

An NLMS Algorithm for the Identification of Bilinear Forms

Constantin Paleologu*, Jacob Benesty†, and Silviu Ciochină*

*University Politehnica of Bucharest, Romania, Email: {pale,silviu}@comm.pub.ro

†INRS-EMT, University of Quebec, Montreal, Canada, Email: benesty@emt.inrs.ca

Abstract—In a recent work, we addressed the identification problem of bilinear forms with the Wiener filter. In this context, a different approach was introduced, by defining the bilinear term with respect to the impulse responses of a spatiotemporal model, which resembles a multiple-input/single-output (MISO) system. However, in practice, the Wiener filter may not be always very efficient or convenient to use. Consequently, in this paper, we further develop a normalized least-mean-square (NLMS) adaptive filter tailored for bilinear forms. Many simulations, which are performed from a system identification perspective, indicate the good performance of the proposed algorithm.

I. INTRODUCTION

In general, the bilinear forms are related to the approximation of nonlinear systems (via a finite sum of the Volterra series expansion between the inputs and outputs of the system). In this framework, the bilinear systems behave similarly (to some extent) to linear models, which further simplify the analysis. Since the bilinear models can approximate a large class of nonlinear systems, they were involved in a wide range of applications, e.g., [1]– [9].

In most of these works, the bilinear term is defined with respect to the data, i.e., in terms of an input-output relation. Recently, we focused on a different approach by defining the bilinear term with respect to the impulse responses of a spatiotemporal model [10], in the context of multiple-input/single-output (MISO) systems. Similar frameworks can be found in [4], [11], [12], in the context of particular applications, e.g., channel equalization and nonlinear acoustic echo cancellation. In [10], the problem was addressed from a system identification perspective and two forms of the Wiener filter (namely direct and iterative) were developed. Since the Wiener filter may not be always convenient to use in practice, we propose in this paper an adaptive filtering approach based on the normalized least-mean-square (NLMS) algorithm. Simulation results (in the context of system identification) indicate the advantages of the proposed solution.

The rest of the paper is organized as follows. In Section II, the system model is introduced. The direct and iterative Wiener filters (developed in [10]) are summarized in Section III. Next, Section IV is dedicated to the proposed NLMS algorithm for bilinear forms. Simulation results are presented in Section V. Finally, Section VI concludes this work.

II. SIGNAL MODEL WITH BILINEAR FORMS

Let us consider the following signal model:

$$d(t) = \mathbf{h}^T \mathbf{X}(t) \mathbf{g} + w(t) = y(t) + w(t), \quad (1)$$

where $d(t)$ is the zero-mean desired (or reference) signal at the discrete-time index t , \mathbf{h} and \mathbf{g} are the two impulse responses of the system of lengths L and M , respectively, the superscript T is the transpose operator, $\mathbf{X}(t) = [\mathbf{x}_1(t) \ \mathbf{x}_2(t) \ \cdots \ \mathbf{x}_M(t)]$ is the zero-mean multiple-input signal matrix of size $L \times M$, $\mathbf{x}_m(t) = [x_m(t) \ x_m(t-1) \ \cdots \ x_m(t-L+1)]^T$ is a vector containing the most recent L samples of the m th ($m = 1, 2, \dots, M$) input signal, $y(t) = \mathbf{h}^T \mathbf{X}(t) \mathbf{g}$ is the bilinear form, and $w(t)$ is the zero-mean additive noise. It is assumed that all the data is real valued and $\mathbf{X}(t)$ and $w(t)$ are uncorrelated.

The impulse responses \mathbf{h} and \mathbf{g} correspond to the temporal and spatial parts of the system, respectively. It is easy to verify that $y(t)$ is bilinear in \mathbf{h} and \mathbf{g} since, for every fixed \mathbf{h} , it is a linear function of \mathbf{g} and for every fixed \mathbf{g} , it is a linear function of \mathbf{h} [13].

Based on the vectorization operation (i.e., converting a matrix into a vector [13]), the matrix $\mathbf{X}(t)$ of size $L \times M$ can be rewritten as an input vector of length ML :

$$\text{vec}[\mathbf{X}(t)] = [\mathbf{x}_1^T(t) \ \mathbf{x}_2^T(t) \ \cdots \ \mathbf{x}_M^T(t)]^T = \tilde{\mathbf{x}}(t). \quad (2)$$

Therefore, the output signal $y(t)$ can be expressed as

$$\begin{aligned} y(t) &= \mathbf{h}^T \mathbf{X}(t) \mathbf{g} = \text{tr} \left[(\mathbf{h} \mathbf{g}^T)^T \mathbf{X}(t) \right] \\ &= \text{vec}^T(\mathbf{h} \mathbf{g}^T) \text{vec}[\mathbf{X}(t)] = (\mathbf{g} \otimes \mathbf{h})^T \tilde{\mathbf{x}}(t) \\ &= \mathbf{f}^T \tilde{\mathbf{x}}(t), \end{aligned} \quad (3)$$

where $\text{tr}[\cdot]$ denotes the trace of a square matrix, \otimes is the Kronecker product, and $\mathbf{f} = \mathbf{g} \otimes \mathbf{h}$ is the spatiotemporal impulse response (of length ML), which is simply the Kronecker product between the two individual impulse responses \mathbf{g} and \mathbf{h} . As a consequence, the signal model in (1) is also

$$d(t) = \mathbf{f}^T \tilde{\mathbf{x}}(t) + w(t), \quad (4)$$

which can be seen as a particular form of a MISO system. In the general case of a MISO system, \mathbf{f} has ML different elements while in this bilinear context, $\mathbf{f} = \mathbf{g} \otimes \mathbf{h}$ is formed with $M+L$ different elements only even though it is of length ML .

In this context, the main objective is to identify the temporal and spatial impulse responses \mathbf{h} and \mathbf{g} with two filters $\hat{\mathbf{h}}$ and $\hat{\mathbf{g}}$ of lengths L and M , respectively, and the spatiotemporal impulse response $\mathbf{f} = \mathbf{g} \otimes \mathbf{h}$ with the long filter $\hat{\mathbf{f}} = \hat{\mathbf{g}} \otimes \hat{\mathbf{h}}$ of length ML . Let $\eta \neq 0$ be a real-valued number. It is clear from (1) that $(\eta\mathbf{h})^T \mathbf{X}(t) (\mathbf{g}/\eta) = \mathbf{h}^T \mathbf{X}(t) \mathbf{g} = y(t)$, so that the pair $\eta\mathbf{h}$ and \mathbf{g}/η is equivalent to the pair \mathbf{h} and \mathbf{g} in the bilinear form. This implies that we can only identify $\hat{\mathbf{h}}$ and $\hat{\mathbf{g}}$ up to a scaling factor. A similar discussion can be found in [4] in the context of blind identification/equalization. However, since $\mathbf{f} = \mathbf{g} \otimes \mathbf{h} = (\mathbf{g}/\eta) \otimes (\eta\mathbf{h})$, the spatiotemporal impulse response will be identified with no scaling ambiguity. Therefore, to evaluate the identification of the temporal and spatial filters, we should use the normalized projection misalignment (NPM) [14]:

$$\text{NPM}(\mathbf{h}, \hat{\mathbf{h}}) = 1 - \left(\frac{\mathbf{h}^T \hat{\mathbf{h}}}{\|\mathbf{h}\| \|\hat{\mathbf{h}}\|} \right)^2, \quad (5)$$

$$\text{NPM}(\mathbf{g}, \hat{\mathbf{g}}) = 1 - \left(\frac{\mathbf{g}^T \hat{\mathbf{g}}}{\|\mathbf{g}\| \|\hat{\mathbf{g}}\|} \right)^2, \quad (6)$$

and the identification of the spatiotemporal filter should be evaluated with the usual normalized misalignment (NM):

$$\text{NM}(\mathbf{f}, \hat{\mathbf{f}}) = \frac{\|\mathbf{f} - \hat{\mathbf{f}}\|^2}{\|\mathbf{f}\|^2}, \quad (7)$$

where $\|\cdot\|$ denotes the Euclidean norm.

III. IDENTIFICATION OF BILINEAR FORMS WITH THE WIENER FILTER

In this section, we summarize the solutions proposed in [10], i.e., the direct and iterative Wiener filters with bilinear forms. Given the filters $\hat{\mathbf{h}}$ and $\hat{\mathbf{g}}$, the estimated signal is given by $\hat{y}(t) = \hat{\mathbf{h}}^T \mathbf{X}(t) \hat{\mathbf{g}}$. As a result, the error signal between the desired and estimated signals is defined as

$$\begin{aligned} e(t) &= d(t) - \hat{y}(t) = d(t) - \hat{\mathbf{h}}^T \mathbf{X}(t) \hat{\mathbf{g}} \\ &= d(t) - \left(\hat{\mathbf{g}} \otimes \hat{\mathbf{h}} \right)^T \tilde{\mathbf{x}}(t) = d(t) - \hat{\mathbf{f}}^T \tilde{\mathbf{x}}(t). \end{aligned} \quad (8)$$

In order to find the optimal spatiotemporal filter, $\hat{\mathbf{f}} = \hat{\mathbf{g}} \otimes \hat{\mathbf{h}}$, we need to minimize a cost function that is usually constructed from the error signal in (8). The most practical choice for this criterion is the mean-squared error (MSE):

$$\begin{aligned} J(\hat{\mathbf{h}}, \hat{\mathbf{g}}) &= E[e^2(t)] \\ &= \sigma_d^2 - 2 \left(\hat{\mathbf{g}} \otimes \hat{\mathbf{h}} \right)^T \mathbf{p} + \left(\hat{\mathbf{g}} \otimes \hat{\mathbf{h}} \right)^T \mathbf{R} \left(\hat{\mathbf{g}} \otimes \hat{\mathbf{h}} \right) \\ &= \sigma_d^2 - 2\hat{\mathbf{f}}^T \mathbf{p} + \hat{\mathbf{f}}^T \mathbf{R} \hat{\mathbf{f}} = J(\hat{\mathbf{f}}), \end{aligned} \quad (9)$$

where $\mathbf{R} = E[\tilde{\mathbf{x}}(t)\tilde{\mathbf{x}}^T(t)]$ is the covariance matrix of $\tilde{\mathbf{x}}(t)$ [with $E(\cdot)$ denoting mathematical expectation] and $\mathbf{p} = E[\tilde{\mathbf{x}}(t)d(t)]$ is the cross-correlation vector between $\tilde{\mathbf{x}}(t)$ and $d(t)$. From (9), we observe that the first obvious possibility

to derive the Wiener spatiotemporal filter is from the minimization of $J(\hat{\mathbf{f}})$ with respect to $\hat{\mathbf{f}}$. Thus, the straightforward solution is

$$\hat{\mathbf{f}}_W = \mathbf{R}^{-1} \mathbf{p}. \quad (10)$$

The previous expression can be rewritten as

$$\begin{aligned} \hat{\mathbf{f}}_W &= \hat{\mathbf{g}}_W \otimes \hat{\mathbf{h}}_W = \left[\hat{\mathbf{h}}_W^T \quad \hat{g}_{2,W} \hat{\mathbf{h}}_W^T \quad \cdots \quad \hat{g}_{M,W} \hat{\mathbf{h}}_W^T \right]^T \\ &= \left[\hat{\mathbf{f}}_{1,W}^T \quad \hat{\mathbf{f}}_{2,W}^T \quad \cdots \quad \hat{\mathbf{f}}_{M,W}^T \right]^T, \end{aligned} \quad (11)$$

where $\hat{g}_{m,W}$, $m = 1, 2, \dots, M$ are the components of $\hat{\mathbf{g}}_W$ and, without loss of generality, we absorb the scaling ambiguity in the first element of $\hat{\mathbf{g}}_W$ by assuming it to be equal to 1, i.e., $\hat{g}_{1,W} = 1$. It is clear from (11) that the first L elements of $\hat{\mathbf{f}}_W$ correspond to $\hat{\mathbf{h}}_W$. Now, to find the components $\hat{g}_{i,W}$, $i = 2, 3, \dots, M$, we need to minimize $\|\hat{\mathbf{f}}_{i,W} - \hat{g}_{i,W} \hat{\mathbf{h}}_W\|^2$. Thus, the solution is

$$\hat{g}_{i,W} = \frac{\hat{\mathbf{f}}_{i,W}^T \hat{\mathbf{h}}_W}{\hat{\mathbf{h}}_W^T \hat{\mathbf{h}}_W}. \quad (12)$$

However, there is a fundamental limitation with this direct approach. Indeed, since the covariance matrix, \mathbf{R} , of the multi-input signal is very large (of size $ML \times ML$), a large amount of data is required in order to have a good estimate of it. Otherwise, the estimate of \mathbf{R} will be very ill conditioned or will not even be full rank. As a consequence, $\hat{\mathbf{f}}_W$ will be very inaccurate or even useless. This closed-form method tries to identify ML coefficients while we have only $M + L$ coefficients to identify. Furthermore, if there is a mismatch between the assumed model and the real one, this technique will likely fail.

The second possibility, which is iterative in nature, is based on much smaller covariance matrices (of sizes $L \times L$ and $M \times M$). In this way, we fully exploit the fact that there are only $M + L$ coefficients to identify in the bilinear model. As a result, even with a small amount of data, we can have very good estimates of the unknown impulse responses.

In the following, let us consider the relationships [13]:

$$\hat{\mathbf{g}} \otimes \hat{\mathbf{h}} = (\hat{\mathbf{g}} \otimes \mathbf{I}_L) \hat{\mathbf{h}} \quad (13)$$

$$= (\mathbf{I}_M \otimes \hat{\mathbf{h}}) \hat{\mathbf{g}}, \quad (14)$$

where \mathbf{I}_L and \mathbf{I}_M are the identity matrices of sizes $L \times L$ and $M \times M$, respectively. Using (13) and (14), we can express the MSE criterion as

$$J(\hat{\mathbf{h}}, \hat{\mathbf{g}}) = \sigma_d^2 - 2\hat{\mathbf{h}}^T \mathbf{p}_{\hat{\mathbf{g}}} + \hat{\mathbf{h}}^T \mathbf{R}_{\hat{\mathbf{g}}} \hat{\mathbf{h}} \quad (15)$$

$$= \sigma_d^2 - 2\hat{\mathbf{g}}^T \mathbf{p}_{\hat{\mathbf{h}}} + \hat{\mathbf{g}}^T \mathbf{R}_{\hat{\mathbf{h}}} \hat{\mathbf{g}}, \quad (16)$$

where

$$\mathbf{p}_{\hat{\mathbf{g}}} = (\hat{\mathbf{g}} \otimes \mathbf{I}_L)^T \mathbf{p}, \quad (17)$$

$$\mathbf{R}_{\hat{\mathbf{g}}} = (\hat{\mathbf{g}} \otimes \mathbf{I}_L)^T \mathbf{R} (\hat{\mathbf{g}} \otimes \mathbf{I}_L), \quad (18)$$

and

$$\mathbf{p}_{\hat{\mathbf{h}}} = (\mathbf{I}_M \otimes \hat{\mathbf{h}})^T \mathbf{p}, \quad (19)$$

$$\mathbf{R}_{\hat{\mathbf{h}}} = (\mathbf{I}_M \otimes \hat{\mathbf{h}})^T \mathbf{R} (\mathbf{I}_M \otimes \hat{\mathbf{h}}). \quad (20)$$

When $\hat{\mathbf{g}}$ is fixed, we write (15) as

$$J_{\hat{\mathbf{g}}}(\hat{\mathbf{h}}) = \sigma_d^2 - 2\hat{\mathbf{h}}^T \mathbf{p}_{\hat{\mathbf{g}}} + \hat{\mathbf{h}}^T \mathbf{R}_{\hat{\mathbf{g}}} \hat{\mathbf{h}}, \quad (21)$$

and when $\hat{\mathbf{h}}$ is fixed, we express (16) as

$$J_{\hat{\mathbf{h}}}(\hat{\mathbf{g}}) = \sigma_d^2 - 2\hat{\mathbf{g}}^T \mathbf{p}_{\hat{\mathbf{h}}} + \hat{\mathbf{g}}^T \mathbf{R}_{\hat{\mathbf{h}}} \hat{\mathbf{g}}. \quad (22)$$

Based on the previous considerations, the iterative Wiener [10] starts with the initialization:

$$\begin{aligned} \hat{\mathbf{g}}^{(0)} &= \frac{1}{M} [1 \ 1 \ \dots \ 1]^T, \\ \mathbf{p}_{\hat{\mathbf{g}}}^{(0)} &= (\hat{\mathbf{g}}^{(0)} \otimes \mathbf{I}_L)^T \mathbf{p}, \\ \mathbf{R}_{\hat{\mathbf{g}}}^{(0)} &= (\hat{\mathbf{g}}^{(0)} \otimes \mathbf{I}_L)^T \mathbf{R} (\hat{\mathbf{g}}^{(0)} \otimes \mathbf{I}_L). \end{aligned}$$

Then, at iteration 1:

$$\begin{aligned} \hat{\mathbf{h}}^{(1)} &= (\mathbf{R}_{\hat{\mathbf{g}}}^{(0)})^{-1} \mathbf{p}_{\hat{\mathbf{g}}}^{(0)}, \\ \mathbf{p}_{\hat{\mathbf{h}}}^{(1)} &= (\mathbf{I}_M \otimes \hat{\mathbf{h}}^{(1)})^T \mathbf{p}, \\ \mathbf{R}_{\hat{\mathbf{h}}}^{(1)} &= (\mathbf{I}_M \otimes \hat{\mathbf{h}}^{(1)})^T \mathbf{R} (\mathbf{I}_M \otimes \hat{\mathbf{h}}^{(1)}), \\ \hat{\mathbf{g}}^{(1)} &= (\mathbf{R}_{\hat{\mathbf{h}}}^{(1)})^{-1} \mathbf{p}_{\hat{\mathbf{h}}}^{(1)}. \end{aligned}$$

Continuing to iterate up to iteration n , we get the estimate of the temporal impulse response:

$$\hat{\mathbf{h}}^{(n)} = (\mathbf{R}_{\hat{\mathbf{g}}}^{(n-1)})^{-1} \mathbf{p}_{\hat{\mathbf{g}}}^{(n-1)}, \quad (23)$$

where

$$\mathbf{R}_{\hat{\mathbf{g}}}^{(n-1)} = (\hat{\mathbf{g}}^{(n-1)} \otimes \mathbf{I}_L)^T \mathbf{R} (\hat{\mathbf{g}}^{(n-1)} \otimes \mathbf{I}_L), \quad (24)$$

$$\mathbf{p}_{\hat{\mathbf{g}}}^{(n-1)} = (\hat{\mathbf{g}}^{(n-1)} \otimes \mathbf{I}_L)^T \mathbf{p}, \quad (25)$$

and the estimate of the spatial impulse response:

$$\hat{\mathbf{g}}^{(n)} = (\mathbf{R}_{\hat{\mathbf{h}}}^{(n)})^{-1} \mathbf{p}_{\hat{\mathbf{h}}}^{(n)}, \quad (26)$$

where

$$\mathbf{R}_{\hat{\mathbf{h}}}^{(n)} = (\mathbf{I}_M \otimes \hat{\mathbf{h}}^{(n)})^T \mathbf{R} (\mathbf{I}_M \otimes \hat{\mathbf{h}}^{(n)}), \quad (27)$$

$$\mathbf{p}_{\hat{\mathbf{h}}}^{(n)} = (\mathbf{I}_M \otimes \hat{\mathbf{h}}^{(n)})^T \mathbf{p}. \quad (28)$$

Finally, we deduce that the Wiener spatiotemporal filter at iteration n is

$$\hat{\mathbf{f}}_W^{(n)} = \hat{\mathbf{g}}^{(n)} \otimes \hat{\mathbf{h}}^{(n)}, \quad (29)$$

where $\hat{\mathbf{h}}^{(n)}$ and $\hat{\mathbf{g}}^{(n)}$ are defined in (23) and (26), respectively. As shown in [10], this iterative solution is much more attractive as it leads to good estimates of the impulse responses, even when a small amount of data is used for the estimation of the statistics (i.e., \mathbf{R} and \mathbf{p}).

IV. NLMS ALGORITHM FOR BILINEAR FORMS

The iterative approach presented in the previous section may not be very efficient or convenient to use in practice (especially in real-time applications), due to the well-known limitations of the Wiener filter, i.e., matrix inversion operations, estimation of the statistics, etc). Consequently, we may approach the problem in terms of adaptive filtering. This is explored in this section, in the framework of the NLMS algorithm.

Let us consider two adaptive filters defined by the impulse responses $\hat{\mathbf{h}}(t)$ and $\hat{\mathbf{g}}(t)$, and the corresponding a priori error signals:

$$e_{\hat{\mathbf{g}}}(t) = d(t) - \hat{\mathbf{h}}^T(t-1) \tilde{\mathbf{x}}_{\hat{\mathbf{g}}}(t) \quad (30)$$

and

$$e_{\hat{\mathbf{h}}}(t) = d(t) - \hat{\mathbf{g}}^T(t-1) \tilde{\mathbf{x}}_{\hat{\mathbf{h}}}(t), \quad (31)$$

where

$$\tilde{\mathbf{x}}_{\hat{\mathbf{g}}}(t) = [\hat{\mathbf{g}}(t-1) \otimes \mathbf{I}_L]^T \tilde{\mathbf{x}}(t), \quad (32)$$

$$\tilde{\mathbf{x}}_{\hat{\mathbf{h}}}(t) = [\mathbf{I}_M \otimes \hat{\mathbf{h}}(t-1)]^T \tilde{\mathbf{x}}(t). \quad (33)$$

It can be verified that $e_{\hat{\mathbf{g}}}(t) = e_{\hat{\mathbf{h}}}(t)$. However, for the clarity of the coming developments, we prefer to keep the notation from (30) and (31). The two filters can be recursively updated as

$$\begin{aligned} \hat{\mathbf{h}}(t) &= \hat{\mathbf{h}}(t-1) - \frac{\mu_{\hat{\mathbf{h}}}(t)}{2} \times \frac{\partial e_{\hat{\mathbf{g}}}^2(t)}{\partial \hat{\mathbf{h}}(t-1)} \\ &= \hat{\mathbf{h}}(t-1) + \mu_{\hat{\mathbf{h}}}(t) \tilde{\mathbf{x}}_{\hat{\mathbf{g}}}(t) e_{\hat{\mathbf{g}}}(t) \end{aligned} \quad (34)$$

and

$$\begin{aligned} \hat{\mathbf{g}}(t) &= \hat{\mathbf{g}}(t-1) - \frac{\mu_{\hat{\mathbf{g}}}(t)}{2} \times \frac{\partial e_{\hat{\mathbf{h}}}^2(t)}{\partial \hat{\mathbf{g}}(t-1)} \\ &= \hat{\mathbf{g}}(t-1) + \mu_{\hat{\mathbf{g}}}(t) \tilde{\mathbf{x}}_{\hat{\mathbf{h}}}(t) e_{\hat{\mathbf{h}}}(t), \end{aligned} \quad (35)$$

where $\mu_{\hat{\mathbf{h}}}(t) > 0$ and $\mu_{\hat{\mathbf{g}}}(t) > 0$ are the step-size parameters.

A reasonable way to derive the step-size parameters, taking into account the stability conditions, is to cancel the a posteriori error signals [15], which can be defined as

$$\varepsilon_{\hat{\mathbf{g}}}(t) = d(t) - \hat{\mathbf{h}}^T(t) \tilde{\mathbf{x}}_{\hat{\mathbf{g}}}(t) \quad (36)$$

and

$$\varepsilon_{\hat{\mathbf{h}}}(t) = d(t) - \hat{\mathbf{g}}^T(t) \tilde{\mathbf{x}}_{\hat{\mathbf{h}}}(t). \quad (37)$$

Replacing (34) in (36), and (35) in (37), with the requirements $\varepsilon_{\hat{\mathbf{g}}}(t) = 0$ and $\varepsilon_{\hat{\mathbf{h}}}(t) = 0$, respectively, it results that

$$\varepsilon_{\hat{\mathbf{g}}}(t) = e_{\hat{\mathbf{g}}}(t) [1 - \mu_{\hat{\mathbf{h}}}(t) \tilde{\mathbf{x}}_{\hat{\mathbf{g}}}^T(t) \tilde{\mathbf{x}}_{\hat{\mathbf{g}}}(t)] = 0 \quad (38)$$

and

$$\varepsilon_{\hat{\mathbf{h}}}(t) = e_{\hat{\mathbf{h}}}(t) [1 - \mu_{\hat{\mathbf{g}}}(t) \tilde{\mathbf{x}}_{\hat{\mathbf{h}}}^T(t) \tilde{\mathbf{x}}_{\hat{\mathbf{h}}}(t)] = 0. \quad (39)$$

Next, assuming that $e_{\hat{\mathbf{g}}}(t) \neq 0$ and $e_{\hat{\mathbf{h}}}(t) \neq 0$, we find

$$\mu_{\hat{\mathbf{h}}}(t) = \frac{1}{\tilde{\mathbf{x}}_{\hat{\mathbf{g}}}^T(t) \tilde{\mathbf{x}}_{\hat{\mathbf{g}}}(t)} \quad (40)$$

and

$$\mu_{\hat{\mathbf{g}}}(t) = \frac{1}{\tilde{\mathbf{x}}_{\hat{\mathbf{h}}}^T(t)\tilde{\mathbf{x}}_{\hat{\mathbf{h}}}(t)}, \quad (41)$$

which represent the step-sizes of the NLMS algorithm for bilinear forms.

In practice, two positive constants $\alpha_{\hat{\mathbf{h}}}$ and $\alpha_{\hat{\mathbf{g}}}$ (usually smaller than 1), multiply these step-sizes to achieve a proper compromise between the convergence rate and the misadjustment [16]. Also, the new NLMS algorithm needs to be regularized by adding two positive constants, $\delta_{\hat{\mathbf{h}}}$ and $\delta_{\hat{\mathbf{g}}}$, to the denominators of the step-size parameters. In general, these regularization parameters are chosen proportional to the input signal variance [17]. Therefore, the NLMS algorithm is defined by the updates:

$$\hat{\mathbf{h}}(t) = \hat{\mathbf{h}}(t-1) + \frac{\alpha_{\hat{\mathbf{h}}}\tilde{\mathbf{x}}_{\hat{\mathbf{g}}}(t)e_{\hat{\mathbf{g}}}(t)}{\tilde{\mathbf{x}}_{\hat{\mathbf{g}}}(t)\tilde{\mathbf{x}}_{\hat{\mathbf{g}}}(t) + \delta_{\hat{\mathbf{h}}}}, \quad (42)$$

$$\hat{\mathbf{g}}(t) = \hat{\mathbf{g}}(t-1) + \frac{\alpha_{\hat{\mathbf{g}}}\tilde{\mathbf{x}}_{\hat{\mathbf{h}}}(t)e_{\hat{\mathbf{h}}}(t)}{\tilde{\mathbf{x}}_{\hat{\mathbf{h}}}(t)\tilde{\mathbf{x}}_{\hat{\mathbf{h}}}(t) + \delta_{\hat{\mathbf{g}}}}. \quad (43)$$

For the initialization, we may choose

$$\begin{aligned} \hat{\mathbf{h}}(0) &= [1 \ 0 \ \cdots \ 0]^T, \\ \hat{\mathbf{g}}(0) &= \frac{1}{M} [1 \ 1 \ \cdots \ 1]^T. \end{aligned}$$

Finally, we can obtain the spatiotemporal filter as

$$\hat{\mathbf{f}}(t) = \hat{\mathbf{g}}(t) \otimes \hat{\mathbf{h}}(t). \quad (44)$$

where $\hat{\mathbf{h}}(t)$ and $\hat{\mathbf{g}}(t)$ result from (42) and (43), respectively.

Alternatively, we may use the regular NLMS algorithm to identify the spatiotemporal impulse response, \mathbf{f} , i.e.,

$$e_{\hat{\mathbf{f}}}(t) = d(t) - \hat{\mathbf{f}}^T(t-1)\tilde{\mathbf{x}}(t), \quad (45)$$

$$\hat{\mathbf{f}}(t) = \hat{\mathbf{f}}(t-1) + \frac{\alpha_{\hat{\mathbf{f}}}\tilde{\mathbf{x}}e_{\hat{\mathbf{f}}}(t)}{\tilde{\mathbf{x}}^T(t)\tilde{\mathbf{x}}(t) + \delta_{\hat{\mathbf{f}}}}, \quad (46)$$

where $\alpha_{\hat{\mathbf{f}}}$ and $\delta_{\hat{\mathbf{f}}}$ denote the normalized step-size and the regularization parameter, respectively. Then, the components $\hat{\mathbf{h}}(t)$ and $\hat{\mathbf{g}}(t)$ can be obtained similarly to the case of the direct Wiener filter, i.e., following the procedure related to (11) and (12).

Nevertheless, we should note that the solution based on the regular NLMS algorithm [from (45) and (46)] involves an adaptive filter of length ML , while the new NLMS algorithm [defined by (30)–(31) and (42)–(43)] uses two shorter filters of lengths L and M , respectively. Consequently, a faster converge rate is expected for the proposed NLMS algorithm as compared to the conventional approach.

V. SIMULATION RESULTS

In this section, the performance of the regular and proposed NLMS algorithms are evaluated from a system identification perspective. In our simulations, the impulse response \mathbf{h} is randomly generated (with Gaussian distribution) and its length is set to $L = 64$. The coefficients of the impulse response \mathbf{g} (of length M) are also randomly generated, using different

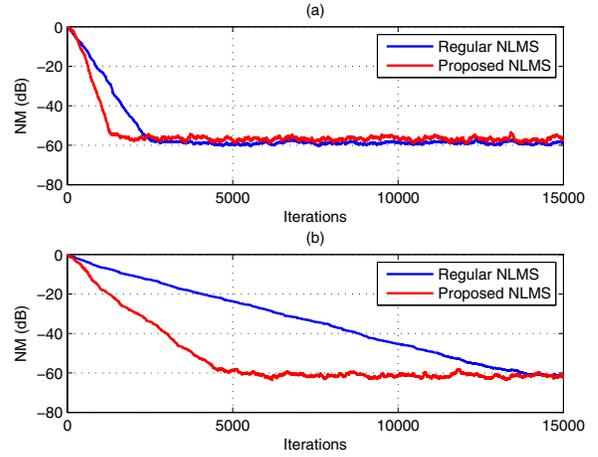


Fig. 1. Normalized misalignment (NM) of the regular and proposed NLMS algorithms for the estimation of the impulse response, \mathbf{f} , of length $ML = 128$. The input signals are (a) white Gaussian noises and (b) AR(1) processes.

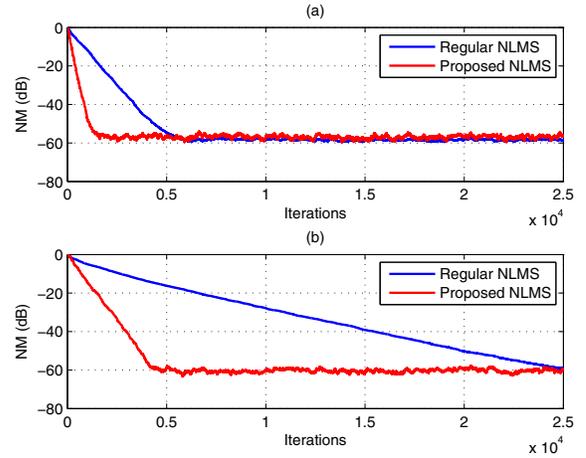


Fig. 2. Normalized misalignment (NM) of the regular and proposed NLMS algorithms for the estimation of the impulse response, \mathbf{f} , of length $ML = 256$. The input signals are (a) white Gaussian noises and (b) AR(1) processes.

values of M . The input signals $x_m(t)$, $m = 1, 2, \dots, M$ are either white Gaussian noises or AR(1) processes [each one of them is generated by filtering a white Gaussian noise through a first-order system $1/(1 - 0.8z^{-1})$]. The additive noise $w(t)$ is white and Gaussian, with the variance $\sigma_w^2 = 0.001$.

First, the performances of the regular and proposed NLMS algorithms are evaluated in terms of identifying the spatiotemporal impulse response, \mathbf{f} . The performance measure is the NM (in dB) evaluated based on (7). In this framework, the length of the impulse response \mathbf{g} takes different values, i.e., $M = 2, 4$, and 8 ; consequently, the length of \mathbf{f} is $ML = 128, 256$, and 512 , respectively. The normalized step-sizes are set to $\alpha_{\hat{\mathbf{f}}} = \alpha_{\hat{\mathbf{h}}} = \alpha_{\hat{\mathbf{g}}} = 0.5$ and the regularization parameters are chosen as $\delta_{\hat{\mathbf{f}}} = \delta_{\hat{\mathbf{h}}} = \delta_{\hat{\mathbf{g}}} = 20\sigma_{\tilde{\mathbf{x}}}^2$, where $\sigma_{\tilde{\mathbf{x}}}^2$ is the variance of the input signal $\tilde{\mathbf{x}}(t)$.

The results presented in Figs. 1–3 support the discussion

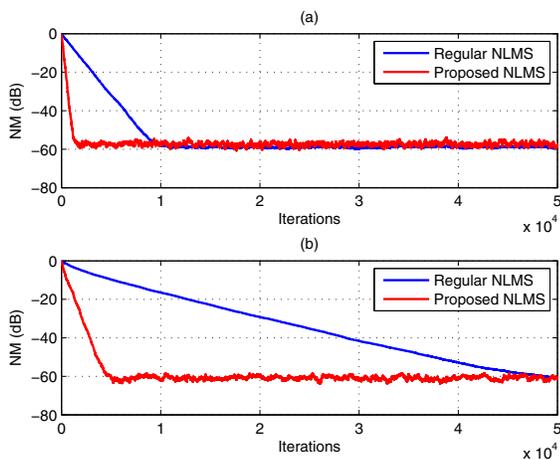


Fig. 3. Normalized misalignment (NM) of the regular and proposed NLMS algorithms for the estimation of the impulse response, \mathbf{f} , of length $ML = 512$. The input signals are (a) white Gaussian noises and (b) AR(1) processes.

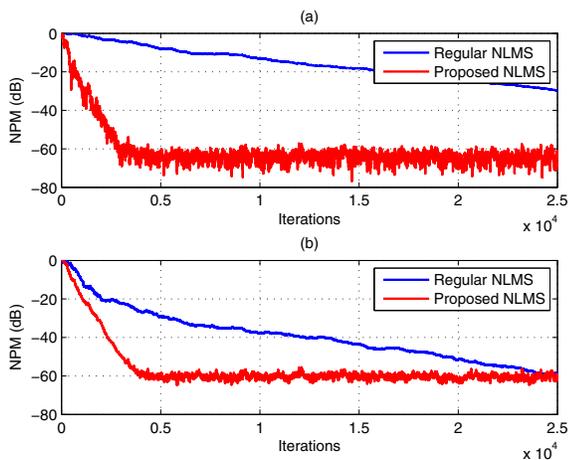


Fig. 4. Normalized projection misalignment (NPM) of the regular and proposed NLMS algorithms for the estimation of (a) the impulse response \mathbf{h} (of length $L = 64$) and (b) the impulse response \mathbf{g} (of length $M = 8$). The input signals are AR(1) processes.

from the end of Section IV, related to the advantage of the proposed NLMS algorithm. As we can notice, the new algorithm outperforms the regular NLMS [based on (45) and (46)] in terms of the convergence rate. The gain is more apparent when the length (ML) of the spatiotemporal filter increases.

Next, we evaluate the estimation of the impulse responses \mathbf{h} and \mathbf{g} with the regular NLMS algorithm [similar to (11) and (12)] and the proposed NLMS [using (42) and (43)]. The results are shown in Fig. 4, where $M = 8$ (i.e., $ML = 512$). The performance is evaluated in terms of the NPM (in dB), which is computed based on (5) and (6). It can be noticed that the proposed NLMS algorithm outperforms its regular counterpart in terms of convergence rate.

VI. CONCLUSIONS

In this paper, we have addressed the problem of identifying bilinear forms, in the context of MISO systems. Our derivation focusses on a different model (as compared to most previous works related to bilinear systems), by defining the bilinear term with respect to the impulse responses of the spatiotemporal model. In this framework, two versions of the Wiener filter have been presented, followed by a more practical solution based on the NLMS algorithm. Simulation results indicate that the proposed NLMS algorithm could represent an appealing solution for bilinear system identification problems.

ACKNOWLEDGMENT

This work was supported by UEFISCDI Romania under Grant PN-II-RU-TE-2014-4-1880 and by UPB Romania under Grants UPB-GEX 99/2016 code 533 & 100/2016 code 220.

REFERENCES

- [1] U. Forssén, "Adaptive bilinear digital filters," *IEEE Trans. Circuits Systems-II: Analog and Digital Signal Process.*, vol. 40, pp. 729–735, Nov. 1993.
- [2] G.-K. Ma, J. Lee, and V. J. Mathews, "A RLS bilinear filter for channel equalization," in *Proc. IEEE ICASSP*, 1994, pp. III-257–III-260.
- [3] R. Hu and H. M. Hassan, "Echo cancellation in high speed data transmission systems using adaptive layered bilinear filters," *IEEE Trans. Communications*, vol. 42, pp. 655–663, Feb./Mar./Apr. 1994.
- [4] D. Gesbert and P. Duhamel, "Robust blind joint data/channel estimation based on bilinear optimization," in *Proc. IEEE WSSAP*, 1996, pp. 168–171.
- [5] Z. Zhu and H. Leung, "Adaptive identification of nonlinear systems with application to chaotic communications," *IEEE Trans. Circuits Systems-I: Fundamental Theory and Applications*, vol. 47, pp. 1072–1080, July 2000.
- [6] S. M. Kuo and H.-T. Wu, "Nonlinear adaptive bilinear filters for active noise control systems," *IEEE Trans. Circuits Systems-I: Regular Papers*, vol. 52, pp. 617–624, Mar. 2005.
- [7] P. Lopes dos Santos, J. A. Ramos, and J. L. Martins de Carvalho, "Identification of bilinear systems with white noise inputs: an iterative deterministic-stochastic subspace approach," *IEEE Trans. Control Systems Technology*, vol. 17, pp. 1145–1153, Sept. 2009.
- [8] H. Zhao, X. Zeng, and Z. He, "Low-complexity nonlinear adaptive filter based on a pipelined bilinear recurrent neural network," *IEEE Trans. Neural Networks*, vol. 22, pp. 1494–1507, Sept. 2011.
- [9] L. Tan and J. Jiang, "Nonlinear active noise control using diagonal-channel LMS and RLS bilinear filters," in *Proc. IEEE MWSCAS*, 2014, pp. 789–792.
- [10] J. Benesty, C. Paleologu, and S. Ciochină, "On the identification of bilinear forms with the Wiener filter," *IEEE Signal Process. Lett.*, vol. 24, pp. 653–657, May 2017.
- [11] A. Stenger, W. Kellermann, R. Rabenstein, "Adaptation of acoustic echo cancellers incorporating a memoryless nonlinearity," in *Proc. IEEE IWAENC*, 1999 (4 pages).
- [12] A. Stenger and W. Kellermann, "Adaptation of a memoryless preprocessor for nonlinear acoustic echo cancelling," *Signal Processing*, vol. 80, pp. 1747–1760, Sept. 2000.
- [13] D. A. Harville, *Matrix Algebra From a Statistician's Perspective*. New York: Springer-Verlag, 1997.
- [14] D. R. Morgan, J. Benesty, and M. M. Sondhi, "On the evaluation of estimated impulse responses," *IEEE Signal Process. Lett.*, vol. 5, pp. 174–176, July 1998.
- [15] D. R. Morgan and S. G. Kratzer, "On a class of computationally efficient, rapidly converging, generalized NLMS algorithms," *IEEE Signal Process. Lett.*, vol. 3, pp. 245–247, Aug. 1996.
- [16] S. Haykin, *Adaptive Filter Theory*. Fourth Edition, Upper Saddle River, NJ: Prentice-Hall, 2002.
- [17] J. Benesty, C. Paleologu, and S. Ciochină, "On regularization in adaptive filtering," *IEEE Trans. Audio, Speech, Language Processing*, vol. 19, pp. 1734–1742, Aug. 2011.