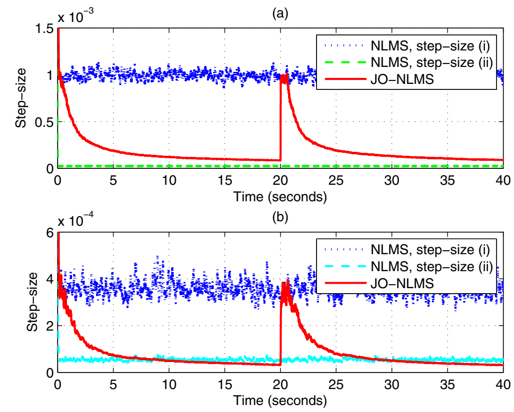

Fig. 2. Evolution of the step-size parameters of the NLMS and JO-NLMS algorithms. Other conditions same as in Fig. 1.

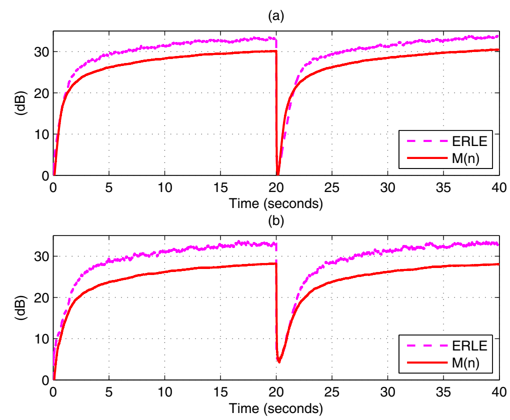

Fig. 3. Evolution of the parameter $M(n) = \left\| \widehat{\mathbf{h}}(n) \right\|_2^2 / m(n)$ and the ERLE of the JO-NLMS algorithm. Other conditions same as in Fig. 1.

This behavior is also supported in Fig. 2, where the evolution of the step-size parameters is presented. It can be noticed that the step-size of the JO-NLMS algorithm goes to the fastest convergence mode in the initial convergence phase and when the system changes, while its value decreases in the steady-state.

An interesting parameter of the JO-NLMS algorithm is $m(n)$. This can be used to evaluate (in an online manner) the overall performance of the algorithm. In echo cancellation, a well-known performance measure is the echo-return loss enhancement (ERLE) [9], which requires the a priori knowledge of the impulse response $\mathbf{h}(n)$ or the echo signal. In the case of the JO-NLMS algorithm, we propose to monitor the parameter $M(n) = \left\| \widehat{\mathbf{h}}(n) \right\|_2^2 / m(n)$, which should have a similar significance with the ERLE. This issue is supported in Fig. 3, where the evolution of the ERLE and $M(n)$ (both in dB) is depicted. However, contrary to the ERLE, the parameter $M(n)$ could be evaluated within the JO-NLMS algorithm.

In the second set of experiments, the input signal is a speech sequence. In Fig. 4, the JO-NLMS and OLMS-ICF algorithms

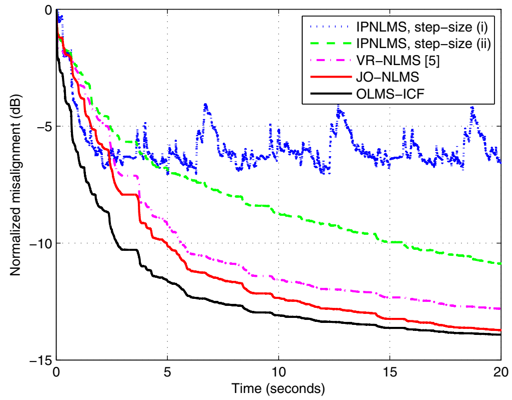

Fig. 4. Misalignment of the IPNLMS algorithm using different step-sizes, i.e., (i) $1/[\delta'+\mathbf{x}^T(n)\mathbf{x}(n)]$ and (ii) $0.05/[\delta'+\mathbf{x}^T(n)\mathbf{x}(n)]$, and misalignment of the VR-NLMS [5], JO-NLMS, and OLMS-ICF algorithms. The input signal is speech, $L = 1000$, and SNR = 20 dB.

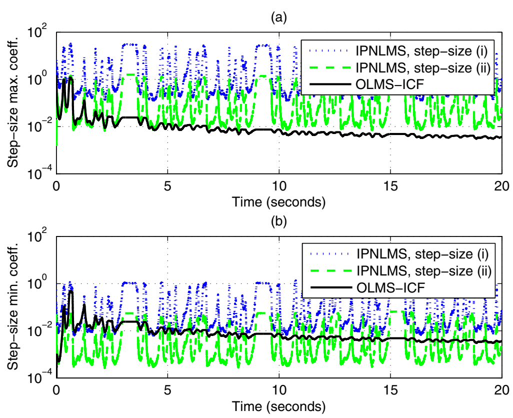

Fig. 5. Evolution of the step-size parameters of the IPNLMS and OLMS-ICF algorithms, corresponding to (a) the maximum coefficient and (b) the minimum coefficient (in terms of magnitude). Other conditions same as in Fig. 4.

are compared to the improved proportionate NLMS (IPNLMS) algorithm [19] (with different step-sizes and including a regularization parameter $\delta' = 20\sigma_x^2/L$). Also, we introduce for comparison the variable-regularized (VR) affine projection algorithm proposed in [5], using a projection order equal to one, which is equivalent to a VR-NLMS algorithm. As we can notice, the OLMS-ICF outperforms the other algorithms.

It would be interesting to observe the individual step-sizes of the OLMS-ICF algorithm. In Fig. 5, we depicted the evolution of the individual step-sizes of two coefficients, i.e., the maximum and the minimum one (in terms of magnitude), comparing the IPNLMS and OLMS-ICF algorithms. As expected, the IPNLMS algorithm allocates a significant gain to the largest coefficient. In the OLMS-ICF algorithm, this difference is valid only in the initial convergence phase, since the misalignments of the individual coefficients [i.e., the elements of $\boldsymbol{\gamma}(n)$] tend to become similar in the steady-state.

VII. Conclusions

In this paper, we have presented a family of optimized LMS-based algorithms in the context of a state variable model for system identification. The optimization criterion is the minimization of the system misalignment, which is a natural approach in system identification problems. The resulting algorithms do not require additional features to monitor their behavior (e.g., system change detectors, stability thresholds, etc.), thus being reliable candidates for real-world applications. Simulations performed in the context of acoustic echo cancellation support the theoretical findings and indicate the good performance of these algorithms.

Acknowledgment

This work was supported by the UEFISCDI under Grants PN-II-RU-TE-2014-4-1880 and PN-II-ID-PCE-2011-3-0097.

References

[1] A. H. Sayed, *Adaptive Filters*. New York, NY: Wiley, 2008.
[2] P. S. R. Diniz, *Adaptive Filtering: Algorithms and Practical Implementation*. Fourth Edition, New York, NY: Springer-Verlag, 2013.
[3] H.-C. Shin, A. H. Sayed, and W.-J. Song, "Variable step-size NLMS and affine projection algorithms," *IEEE Signal Processing Lett.*, vol. 11, pp. 132–135, Feb. 2004.
[4] J. Benesty, H. Rey, L. Rey Vega, and S. Tressens, "A nonparametric VSS-NLMS algorithm," *IEEE Signal Processing Lett.*, vol. 13, pp. 581–584, Oct. 2006.
[5] H. Rey, L. Rey Vega, S. Tressens, and J. Benesty, "Variable explicit regularization in affine projection algorithm: robustness issues and optimal choice," *IEEE Trans. Signal Processing*, vol. 55, pp. 2096–2108, May 2007.
[6] S. Werner, J. A. Apolinário Jr., and P. S. R. Diniz, "Set-membership proportionate affine projection algorithms," *EURASIP J. Audio, Speech, Music Processing*, vol. 2007, no. 1, pp. 1–10, 2007.
[7] S. Ciochină, C. Paleologu, and J. Benesty, "An optimized NLMS algorithm for system identification," *Signal Processing*, vol. 118, pp. 115–121, Jan. 2016.
[8] C. Paleologu, J. Benesty, and S. Ciochină, *Sparse Adaptive Filters for Echo Cancellation*. Morgan & Claypool Publishers, Synthesis Lectures on Speech and Audio Processing, 2010.
[9] J. Benesty, T. Gaensler, D. R. Morgan, M. M. Sondhi, and S. L. Gay, *Advances in Network and Acoustic Echo Cancellation*. Berlin, Germany: Springer-Verlag, 2001.
[10] G. Enzner, H. Buchner, A. Favrot, and F. Kuech, "Acoustic echo control," in *Academic Press Library in Signal Processing*, vol. 4, ch. 30, pp. 807–877, Academic Press 2014.
[11] R. E. Kalman, "A new approach to linear filtering and prediction problems," *J. Basic Engineering*, vol. 82, pp. 35–45, Mar. 1960.
[12] A. H. Sayed and T. Kailath, "A state-space approach to adaptive RLS filtering," *IEEE Signal Processing Magazine*, vol. 11, pp. 18–60, July 1994.
[13] C. Paleologu, J. Benesty, and S. Ciochină, "Study of the general Kalman filter for echo cancellation," *IEEE Trans. Audio, Speech, Language Process.*, vol. 21, pp. 1539–1549, Aug. 2013.
[14] P. A. C. Lopes and J. B. Gerald, "New normalized LMS algorithms based on the Kalman filter," in *Proc. IEEE ISCAS*, 2007, pp. 117–120.
[15] G. Enzner, "Bayesian inference model for applications of time-varying acoustic system identification," in *Proc. EUSIPCO*, 2010, pp. 2126–2130.
[16] S. Ciochină, C. Paleologu, J. Benesty, and S. L. Grant, "An optimized proportionate adaptive algorithm for sparse system identification," in Proc. *IEEE Asilomar*, 2015, 5 pages.
[17] C. Paleologu, S. Ciochină, and J. Benesty, "Variable step-size NLMS algorithm for under-modeling acoustic echo cancellation," *IEEE Signal Processing Lett.*, vol. 15, pp. 5–8, 2008.
[18] M. A. Iqbal and S. L. Grant, "Novel variable step size NLMS algorithms for echo cancellation," in *Proc. IEEE ICASSP*, 2008, pp. 241–244.
[19] J. Benesty and S. L. Gay, "An improved PNLMS algorithm," in *Proc. IEEE ICASSP*, 2002, pp. II-1881–II-1884.