Article

A Connection Between the Kalman Filter and an Optimized LMS Algorithm for Bilinear Forms

Laura-Maria Dogariu ¹, Silviu Ciochină ¹, Constantin Paleologu ^{1,*} and Jacob Benesty ²

¹ Department of Telecommunications, University Politehnica of Bucharest, 1-3, Iuliu Maniu Blvd., 061071 Bucharest, Romania; ldogariu@comm.pub.ro (L.-M.D.); silviu@comm.pub.ro (S.C.)

² Energy Materials Telecommunications Research Centre, National Institute of Scientific Research (INRS-EMT), University of Quebec, Montreal, QC H5A 1K6, Canada; benesty@emt.inrs.ca

* Correspondence: pale@comm.pub.ro; Tel.: +40-21-402-4634

Received: 1 December 2018; Accepted: 14 December 2018; Published: 17 December 2018

Abstract: The system identification problem becomes more challenging when the parameter space increases. Recently, several works have focused on the identification of bilinear forms, which are related to the impulse responses of a spatiotemporal model, in the context of a multiple-input/single-output system. In this framework, the problem was addressed in terms of the Wiener filter and different basic adaptive algorithms. This paper studies two types of algorithms tailored for the identification of such bilinear forms, i.e., the Kalman filter (along with its simplified version) and an optimized least-mean-square (LMS) algorithm. Also, a comparison between them is performed, which shows interesting similarities. In addition to the mathematical derivation of the algorithms, we also provide extensive experimental results, which support the theoretical findings and indicate the good performance of the proposed solutions.

Keywords: adaptive filter; Kalman filter; optimized LMS algorithm; bilinear forms; system identification

1. Introduction

Bilinear systems were previously treated in various contexts [1], such as the nonlinear systems approximation. The applications derived from there are numerous, among which we can mention system identification [2–5], design of digital filters [6], echo cancellation [7], chaotic communications [8], active noise control [9], neural networks [10], etc. In all these papers, the bilinear term is understood in terms of an input-output relation (with respect to the data).

More recently, a new approach was studied in [11], where the bilinear term was defined in the context of a multiple-input/single-output (MISO) system, with respect to the impulse responses of a spatiotemporal model. In [11], the Wiener filter solution for the identification of such bilinear forms was proposed, and then, in [12–14], some adaptive solutions based on different basic algorithms were provided. Similar frameworks can be found in [15–18], in conjunction with specific applications, such as channel equalization and nonlinear acoustic echo cancellation; however, these works were not associated with the identification of bilinear forms or analyzed in this context.

The Kalman filter was first introduced in [19] and it has a wide range of applications in technology, as well as in signal processing [20]. Some of its applications in the automotive field are mentioned in [21,22], and the references therein, while [23] also provides a computational complexity analysis.

In this paper, we focus on a comparative study of two different types of algorithms tailored for the identification of bilinear forms. The first one is based on the Kalman filter [19], whose bilinear form was introduced in our previous work [13], together with a simplified (i.e., low complexity) version, which could be more suitable in real-world applications. The second algorithm is a version of the least-mean-square (LMS) adaptive filter, but where the step-size parameter has been optimized in

order to meet a proper compromise between the convergence rate and misadjustment. In addition, as it will be shown, there is a strong similarity between the simplified Kalman filter and the optimized LMS algorithm. Simulations show the good performance of the proposed solutions, in comparison with other related works based on the conventional algorithms. Moreover, the gain could be twofold, in terms of both convergence performance and computational complexity.

The paper structure is as follows. Section 2 introduces the system model in the context of bilinear forms. In this framework, we present the Kalman filter together with its simplified version in Section 3. The optimized LMS algorithm is derived in Section 4. Next, a comparison between the simplified Kalman filter and the optimized LMS algorithm for bilinear forms is presented in Section 5. Section 6 provides some practical considerations. Simulations are performed in the framework of system identification and the results are given in Section 7. Finally, some conclusions are drawn in Section 8. Also, Appendix A provides detailed computations of some terms required within the optimized LMS algorithm, while Appendix B summarizes and briefly explains the main parameters used in the paper, in order to facilitate the reading. For the sake of clarity, the main notation used in this work is provided in Table 1.

Table 1. Notation used throughout the paper.

Symbol	Significance
a	scalar a
$\mathbf{a}$	vector $\mathbf{a}$
$\mathbf{A}$	matrix $\mathbf{A}$
T	transpose operator
$\otimes$	Kronecker product operator
$\ \mathbf{b}\ $	Euclidean norm of the vector $\mathbf{b}$
$\mathbb{E}(\cdot)$	mathematical expectation
$\mathbf{I}_L$	identity matrix of size $L \times L$
$\text{tr}(\mathbf{B})$	trace of the square matrix $\mathbf{B}$
$\text{vec}(\mathbf{C}) = \mathbf{c}$	vectorization operation, i.e., conversion of a matrix ($\mathbf{C}$ of size $L \times M$) into a vector ($\mathbf{c}$ of length ML)

2. System Model

The signal model that shall be used throughout the paper is given by [11]

$$\begin{aligned} d(n) &= \mathbf{h}^T(n)\mathbf{X}(n)\mathbf{g}(n) + v(n) \\ &= y(n) + v(n), \end{aligned} \quad (1)$$

where $d(n)$ is the zero-mean desired (or reference) signal at the discrete-time index n , $\mathbf{h}(n)$ and $\mathbf{g}(n)$ are the two impulse responses of the system of lengths L and M , respectively, the superscript T is the transpose operator, $\mathbf{X}(n) = [\mathbf{x}_1(n) \ \mathbf{x}_2(n) \ \dots \ \mathbf{x}_M(n)]$ is the zero-mean multiple-input signal matrix, $\mathbf{x}_m(n) = [x_m(n) \ x_m(n-1) \ \dots \ x_m(n-L+1)]^T$ is a vector containing the L most recent time samples of the m th ($m = 1, 2, \dots, M$) input signal, $y(n) = \mathbf{h}^T(n)\mathbf{X}(n)\mathbf{g}(n)$ is the output signal and it represents a bilinear form, and $v(n)$ is a zero-mean additive noise (with the variance σ_v^2). It is assumed that all signals are real-valued, and $\mathbf{X}(n)$ and $v(n)$ are independent. As we can notice, $y(n)$ is a bilinear function of $\mathbf{h}(n)$ and $\mathbf{g}(n)$, because for every fixed $\mathbf{h}(n)$, $y(n)$ is a linear function of $\mathbf{g}(n)$, and for every fixed $\mathbf{g}(n)$, it is a linear function of $\mathbf{h}(n)$.

We can rewrite the matrix $\mathbf{X}(n)$, of size $L \times M$, as a vector of length ML , by using the vectorization operation:

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

Therefore, the output signal can be expressed as

$$\begin{aligned} y(n) &= \mathbf{h}^T(n) \mathbf{X}(n) \mathbf{g}(n) \\ &= [\mathbf{g}(n) \otimes \mathbf{h}(n)]^T \tilde{\mathbf{x}}(n) = \mathbf{f}^T(n) \tilde{\mathbf{x}}(n), \end{aligned} \quad (3)$$

where $\otimes$ denotes the Kronecker product between the individual impulse responses, while the vector $\mathbf{f}(n) = \mathbf{g}(n) \otimes \mathbf{h}(n)$, of length ML , represents the spatiotemporal (i.e., global) impulse response of the system. Consequently, the signal model in (1) becomes

$$d(n) = \mathbf{f}^T(n) \tilde{\mathbf{x}}(n) + v(n). \quad (4)$$

The major difference compared to a general MISO system is yielded by the fact that in this bilinear context $\mathbf{f}(n)$ is formed with only $M + L$ different elements, despite being of length ML .

The goal is the identification of the two impulse responses $\mathbf{h}(n)$ and $\mathbf{g}(n)$, and, in this way, of the spatiotemporal impulse response $\mathbf{f}(n)$. For this aim, we can use two adaptive filters, $\hat{\mathbf{h}}(n)$ and $\hat{\mathbf{g}}(n)$; hence, the global impulse response can be evaluated as

$$\hat{\mathbf{f}}(n) = \hat{\mathbf{g}}(n) \otimes \hat{\mathbf{h}}(n). \quad (5)$$

Let $\eta \neq 0$ be a real-valued constant. We can see from (1) that

$$\left[\frac{1}{\eta} \mathbf{h}^T(n) \right] \mathbf{X}(n) [\eta \mathbf{g}(n)] = \mathbf{h}^T(n) \mathbf{X}(n) \mathbf{g}(n) = y(n), \quad (6)$$

meaning that the pairs $[\mathbf{h}(n)/\eta, \eta \mathbf{g}(n)]$ and $[\mathbf{h}(n), \mathbf{g}(n)]$ are equivalent in the bilinear form. This implies that we can only identify $\hat{\mathbf{h}}(n)$ and $\hat{\mathbf{g}}(n)$ up to a scaling factor. A similar discussion can be found in [15,18] in the framework of blind identification/equalization and nonlinear acoustic echo cancellation, respectively. Nevertheless, because

$$\mathbf{f}(n) = \mathbf{g}(n) \otimes \mathbf{h}(n) = [\eta \mathbf{g}(n)] \otimes \left[\frac{1}{\eta} \mathbf{h}(n) \right], \quad (7)$$

the global impulse response can be identified with no scaling ambiguity. Consequently, for the performance evaluation of the identification of the temporal and spatial filters, we can use the normalized projection misalignment (NPM), defined in [24]:

$$\text{NPM} [\mathbf{h}(n), \hat{\mathbf{h}}(n)] = 1 - \left[\frac{\mathbf{h}^T(n) \hat{\mathbf{h}}(n)}{\|\mathbf{h}(n)\| \|\hat{\mathbf{h}}(n)\|} \right]^2, \quad (8)$$

$$\text{NPM} [\mathbf{g}(n), \hat{\mathbf{g}}(n)] = 1 - \left[\frac{\mathbf{g}^T(n) \hat{\mathbf{g}}(n)}{\|\mathbf{g}(n)\| \|\hat{\mathbf{g}}(n)\|} \right]^2, \quad (9)$$

where $\|\cdot\|$ denotes the Euclidean norm. On the other hand, for the identification of the global filter, $\mathbf{f}(n)$, we should use the normalized misalignment:

$$\text{NM} [\mathbf{f}(n), \hat{\mathbf{f}}(n)] = \frac{\|\mathbf{f}(n) - \hat{\mathbf{f}}(n)\|^2}{\|\mathbf{f}(n)\|^2}. \quad (10)$$

In [11], this bilinear system identification problem has been studied in terms of the Wiener filter. Therefore, the assumption was that the two impulse responses that need to be identified are time-invariant (which represents a basic assumption in the context of the Wiener filter). In practice, however, these systems could vary in time. For this reason, in this paper, we approach the system identification problem considering that the systems that need to be identified vary in time. Thus,

we assume that $\mathbf{h}(n)$ and $\mathbf{g}(n)$ are zero-mean random vectors, following a simplified first-order Markov model, i.e.,

$$\mathbf{h}(n) = \mathbf{h}(n-1) + \mathbf{w}_h(n), \quad (11)$$

$$\mathbf{g}(n) = \mathbf{g}(n-1) + \mathbf{w}_g(n), \quad (12)$$

where $\mathbf{w}_h(n)$ and $\mathbf{w}_g(n)$ are zero-mean white Gaussian noise vectors, with correlation matrices $\mathbf{R}_{w_h}(n) = \sigma_{w_h}^2 \mathbf{I}_L$ and $\mathbf{R}_{w_g}(n) = \sigma_{w_g}^2 \mathbf{I}_M$, respectively (with $\mathbf{I}_L$ and $\mathbf{I}_M$ being the identity matrices of sizes $L \times L$ and $M \times M$, respectively). It is considered that $\mathbf{w}_h(n)$ is uncorrelated with $\mathbf{h}(n-1)$ and $v(n)$, while $\mathbf{w}_g(n)$ is uncorrelated with $\mathbf{g}(n-1)$ and $v(n)$. The variances $\sigma_{w_h}^2$ and $\sigma_{w_g}^2$ capture the uncertainties in $\mathbf{h}(n)$ and $\mathbf{g}(n)$, respectively.

3. Kalman Filter for Bilinear Forms

In this section, we address the previously described system identification problem in terms of the Kalman filter, summarizing the findings from [13]. In this context, the signal model from (1) may be interpreted as the observation equation, while the system impulse responses can be considered as state equations. Given the two adaptive filters $\hat{\mathbf{h}}(n)$ and $\hat{\mathbf{g}}(n)$, the estimated signal is given by

$$\hat{y}(n) = \hat{\mathbf{h}}^T(n-1)\mathbf{X}(n)\hat{\mathbf{g}}(n-1). \quad (13)$$

As a result, the a priori error signal between the desired and estimated signals can be defined as

$$\begin{aligned} e(n) &= d(n) - \hat{y}(n) \\ &= d(n) - \hat{\mathbf{h}}^T(n-1)\mathbf{X}(n)\hat{\mathbf{g}}(n-1) \\ &= d(n) - \left[\hat{\mathbf{g}}(n-1) \otimes \hat{\mathbf{h}}(n-1) \right]^T \tilde{\mathbf{x}}(n) \\ &= d(n) - \hat{\mathbf{f}}^T(n-1)\tilde{\mathbf{x}}(n) \\ &= d(n) - \hat{\mathbf{h}}^T(n-1)\mathbf{x}_{\hat{\mathbf{g}}}(n) \\ &= d(n) - \hat{\mathbf{g}}^T(n-1)\mathbf{x}_{\hat{\mathbf{h}}}(n), \end{aligned} \quad (14)$$

where

$$\mathbf{x}_{\hat{\mathbf{g}}}(n) = [\hat{\mathbf{g}}(n-1) \otimes \mathbf{I}_L]^T \tilde{\mathbf{x}}(n), \quad (15)$$

$$\mathbf{x}_{\hat{\mathbf{h}}}(n) = [\mathbf{I}_M \otimes \hat{\mathbf{h}}(n-1)]^T \tilde{\mathbf{x}}(n). \quad (16)$$

In the context of the linear sequential Bayesian approach, the optimal estimates of the state vectors have the forms [25]:

$$\hat{\mathbf{h}}(n) = \hat{\mathbf{h}}(n-1) + \mathbf{k}_h(n)e(n), \quad (17)$$

$$\hat{\mathbf{g}}(n) = \hat{\mathbf{g}}(n-1) + \mathbf{k}_g(n)e(n), \quad (18)$$

where $\mathbf{k}_h(n)$ and $\mathbf{k}_g(n)$ are the Kalman gain vectors. Next, we can define the a posteriori misalignments (which represent the state estimation errors) related to the temporal and spatial impulse responses:

$$\mathbf{c}_h(n) = \frac{1}{\eta} \mathbf{h}(n) - \hat{\mathbf{h}}(n), \quad (19)$$

$$\mathbf{c}_g(n) = \eta \mathbf{g}(n) - \hat{\mathbf{g}}(n), \quad (20)$$

for which their correlation matrices are $\mathbf{R}_{c_h}(n) = \mathbb{E}[\mathbf{c}_h(n)\mathbf{c}_h^T(n)]$ and $\mathbf{R}_{c_g}(n) = \mathbb{E}[\mathbf{c}_g(n)\mathbf{c}_g^T(n)]$, respectively, where $\mathbb{E}[\cdot]$ denotes mathematical expectation. As mentioned in Section 2, we can only identify the impulse responses up to this arbitrary scaling factor η ; however, the pair $\mathbf{h}(n)/\eta$ and $\eta\mathbf{g}(n)$

is equivalent to the pair $\mathbf{h}$ and $\mathbf{g}$ in the bilinear form. Let us also define the a priori misalignments related to the two impulse responses:

$$\begin{aligned}\mathbf{c}_{\mathbf{h}_a}(n) &= \frac{1}{\eta} \mathbf{h}(n) - \hat{\mathbf{h}}(n-1) \\ &= \mathbf{c}_{\mathbf{h}}(n-1) + \frac{1}{\eta} \mathbf{w}_{\mathbf{h}}(n),\end{aligned}\quad (21)$$

$$\begin{aligned}\mathbf{c}_{\mathbf{g}_a}(n) &= \eta \mathbf{g}(n) - \hat{\mathbf{g}}(n-1) \\ &= \mathbf{c}_{\mathbf{g}}(n-1) + \eta \mathbf{w}_{\mathbf{g}}(n),\end{aligned}\quad (22)$$

whose correlation matrices are $\mathbf{R}_{\mathbf{c}_{\mathbf{h}_a}}(n) = \mathbb{E} [\mathbf{c}_{\mathbf{h}_a}(n) \mathbf{c}_{\mathbf{h}_a}^T(n)]$ and $\mathbf{R}_{\mathbf{c}_{\mathbf{g}_a}}(n) = \mathbb{E} [\mathbf{c}_{\mathbf{g}_a}(n) \mathbf{c}_{\mathbf{g}_a}^T(n)]$, respectively.

For the sake of simplicity of the upcoming developments, let us multiply $\mathbf{w}_{\mathbf{h}}(n)$ and $\mathbf{w}_{\mathbf{g}}(n)$ by $1/\eta$ and η , respectively, and introduce the notation:

$$\bar{\mathbf{w}}_{\mathbf{h}}(n) = \frac{1}{\eta} \mathbf{w}_{\mathbf{h}}(n), \quad (23)$$

$$\bar{\mathbf{w}}_{\mathbf{g}}(n) = \eta \mathbf{w}_{\mathbf{g}}(n). \quad (24)$$

These new terms are also zero-mean white Gaussian noise vectors, having the correlation matrices $\mathbf{R}_{\bar{\mathbf{w}}_{\mathbf{h}}}(n) = \sigma_{\bar{\mathbf{w}}_{\mathbf{h}}}^2 \mathbf{I}_L$ and $\mathbf{R}_{\bar{\mathbf{w}}_{\mathbf{g}}}(n) = \sigma_{\bar{\mathbf{w}}_{\mathbf{g}}}^2 \mathbf{I}_M$, respectively. Clearly, we have $\sigma_{\bar{\mathbf{w}}_{\mathbf{h}}}^2 = \sigma_{\mathbf{w}_{\mathbf{h}}}^2 / \eta^2$ and $\sigma_{\bar{\mathbf{w}}_{\mathbf{g}}}^2 = \eta^2 \sigma_{\mathbf{w}_{\mathbf{g}}}^2$. Consequently, using this notation in (21) and (22), we obtain

$$\mathbf{R}_{\mathbf{c}_{\mathbf{h}_a}}(n) = \mathbf{R}_{\mathbf{c}_{\mathbf{h}}}(n-1) + \sigma_{\bar{\mathbf{w}}_{\mathbf{h}}}^2 \mathbf{I}_L, \quad (25)$$

$$\mathbf{R}_{\mathbf{c}_{\mathbf{g}_a}}(n) = \mathbf{R}_{\mathbf{c}_{\mathbf{g}}}(n-1) + \sigma_{\bar{\mathbf{w}}_{\mathbf{g}}}^2 \mathbf{I}_M. \quad (26)$$

In this context, the Kalman gain vectors are computed from the minimization of the criteria:

$$J_{\mathbf{h}}(n) = \frac{1}{L} \text{tr} [\mathbf{R}_{\mathbf{c}_{\mathbf{h}}}(n)], \quad (27)$$

$$J_{\mathbf{g}}(n) = \frac{1}{M} \text{tr} [\mathbf{R}_{\mathbf{c}_{\mathbf{g}}}(n)], \quad (28)$$

with respect to $\mathbf{k}_{\mathbf{h}}(n)$ and $\mathbf{k}_{\mathbf{g}}(n)$, respectively, where $\text{tr}[\cdot]$ denotes the trace of a square matrix. From these minimizations, we find that

$$\mathbf{k}_{\mathbf{h}}(n) = \frac{\mathbf{R}_{\mathbf{c}_{\mathbf{h}_a}}(n) \mathbf{x}_{\hat{\mathbf{g}}}(n)}{\mathbf{x}_{\hat{\mathbf{g}}}^T(n) \mathbf{R}_{\mathbf{c}_{\mathbf{h}_a}}(n) \mathbf{x}_{\hat{\mathbf{g}}}(n) + \sigma_v^2}, \quad (29)$$

$$\mathbf{k}_{\mathbf{g}}(n) = \frac{\mathbf{R}_{\mathbf{c}_{\mathbf{g}_a}}(n) \mathbf{x}_{\hat{\mathbf{h}}}(n)}{\mathbf{x}_{\hat{\mathbf{h}}}^T(n) \mathbf{R}_{\mathbf{c}_{\mathbf{g}_a}}(n) \mathbf{x}_{\hat{\mathbf{h}}}(n) + \sigma_v^2}, \quad (30)$$

and

$$\mathbf{R}_{\mathbf{c}_{\mathbf{h}}}(n) = [\mathbf{I}_L - \mathbf{k}_{\mathbf{h}}(n) \mathbf{x}_{\hat{\mathbf{g}}}^T(n)] \mathbf{R}_{\mathbf{c}_{\mathbf{h}_a}}(n), \quad (31)$$

$$\mathbf{R}_{\mathbf{c}_{\mathbf{g}}}(n) = [\mathbf{I}_M - \mathbf{k}_{\mathbf{g}}(n) \mathbf{x}_{\hat{\mathbf{h}}}^T(n)] \mathbf{R}_{\mathbf{c}_{\mathbf{g}_a}}(n). \quad (32)$$

Summarizing, the Kalman filter for bilinear forms (namely KF-BF) is defined by Equations (17), (18), (25), (26), and (29)–(32). As we can notice, the computational complexity of this algorithm is proportional to $\mathcal{O}(L^2 + M^2)$. For the purpose of reducing the computational complexity of the KF-BF, a simplified version of this algorithm is derived. The idea of this low-complexity version was inspired by the work presented in [26], in the context of echo cancellation. First, let us assume that the KF-BF has reached the steady-state convergence. Consequently, $\mathbf{R}_{\mathbf{c}_{\mathbf{h}_a}}(n)$ and $\mathbf{R}_{\mathbf{c}_{\mathbf{g}_a}}(n)$ tend to become diagonal

matrices, which have all the elements on the main diagonal equal to small positive numbers, $\sigma_{\mathbf{c}_{\text{ha}}}^2(n)$ and $\sigma_{\mathbf{c}_{\text{ga}}}^2(n)$, respectively. Therefore, we can use the approximations:

$$\mathbf{R}_{\mathbf{c}_{\text{ha}}}(n) \approx \sigma_{\mathbf{c}_{\text{ha}}}^2(n) \mathbf{I}_L, \quad (33)$$

$$\mathbf{R}_{\mathbf{c}_{\text{ga}}}(n) \approx \sigma_{\mathbf{c}_{\text{ga}}}^2(n) \mathbf{I}_M. \quad (34)$$

Hence, the Kalman gain vectors for the temporal and spatial impulse responses become

$$\mathbf{k}_{\mathbf{h}}(n) = \frac{\mathbf{x}_{\hat{\mathbf{g}}}(n)}{\mathbf{x}_{\hat{\mathbf{g}}}^T(n) \mathbf{x}_{\hat{\mathbf{g}}}(n) + \delta_{\mathbf{h}}(n)}, \quad (35)$$

$$\mathbf{k}_{\mathbf{g}}(n) = \frac{\mathbf{x}_{\hat{\mathbf{h}}}(n)}{\mathbf{x}_{\hat{\mathbf{h}}}^T(n) \mathbf{x}_{\hat{\mathbf{h}}}(n) + \delta_{\mathbf{g}}(n)}, \quad (36)$$

where $\delta_{\mathbf{h}}(n) = \sigma_v^2 / \sigma_{\mathbf{c}_{\text{ha}}}^2(n)$ and $\delta_{\mathbf{g}}(n) = \sigma_v^2 / \sigma_{\mathbf{c}_{\text{ga}}}^2(n)$ can be seen as variable regularization parameters. Then, we use the Kalman vectors from (35) and (36) in the updates (17) and (18), respectively.

Next, a new simplification can be made, by considering that the matrices appearing in the updates of $\mathbf{R}_{\mathbf{c}_{\text{h}}}(n)$ and $\mathbf{R}_{\mathbf{c}_{\text{g}}}(n)$ can be approximated as

$$\mathbf{I}_L - \mathbf{k}_{\mathbf{h}}(n) \mathbf{x}_{\hat{\mathbf{g}}}^T(n) \approx \left[1 - \frac{1}{L} \mathbf{k}_{\mathbf{h}}^T(n) \mathbf{x}_{\hat{\mathbf{g}}}(n) \right] \mathbf{I}_L, \quad (37)$$

$$\mathbf{I}_M - \mathbf{k}_{\mathbf{g}}(n) \mathbf{x}_{\hat{\mathbf{h}}}^T(n) \approx \left[1 - \frac{1}{M} \mathbf{k}_{\mathbf{g}}^T(n) \mathbf{x}_{\hat{\mathbf{h}}}(n) \right] \mathbf{I}_M. \quad (38)$$

We can perform these approximations because, as the filters start to converge, the misalignments of the individual coefficients tend to become uncorrelated; due to this fact, the matrices $\mathbf{R}_{\mathbf{c}_{\text{h}}}(n)$ and $\mathbf{R}_{\mathbf{c}_{\text{g}}}(n)$ tend to become diagonal. Using the notation:

$$\begin{aligned} \mathbf{R}_{\mathbf{c}_{\text{ha}}}(n) &\approx \sigma_{\mathbf{c}_{\text{ha}}}^2(n) \mathbf{I}_L = r_{\mathbf{c}_{\text{ha}}}(n) \mathbf{I}_L, \\ \mathbf{R}_{\mathbf{c}_{\text{ga}}}(n) &\approx \sigma_{\mathbf{c}_{\text{ga}}}^2(n) \mathbf{I}_M = r_{\mathbf{c}_{\text{ga}}}(n) \mathbf{I}_M, \end{aligned} \quad (39)$$

together with

$$\begin{aligned} \mathbf{R}_{\mathbf{c}_{\text{h}}}(n) &\approx \sigma_{\mathbf{c}_{\text{h}}}^2(n) \mathbf{I}_L = r_{\mathbf{c}_{\text{h}}}(n) \mathbf{I}_L, \\ \mathbf{R}_{\mathbf{c}_{\text{g}}}(n) &\approx \sigma_{\mathbf{c}_{\text{g}}}^2(n) \mathbf{I}_M = r_{\mathbf{c}_{\text{g}}}(n) \mathbf{I}_M, \end{aligned} \quad (40)$$

we can summarize the simplified Kalman filter for bilinear forms (SKF-BF) in Table 2. As we can notice, its computational complexity is proportional to $\mathcal{O}(L + M)$, which represents an important gain as compared to KF-BF.

Table 2. Simplified Kalman filter for bilinear forms (SKF-BF).

Initialization:
$\hat{\mathbf{h}}(0) = [1 \ 0 \ \dots \ 0]^T$, $\hat{\mathbf{g}}(0) = \frac{1}{M}[1 \ 1 \ \dots \ 1]^T$
$r_{\mathbf{c}_h}(0) = \epsilon_h$, $r_{\mathbf{c}_g}(0) = \epsilon_g$ (positive constants)
Parameters: $\sigma_{\bar{w}_h}^2$, $\sigma_{\bar{w}_g}^2$, σ_v^2 known or estimated
Algorithm:
$r_{\mathbf{c}_{ha}}(n) = r_{\mathbf{c}_h}(n-1) + \sigma_{\bar{w}_h}^2(n)$
$r_{\mathbf{c}_{ga}}(n) = r_{\mathbf{c}_g}(n-1) + \sigma_{\bar{w}_g}^2(n)$
$\delta_h(n) = \frac{\sigma_v^2}{r_{\mathbf{c}_{ha}}(n)}$
$\delta_g(n) = \frac{\sigma_v^2}{r_{\mathbf{c}_{ga}}(n)}$
$e(n) = d(n) - \mathbf{x}_g^T(n)\hat{\mathbf{h}}(n-1) = d(n) - \mathbf{x}_h^T(n)\hat{\mathbf{g}}(n-1)$
$\hat{\mathbf{h}}(n) = \hat{\mathbf{h}}(n-1) + \frac{\mathbf{x}_g(n)e(n)}{\mathbf{x}_g^T(n)\mathbf{x}_g(n) + \delta_h(n)}$
$\hat{\mathbf{g}}(n) = \hat{\mathbf{g}}(n-1) + \frac{\mathbf{x}_h(n)e(n)}{\mathbf{x}_h^T(n)\mathbf{x}_h(n) + \delta_g(n)}$
$r_{\mathbf{c}_h}(n) = \left\{ 1 - \frac{\mathbf{x}_g^T(n)\mathbf{x}_g(n)}{L[\mathbf{x}_g^T(n)\mathbf{x}_g(n) + \delta_h(n)]} \right\} r_{\mathbf{c}_{ha}}(n)$
$r_{\mathbf{c}_g}(n) = \left\{ 1 - \frac{\mathbf{x}_h^T(n)\mathbf{x}_h(n)}{M[\mathbf{x}_h^T(n)\mathbf{x}_h(n) + \delta_g(n)]} \right\} r_{\mathbf{c}_{ga}}(n)$

4. Optimized LMS Algorithm for Bilinear Forms

In this section, we approach the system identification problem based on the LMS algorithm, aiming to optimize its step-size parameter in order to address the compromise between the main performance criteria, i.e., convergence rate versus misadjustment [27]. In the following, the proposed optimized LMS algorithm for bilinear forms (namely OLMS-BF) is derived based on the same system model given in Section 2. As will be explained in Section 5, this algorithm has striking resemblances with the SKF-BF, even if their derivations follow different patterns.

Let us consider the two estimated impulse responses $\hat{\mathbf{h}}(n)$ and $\hat{\mathbf{g}}(n)$, such that the estimated signal is given by (13). As a consequence, the a priori error signal between the desired signal and the estimated one can be defined following (14), i.e.,

$$\begin{aligned}
 e(n) &= d(n) - \hat{y}(n) \\
 &= [\mathbf{g}(n) \otimes \mathbf{h}(n)]^T \tilde{\mathbf{x}}(n) + v(n) - [\hat{\mathbf{g}}(n-1) \otimes \hat{\mathbf{h}}(n-1)]^T \tilde{\mathbf{x}}(n) \\
 &= \mathbf{h}^T(n)\mathbf{x}_g(n) + v(n) - \hat{\mathbf{h}}^T(n-1)\mathbf{x}_{\hat{\mathbf{g}}}(n) \tag{41}
 \end{aligned}$$

$$= \mathbf{g}^T(n)\mathbf{x}_h(n) + v(n) - \hat{\mathbf{g}}^T(n-1)\mathbf{x}_{\hat{\mathbf{h}}}(n), \tag{42}$$

where $\mathbf{x}_g(n) = [\mathbf{g}(n) \otimes \mathbf{I}_L]^T \tilde{\mathbf{x}}(n)$ and $\mathbf{x}_h(n) = [\mathbf{I}_M \otimes \mathbf{h}(n)]^T \tilde{\mathbf{x}}(n)$, while $\mathbf{x}_{\hat{\mathbf{g}}}(n)$ and $\mathbf{x}_{\hat{\mathbf{h}}}(n)$ are defined in (15) and (16), respectively.

The desired signal $d(n)$ may also be expressed as

$$\begin{aligned}
 d(n) &= \mathbf{g}^T(n)\mathbf{x}_h(n) + \mathbf{g}^T(n)\mathbf{x}_{\hat{\mathbf{h}}}(n) - \mathbf{g}^T(n)\mathbf{x}_{\hat{\mathbf{h}}}(n) + v(n) \\
 &= \mathbf{g}^T(n)\mathbf{x}_{\hat{\mathbf{h}}}(n) + v(n) + v_g(n), \tag{43}
 \end{aligned}$$

where the term:

$$\begin{aligned}
 v_g(n) &= \mathbf{g}^T(n) [\mathbf{x}_h(n) - \mathbf{x}_{\hat{\mathbf{h}}}(n)] \\
 &= \mathbf{c}_{ha}^T(n)\mathbf{x}_g(n) \\
 &= [\mathbf{c}_h(n-1) + \bar{\mathbf{w}}_h(n)]^T \mathbf{x}_g(n) \tag{44}
 \end{aligned}$$

can be seen as an additional “noise” term, of variance $\sigma_{v_g}^2(n)$, introduced by the system $\mathbf{g}$. The terms $\mathbf{c}_h$ and $\mathbf{c}_{h_a}$ were previously defined in (19) and (21), respectively. For the sake of simplicity, the scaling factor η does not appear explicitly in the following. As explained before, this parameter is included in the expression of the uncertainty parameter, thus leading to the notation from (23) and (24).

In a similar way, for the second system we have

$$\begin{aligned} d(n) &= \mathbf{h}^T(n)\mathbf{x}_g(n) + \mathbf{h}^T(n)\mathbf{x}_{\hat{g}}(n) - \mathbf{h}^T(n)\mathbf{x}_{\hat{g}}(n) + v(n) \\ &= \mathbf{h}^T(n)\mathbf{x}_g(n) + v(n) + v_h(n), \end{aligned} \quad (45)$$

where

$$\begin{aligned} v_h(n) &= \mathbf{h}^T(n) [\mathbf{x}_g(n) - \mathbf{x}_{\hat{g}}(n)] \\ &= \mathbf{c}_{g_a}^T(n)\mathbf{x}_h(n) \\ &= [\mathbf{c}_g(n-1) + \overline{\mathbf{w}}_g(n)]^T \mathbf{x}_h(n) \end{aligned} \quad (46)$$

can be interpreted as an additional “noise” term, of variance $\sigma_{v_h}^2(n)$, related to the system $\mathbf{h}$. Here, the a posteriori and a priori misalignments corresponding to the system $\mathbf{g}$ were defined in (20) and (22), respectively. In Figures 1 and 2, the equivalent system identification scheme is represented in terms of the two components, $\mathbf{g}(n)$ and $\mathbf{h}(n)$, respectively; it can be observed that each system influences the other one through the additional “noise” term.

Figure 1. Equivalent system identification scheme when considering the system $\mathbf{g}(n)$ and the input $\mathbf{x}_h(n)$.

Figure 2. Equivalent system identification scheme when considering the system $\mathbf{h}(n)$ and the input $\mathbf{x}_g(n)$.

In the framework of the LMS algorithm for bilinear forms, namely LMS-BF [12], the updates are the following:

$$\hat{\mathbf{g}}(n) = \hat{\mathbf{g}}(n-1) + \mu_{\hat{\mathbf{g}}} \mathbf{x}_{\hat{\mathbf{h}}}(n) e(n), \quad (47)$$

$$\hat{\mathbf{h}}(n) = \hat{\mathbf{h}}(n-1) + \mu_{\hat{\mathbf{h}}} \mathbf{x}_{\hat{\mathbf{g}}}(n) e(n), \quad (48)$$

where $\mu_{\hat{\mathbf{g}}}$ and $\mu_{\hat{\mathbf{h}}}$ are the step-size parameters. In this context, the vectors corresponding to the a posteriori misalignments become

$$\begin{aligned} \mathbf{c}_{\mathbf{g}}(n) &= \mathbf{g}(n) - \hat{\mathbf{g}}(n-1) - \mu_{\hat{\mathbf{g}}} \mathbf{x}_{\hat{\mathbf{h}}}(n) e(n) \\ &= \mathbf{c}_{\mathbf{g}}(n-1) + \bar{\mathbf{w}}_{\mathbf{g}}(n) - \mu_{\hat{\mathbf{g}}} \mathbf{x}_{\hat{\mathbf{h}}}(n) e(n), \end{aligned} \quad (49)$$

$$\begin{aligned} \mathbf{c}_{\mathbf{h}}(n) &= \mathbf{h}(n) - \hat{\mathbf{h}}(n-1) - \mu_{\hat{\mathbf{h}}} \mathbf{x}_{\hat{\mathbf{g}}}(n) e(n) \\ &= \mathbf{c}_{\mathbf{h}}(n-1) + \bar{\mathbf{w}}_{\mathbf{h}}(n) - \mu_{\hat{\mathbf{h}}} \mathbf{x}_{\hat{\mathbf{g}}}(n) e(n). \end{aligned} \quad (50)$$

At this point, let us introduce the notation $m_{\mathbf{g}}(n) = \mathbb{E} [\|\mathbf{c}_{\mathbf{g}}(n)\|^2]$ and $m_{\mathbf{h}}(n) = \mathbb{E} [\|\mathbf{c}_{\mathbf{h}}(n)\|^2]$. Taking the square ℓ_2 norms in both sides of (49) and (50), respectively, we can recursively evaluate

$$\begin{aligned} m_{\mathbf{g}}(n) &= m_{\mathbf{g}}(n-1) - 2\mu_{\hat{\mathbf{g}}} \mathbb{E} \{ [\mathbf{c}_{\mathbf{g}}^T(n-1) + \bar{\mathbf{w}}_{\mathbf{g}}^T(n)] \mathbf{x}_{\hat{\mathbf{h}}}(n) e(n) \} + \mu_{\hat{\mathbf{g}}}^2 \mathbb{E} [\mathbf{x}_{\hat{\mathbf{h}}}^T(n) \mathbf{x}_{\hat{\mathbf{h}}}(n) e^2(n)] + M\sigma_{\bar{\mathbf{w}}_{\mathbf{g}}}^2 \\ &= m_{\mathbf{g}}(n-1) - 2A_{\mathbf{g}}\mu_{\hat{\mathbf{g}}} + \mu_{\hat{\mathbf{g}}}^2 B_{\mathbf{g}} + M\sigma_{\bar{\mathbf{w}}_{\mathbf{g}}}^2, \end{aligned} \quad (51)$$

$$\begin{aligned} m_{\mathbf{h}}(n) &= m_{\mathbf{h}}(n-1) - 2\mu_{\hat{\mathbf{h}}} \mathbb{E} \{ [\mathbf{c}_{\mathbf{h}}^T(n-1) + \bar{\mathbf{w}}_{\mathbf{h}}^T(n)] \mathbf{x}_{\hat{\mathbf{g}}}(n) e(n) \} + \mu_{\hat{\mathbf{h}}}^2 \mathbb{E} [\mathbf{x}_{\hat{\mathbf{g}}}^T(n) \mathbf{x}_{\hat{\mathbf{g}}}(n) e^2(n)] + L\sigma_{\bar{\mathbf{w}}_{\mathbf{h}}}^2 \\ &= m_{\mathbf{h}}(n-1) - 2A_{\mathbf{h}}\mu_{\hat{\mathbf{h}}} + \mu_{\hat{\mathbf{h}}}^2 B_{\mathbf{h}} + L\sigma_{\bar{\mathbf{w}}_{\mathbf{h}}}^2, \end{aligned} \quad (52)$$

where

$$A_{\mathbf{g}} = \mathbb{E} \{ [\mathbf{c}_{\mathbf{g}}^T(n-1) + \bar{\mathbf{w}}_{\mathbf{g}}^T(n)] \mathbf{x}_{\hat{\mathbf{h}}}(n) e(n) \}, \quad (53)$$

$$B_{\mathbf{g}} = \mathbb{E} [\mathbf{x}_{\hat{\mathbf{h}}}^T(n) \mathbf{x}_{\hat{\mathbf{h}}}(n) e^2(n)], \quad (54)$$

$$A_{\mathbf{h}} = \mathbb{E} \{ [\mathbf{c}_{\mathbf{h}}^T(n-1) + \bar{\mathbf{w}}_{\mathbf{h}}^T(n)] \mathbf{x}_{\hat{\mathbf{g}}}(n) e(n) \}, \quad (55)$$

$$B_{\mathbf{h}} = \mathbb{E} [\mathbf{x}_{\hat{\mathbf{g}}}^T(n) \mathbf{x}_{\hat{\mathbf{g}}}(n) e^2(n)]. \quad (56)$$

It is very difficult to further process the expectation terms from (53)–(56) (and, consequently, (51) and (52)) without any supporting assumptions on the character of the input signals. Hence, let us consider that the covariance matrices of the inputs are close to a diagonal one. This is a fairly restrictive assumption on the input signals, which has been widely used to simplify the convergence analysis of many adaptive algorithms [27,28]. Also, let us consider that the input signals are independent and have the same power. In this context, the computations of the expectation terms from (53)–(56) are detailed in Appendix A. Summarizing the results from this appendix, these terms result in

$$A_{\mathbf{g}} = p_{\mathbf{g}}(n) + \sigma_{\tilde{\mathbf{x}}}^2 \mathbb{E} [\|\hat{\mathbf{h}}(n-1)\|^2] [m_{\mathbf{g}}(n-1) + M\sigma_{\bar{\mathbf{w}}_{\mathbf{g}}}^2], \quad (57)$$

$$A_{\mathbf{h}} = p_{\mathbf{h}}(n) + \sigma_{\tilde{\mathbf{x}}}^2 \mathbb{E} [\|\hat{\mathbf{g}}(n-1)\|^2] [m_{\mathbf{h}}(n-1) + L\sigma_{\bar{\mathbf{w}}_{\mathbf{h}}}^2], \quad (58)$$

$$B_{\mathbf{g}} = M\sigma_{\tilde{\mathbf{x}}}^2 \mathbb{E} [\|\hat{\mathbf{h}}(n-1)\|^2] \left\{ \sigma_v^2 + 2p_{\mathbf{g}}(n) + \frac{1}{L}\sigma_{\bar{\mathbf{w}}_{\mathbf{g}}}^2(n) + \sigma_{\tilde{\mathbf{x}}}^2 \mathbb{E} [\|\hat{\mathbf{h}}(n-1)\|^2] [m_{\mathbf{g}}(n-1) + M\sigma_{\bar{\mathbf{w}}_{\mathbf{g}}}^2] \right\}, \quad (59)$$

$$B_{\mathbf{h}} = L\sigma_{\tilde{\mathbf{x}}}^2 \mathbb{E} [\|\hat{\mathbf{g}}(n-1)\|^2] \left\{ \sigma_v^2 + 2p_{\mathbf{h}}(n) + \frac{1}{M}\sigma_{\bar{\mathbf{w}}_{\mathbf{h}}}^2(n) + \sigma_{\tilde{\mathbf{x}}}^2 \mathbb{E} [\|\hat{\mathbf{g}}(n-1)\|^2] [m_{\mathbf{h}}(n-1) + L\sigma_{\bar{\mathbf{w}}_{\mathbf{h}}}^2] \right\}, \quad (60)$$

where $\sigma_{\tilde{\mathbf{x}}}^2 = \mathbb{E} [\|\tilde{\mathbf{x}}(n)\|^2]$ and the terms denoted by $p_{\mathbf{g}}(n)$ and $p_{\mathbf{h}}(n)$ are evaluated as

$$\begin{aligned} p_{\mathbf{g}}(n) &= \mathbb{E} \{ \mathbf{c}_{\mathbf{g}}^T(n-1) \mathbf{x}_{\hat{\mathbf{h}}}(n) \tilde{\mathbf{x}}^T(n) [\mathbf{g} \otimes \mathbf{c}_{\mathbf{h}}(n-1)] \} \\ &= \mathbb{E} \{ \mathbf{c}_{\mathbf{g}}^T(n-1) \mathbf{x}_{\hat{\mathbf{h}}}(n) \mathbf{c}_{\mathbf{h}}^T(n-1) \mathbf{x}_{\hat{\mathbf{g}}}(n) \}, \end{aligned} \quad (61)$$

$$\begin{aligned} p_{\mathbf{h}}(n) &= \mathbb{E}\{\mathbf{c}_{\mathbf{h}}^T(n-1)\mathbf{x}_{\hat{\mathbf{g}}}(n)\tilde{\mathbf{x}}^T(n)[\mathbf{c}_{\mathbf{g}}(n-1) \otimes \mathbf{h}]\} \\ &= \mathbb{E}\{\mathbf{c}_{\mathbf{h}}^T(n-1)\mathbf{x}_{\hat{\mathbf{g}}}(n)\mathbf{c}_{\hat{\mathbf{g}}}^T(n-1)\mathbf{x}_{\mathbf{h}}(n)\}. \end{aligned} \quad (62)$$

The expressions of the variances $\sigma_{v_{\mathbf{g}}}^2(n)$ and $\sigma_{v_{\mathbf{h}}}^2(n)$ are also provided in Appendix A. Consequently, using (57)–(60) in (51) and (52), we obtain

$$\begin{aligned} m_{\mathbf{g}}(n) &= m_{\mathbf{g}}(n-1) \left\{ 1 - 2\mu_{\hat{\mathbf{g}}}\sigma_{\mathbf{x}}^2\mathbb{E}[\|\hat{\mathbf{h}}(n-1)\|^2] + \mu_{\hat{\mathbf{g}}}^2\sigma_{\mathbf{x}}^4M\{\mathbb{E}[\|\hat{\mathbf{h}}(n-1)\|^2]\}^2 \right\} - 2\mu_{\hat{\mathbf{g}}}p_{\mathbf{g}}(n) \\ &\quad + \mu_{\hat{\mathbf{g}}}^2M\sigma_{\mathbf{x}}^2\mathbb{E}[\|\hat{\mathbf{h}}(n-1)\|^2] \left\{ \sigma_v^2 + \frac{1}{L}\sigma_{v_{\mathbf{g}}}^2(n) + M\sigma_{\hat{\mathbf{w}}_{\mathbf{g}}}^2\sigma_{\mathbf{x}}^2\mathbb{E}[\|\hat{\mathbf{h}}(n-1)\|^2] + 2p_{\mathbf{g}}(n) \right\} \\ &\quad + M\sigma_{\hat{\mathbf{w}}_{\mathbf{g}}}^2 \left\{ 1 - 2\mu_{\hat{\mathbf{g}}}\sigma_{\mathbf{x}}^2\mathbb{E}[\|\hat{\mathbf{h}}(n-1)\|^2] \right\}, \end{aligned} \quad (63)$$

$$\begin{aligned} m_{\mathbf{h}}(n) &= m_{\mathbf{h}}(n-1) \left\{ 1 - 2\mu_{\hat{\mathbf{h}}}\sigma_{\mathbf{x}}^2\mathbb{E}[\|\hat{\mathbf{g}}(n-1)\|^2] + \mu_{\hat{\mathbf{h}}}^2\sigma_{\mathbf{x}}^4L\{\mathbb{E}[\|\hat{\mathbf{g}}(n-1)\|^2]\}^2 \right\} - 2\mu_{\hat{\mathbf{h}}}p_{\mathbf{h}}(n) \\ &\quad + \mu_{\hat{\mathbf{h}}}^2L\sigma_{\mathbf{x}}^2\mathbb{E}[\|\hat{\mathbf{g}}(n-1)\|^2] \left\{ \sigma_v^2 + \frac{1}{M}\sigma_{v_{\mathbf{h}}}^2(n) + L\sigma_{\hat{\mathbf{w}}_{\mathbf{h}}}^2\sigma_{\mathbf{x}}^2\mathbb{E}[\|\hat{\mathbf{g}}(n-1)\|^2] + 2p_{\mathbf{h}}(n) \right\} \\ &\quad + L\sigma_{\hat{\mathbf{w}}_{\mathbf{h}}}^2 \left\{ 1 - 2\mu_{\hat{\mathbf{h}}}\sigma_{\mathbf{x}}^2\mathbb{E}[\|\hat{\mathbf{g}}(n-1)\|^2] \right\}. \end{aligned} \quad (64)$$

In the context of system identification problems, the main goal is to reduce the system misalignment, which basically represents the difference between the true impulse response and the estimated one. Therefore, in our framework, the optimal step-size parameters (denoted in the following by $\mu_{\hat{\mathbf{g}},o}$ and $\mu_{\hat{\mathbf{h}},o}$) can be found by minimizing (63) and (64). This is done by canceling the derivatives of (63) and (64) with respect to the step-sizes, which result in:

$$\frac{\partial m_{\mathbf{g}}(n)}{\partial \mu_{\hat{\mathbf{g}}}} = -2A_{\mathbf{g}} + 2B_{\mathbf{g}}\mu_{\hat{\mathbf{g}}} = 0 \Rightarrow \mu_{\hat{\mathbf{g}},o} = \frac{A_{\mathbf{g}}}{B_{\mathbf{g}}}, \quad (65)$$

$$\frac{\partial m_{\mathbf{h}}(n)}{\partial \mu_{\hat{\mathbf{h}}}} = -2A_{\mathbf{h}} + 2B_{\mathbf{h}}\mu_{\hat{\mathbf{h}}} = 0 \Rightarrow \mu_{\hat{\mathbf{h}},o} = \frac{A_{\mathbf{h}}}{B_{\mathbf{h}}}. \quad (66)$$

By replacing $A_{\mathbf{g}}$, $B_{\mathbf{g}}$, $A_{\mathbf{h}}$, and $B_{\mathbf{h}}$ with their expressions (see (57)–(60)), the step-size parameters of the proposed optimized LMS algorithm for bilinear forms (namely OLMS-BF) are found. Finally, introducing these parameters in (47) and (48), the updates of the OLMS-BF algorithm become

$$\begin{aligned} \hat{\mathbf{g}}(n) &= \hat{\mathbf{g}}(n-1) + \mu_{\hat{\mathbf{g}},o}(n)x_{\hat{\mathbf{h}}}(n)e(n) \\ &= \hat{\mathbf{g}}(n-1) + \frac{x_{\hat{\mathbf{h}}}(n)e(n)}{M\sigma_{\mathbf{x}}^2\mathbb{E}[\|\hat{\mathbf{h}}(n-1)\|^2] \left\{ 1 + \frac{p_{\mathbf{g}}(n) + \sigma_v^2 + \frac{1}{L}\sigma_{v_{\mathbf{g}}}^2(n)}{p_{\mathbf{g}}(n) + \sigma_{\mathbf{x}}^2\mathbb{E}[\|\hat{\mathbf{h}}(n-1)\|^2] [m_{\mathbf{g}}(n-1) + M\sigma_{\hat{\mathbf{w}}_{\mathbf{g}}}^2]} \right\}}, \end{aligned} \quad (67)$$

$$\begin{aligned} \hat{\mathbf{h}}(n) &= \hat{\mathbf{h}}(n-1) + \mu_{\hat{\mathbf{h}},o}(n)x_{\hat{\mathbf{g}}}(n)e(n) \\ &= \hat{\mathbf{h}}(n-1) + \frac{x_{\hat{\mathbf{g}}}(n)e(n)}{L\sigma_{\mathbf{x}}^2\mathbb{E}[\|\hat{\mathbf{g}}(n-1)\|^2] \left\{ 1 + \frac{p_{\mathbf{h}}(n) + \sigma_v^2 + \frac{1}{M}\sigma_{v_{\mathbf{h}}}^2(n)}{p_{\mathbf{h}}(n) + \sigma_{\mathbf{x}}^2\mathbb{E}[\|\hat{\mathbf{g}}(n-1)\|^2] [m_{\mathbf{h}}(n-1) + L\sigma_{\hat{\mathbf{w}}_{\mathbf{h}}}^2]} \right\}}. \end{aligned} \quad (68)$$

The most problematic terms in (67) and (68) are $p_{\mathbf{g}}(n)$ and $p_{\mathbf{h}}(n)$ (from (61) and (62), respectively), which depend on the true impulse responses. However, as shown in the next section, these terms could be omitted in practice.

5. SKF-BF versus OLMS-BF

The SKF-BF and OLMS-BF algorithms were developed following different theoretical patterns. However, there are strong similarities between these two algorithms, as will be explained in this section.

The update equations of the SKF-BF are given by (17) and (18), where the Kalman gain vectors have the expressions in (35) and (36). It can be noticed that the updates of the SKF-BF can be expressed as

$$\hat{\mathbf{g}}(n) = \hat{\mathbf{g}}(n-1) + \mu_{\hat{\mathbf{g}},\mathbf{K}}(n)x_{\hat{\mathbf{h}}}(n)e(n), \quad (69)$$

$$\hat{\mathbf{h}}(n) = \hat{\mathbf{h}}(n-1) + \mu_{\hat{\mathbf{h}},\mathbf{K}}(n)x_{\hat{\mathbf{g}}}(n)e(n), \quad (70)$$

where the Kalman step-size parameters are

$$\mu_{\hat{\mathbf{g}},\mathbf{K}}(n) = \frac{1}{\mathbf{x}_{\hat{\mathbf{h}}}^T(n)\mathbf{x}_{\hat{\mathbf{h}}}(n) \left\{ 1 + \frac{M\sigma_v^2}{[m_{\mathbf{g}}(n-1) + M\sigma_{\bar{w}_{\mathbf{g}}}^2] \mathbf{x}_{\hat{\mathbf{h}}}^T(n)\mathbf{x}_{\hat{\mathbf{h}}}(n)} \right\}}, \quad (71)$$

$$\mu_{\hat{\mathbf{h}},\mathbf{K}}(n) = \frac{1}{\mathbf{x}_{\hat{\mathbf{g}}}^T(n)\mathbf{x}_{\hat{\mathbf{g}}}(n) \left\{ 1 + \frac{L\sigma_v^2}{[m_{\mathbf{h}}(n-1) + L\sigma_{\bar{w}_{\mathbf{h}}}^2] \mathbf{x}_{\hat{\mathbf{g}}}^T(n)\mathbf{x}_{\hat{\mathbf{g}}}(n)} \right\}}. \quad (72)$$

Comparing these parameters with the optimal step-sizes from (67) and (68) (also taking (A7) and (A8) into account; see Appendix A), we can notice striking resemblances between SKF-BF and OLMS-BF. In fact, these two algorithms are very similar when

$$p_{\mathbf{g}} = p_{\mathbf{h}} = 0. \quad (73)$$

On the other hand, as it was indicated in [12], this could represent a reasonable assumption, since

$$\lim_{n \rightarrow \infty} p_{\mathbf{g}}(n) = \lim_{n \rightarrow \infty} p_{\mathbf{h}}(n) = 0, \quad (74)$$

suggesting that in the steady-state of the algorithm, the influence of the terms $p_{\mathbf{h}}$ and $p_{\mathbf{g}}$ on the step-size parameters diminishes. As will be supported in simulations, (73) can be fairly imposed within the OLMS-BF algorithm, while still leading to a very good compromise between the performance criteria (e.g., convergence rate versus misadjustment). Under these considerations, the OLMS-BF algorithm is summarized in Table 3 (in a practical form that facilitates its implementation).

Table 3. Optimized LMS algorithm for bilinear forms (OLMS-BF) (practical version, i.e., using (73)).

Initialization:
$\hat{\mathbf{h}}(0) = [1 \ 0 \ \dots \ 0]^T$, $\hat{\mathbf{g}}(0) = \frac{1}{M}[1 \ 1 \ \dots \ 1]^T$
$m_{\mathbf{h}}(0) = \epsilon_{\mathbf{h}}$, $m_{\mathbf{g}}(0) = \epsilon_{\mathbf{g}}$ (positive constants)
Parameters: $\mathbb{E}[\ \mathbf{h}(n)\ ^2]$, $\mathbb{E}[\ \mathbf{g}(n)\ ^2]$, $\sigma_{\bar{w}_{\mathbf{h}}}^2$, $\sigma_{\bar{w}_{\mathbf{g}}}^2$, σ_v^2 known or estimated
Algorithm:
$\sigma_{v_{\mathbf{h}}}^2(n) = M\sigma_{\mathbf{x}}^2\mathbb{E}[\ \mathbf{h}(n)\ ^2] \left[m_{\mathbf{g}}(n-1) + M\sigma_{\bar{w}_{\mathbf{g}}}^2 \right]$
$\sigma_{v_{\mathbf{g}}}^2(n) = L\sigma_{\mathbf{x}}^2\mathbb{E}[\ \mathbf{g}(n)\ ^2] \left[m_{\mathbf{h}}(n-1) + L\sigma_{\bar{w}_{\mathbf{h}}}^2 \right]$
$A_{\mathbf{h}} = \sigma_{\mathbf{x}}^2\mathbb{E}[\ \hat{\mathbf{g}}(n-1)\ ^2] \left[m_{\mathbf{h}}(n-1) + L\sigma_{\bar{w}_{\mathbf{h}}}^2 \right]$
$B_{\mathbf{h}} = L\sigma_{\mathbf{x}}^2\mathbb{E}[\ \hat{\mathbf{g}}(n-1)\ ^2] \left\{ \sigma_v^2 + \frac{1}{M}\sigma_{v_{\mathbf{h}}}^2(n) + \sigma_{\mathbf{x}}^2\mathbb{E}[\ \hat{\mathbf{g}}(n-1)\ ^2] \left[m_{\mathbf{h}}(n-1) + L\sigma_{\bar{w}_{\mathbf{h}}}^2 \right] \right\}$
$A_{\mathbf{g}} = \sigma_{\mathbf{x}}^2\mathbb{E}[\ \hat{\mathbf{h}}(n-1)\ ^2] \left[m_{\mathbf{g}}(n-1) + M\sigma_{\bar{w}_{\mathbf{g}}}^2 \right]$
$B_{\mathbf{g}} = M\sigma_{\mathbf{x}}^2\mathbb{E}[\ \hat{\mathbf{h}}(n-1)\ ^2] \left\{ \sigma_v^2 + \frac{1}{L}\sigma_{v_{\mathbf{g}}}^2(n) + \sigma_{\mathbf{x}}^2\mathbb{E}[\ \hat{\mathbf{h}}(n-1)\ ^2] \left[m_{\mathbf{g}}(n-1) + M\sigma_{\bar{w}_{\mathbf{g}}}^2 \right] \right\}$
$\mu_{\hat{\mathbf{h}},o} = \frac{A_{\mathbf{h}}}{B_{\mathbf{h}}}$
$\mu_{\hat{\mathbf{g}},o} = \frac{A_{\mathbf{g}}}{B_{\mathbf{g}}}$
$e(n) = d(n) - \mathbf{x}_{\hat{\mathbf{g}}}^T(n)\hat{\mathbf{h}}(n-1) = d(n) - \mathbf{x}_{\hat{\mathbf{h}}}^T(n)\hat{\mathbf{g}}(n-1)$
$\hat{\mathbf{h}}(n) = \hat{\mathbf{h}}(n-1) + \mu_{\hat{\mathbf{h}},o}x_{\hat{\mathbf{g}}}(n)e(n)$
$\hat{\mathbf{g}}(n) = \hat{\mathbf{g}}(n-1) + \mu_{\hat{\mathbf{g}},o}x_{\hat{\mathbf{h}}}(n)e(n)$
$m_{\mathbf{h}}(n) = m_{\mathbf{h}}(n-1) - 2A_{\mathbf{h}}\mu_{\hat{\mathbf{h}},o} + \mu_{\hat{\mathbf{h}},o}^2B_{\mathbf{h}} + L\sigma_{\bar{w}_{\mathbf{h}}}^2$
$m_{\mathbf{g}}(n) = m_{\mathbf{g}}(n-1) - 2A_{\mathbf{g}}\mu_{\hat{\mathbf{g}},o} + \mu_{\hat{\mathbf{g}},o}^2B_{\mathbf{g}} + M\sigma_{\bar{w}_{\mathbf{g}}}^2$

6. Practical Considerations

The previously developed algorithms are designed to identify the individual impulse responses of the bilinear form. The global (spatiotemporal) impulse response can be computed based on the Kronecker product between them. An alternative solution is to use the regular Kalman filter to identify the spatiotemporal impulse response directly, relying on the observation Equation (4) and identifying the state equation:

$$\mathbf{f}(n) = \mathbf{f}(n-1) + \mathbf{w}(n), \quad (75)$$

where $\mathbf{w}(n)$ is a zero-mean white Gaussian noise signal vector. The covariance matrix of $\mathbf{w}(n)$ is $\mathbf{R}_w(n) = \sigma_w^2 \mathbf{I}_{ML}$, where $\mathbf{I}_{ML}$ is the identity matrix of size $ML \times ML$ and the variance σ_w^2 captures the uncertainties in $\mathbf{f}(n)$.

In this way, following the approach from [26], we can easily derive the regular Kalman filter (KF) and its simplified version (namely SKF), which can identify the global impulse response using a single adaptive filter $\hat{\mathbf{f}}(n)$; for further details, please see Sections VI and VII in [26]. However, we need to mention that the solution found using the regular KF and SKF involves an adaptive filter of length ML , whereas their counterparts tailored for bilinear forms (i.e., KF-BF and SKF-BF) use two shorter filters of lengths L and M , respectively. As a consequence, besides a lower computational complexity, a much faster converge rate and tracking are expected for the bilinear algorithms with respect to the conventional ones. The same ideas apply for the OLMS-BF algorithm, as compared to its regular counterpart, i.e., the joint-optimized normalized LMS (JO-NLMS) algorithm [29], which could be used to identify the global impulse response $\hat{\mathbf{f}}(n)$.

The computational complexity of the previously discussed algorithms is summarized in Table 4. It can be easily seen that the SKF-BF offer a great reduction in terms of complexity with respect to KF-BF. Also, the SKF-BF and OLMS-BF differ only by a small number of operations, thus confirming the similarity that was highlighted in Section 5. Finally, when $ML \gg M + L$ (which is usually the case in practice), we can notice that the algorithms tailored for bilinear forms (namely KF-BF, SKF-BF, and OLMS-BF) offer lower computational complexities as compared to their regular counterparts (i.e., KF, SKF, and JO-NLMS, respectively).

Table 4. Computational complexity of the algorithms.

Algorithms	$\times$	$+$	$/$
KF-BF [13]	$3(L^2 + M^2) + 2ML + 3L + 4M$	$2(L^2 + M^2) + 2ML + L + 2M$	2
KF [26]	$3(ML)^2 + 4ML$	$2(ML)^2 + 3ML$	1
SKF-BF (Table 2)	$2ML + 2L + 3M + 6$	$2ML + L + 2M + 4$	2
SKF [26]	$3ML + 6$	$3ML + 4$	1
OLMS-BF (Table 3)	$2ML + 2L + 3M + 12$	$2ML + L + 2M + 8$	2
JO-NLMS [29]	$3ML + 6$	$3ML + 5$	1

Next, a few important observations ought to be made regarding the specific parameters that must be set within the algorithms. Here, the noise power σ_v^2 is required in order to compute the Kalman gain vectors (for KF-BF and SKF-BF) or the optimal step-sizes (for OLMS-BF). In practice, we can estimate this parameter in different ways; some simple and efficient methods for this purpose are presented in [30,31]. Although there are different other methods that can be used to estimate the noise power, the analysis of their influence on the performance of the algorithms lies beyond the scope of this paper.

The parameters related to the uncertainties in the unknown systems also need to be set or estimated, i.e., $\sigma_{w_h}^2$ and $\sigma_{w_g}^2$. Choosing small values for these parameters yields a small misalignment, but at the same time a poor tracking. On the other hand, large values (meaning that there are high uncertainties in the unknown systems) lead to a good tracking but also a high misalignment. This means that we always need to have a good compromise between fast tracking and low misalignment. In practice, if we have some a priori information about the systems which we need to

identify, we can take it into consideration when setting the values of these parameters. For example, if we assume the spatial impulse response to be time-invariant, we could fix $\sigma_{\bar{w}_g}^2 = 0$ and tune only the parameter related to the temporal impulse response. Thus, based on the state equation related to $\mathbf{h}(n)$, together with the approximation $\|\bar{\mathbf{w}}_{\mathbf{h}}(n)\|^2 \approx L\sigma_{\bar{w}_h}^2$ (which is valid when $L \gg 1$), and replacing $\mathbf{h}(n)$ and $\mathbf{h}(n-1)$ by their estimates, we can evaluate

$$\hat{\sigma}_{\bar{w}_h}^2(n) = \frac{1}{L} \|\hat{\mathbf{h}}(n) - \hat{\mathbf{h}}(n-1)\|^2. \quad (76)$$

It can be noticed that the estimation from (76) is designed to achieve a proper compromise between good tracking and low misalignment. When the algorithm starts to converge or when there is an abrupt change of the system, the difference between $\hat{\mathbf{h}}(n)$ and $\hat{\mathbf{h}}(n-1)$ is significant, leading to large values of the parameter $\hat{\sigma}_{\bar{w}_h}^2(n)$, therefore providing fast convergence and tracking. On the contrary, when the algorithm is converging to its steady-state, the difference between $\hat{\mathbf{h}}(n)$ and $\hat{\mathbf{h}}(n-1)$ reduces, thus leading to the parameter $\hat{\sigma}_{\bar{w}_h}^2(n)$ taking small values and, consequently, to a low misalignment.

7. Results

Experiments are performed in the context of system identification, in order to highlight the performance of the Kalman-based algorithms for bilinear forms (referred to as KF-BF and SKF-BF), in comparison with their regular counterparts (KF and SKF, as mentioned in Section 6). Also, we aim to evaluate the features of the OLMS-BF algorithm, as compared to other existing solutions, e.g., the normalized LMS-BF (NLMS-BF) and the JO-NLMS algorithms, which were introduced in [12] and [29], respectively.

In most of the experiments, both the temporal and the spatial impulse responses are randomly generated from a Gaussian distribution, having the lengths equal to $L = 64$ and $M = 8$, respectively. This leads to a length of the spatiotemporal impulse response equal to $ML = 8 \times 64 = 512$. It is also useful to evaluate the tracking capabilities of the algorithms; to this purpose, a sudden change in the temporal impulse response is applied in the middle of simulations, by generating a new random vector of length $L = 64$, also from a Gaussian distribution. Only in the last experiment, the impulse response $\mathbf{h}(n)$ is an acoustic echo path of length $L = 512$, while the coefficients of $\mathbf{g}(n)$ are computed as $g_m(n) = 0.5^m$, with $m = 1, \dots, M$ and $M = 4$; in this case, the length of the global system is $ML = 4 \times 512 = 2048$.

The input signals $x_m(n)$, $m = 1, 2, \dots, M$ are either white Gaussian noises (WGNs) or AR(1) processes (which were obtained after passing a white Gaussian noise through a first-order system with the transfer function $1/(1 - 0.8z^{-1})$). The additive noise $v(n)$ is white and Gaussian, having the variance $\sigma_v^2 = 0.01$; we assume that this parameter is available in the experiments. In most of simulations, the measure of the performance is the NM (in dB) (see (10)), to evaluate the identification of the global impulse response. In addition, in the second set of experiments (focusing on the OLMS-BF algorithm), we also involve the NPMs (based on (8) and (9)), related to the individual impulse responses.

In Figures 3 and 4, the KF-BF is compared to the regular KF for WGN and AR(1) input signals, respectively. The specific parameters of the algorithms are set to $\sigma_{\bar{w}_h}^2 = \sigma_{\bar{w}_g}^2 = \sigma_w^2 = 10^{-9}$. It can be noticed from both figures that the KF-BF achieves a faster convergence rate as compared to the regular KF, for both types of input signals, providing also a better tracking capability. The gain is even more apparent in case of AR(1) inputs.

Figure 3. Normalized misalignment of the KF-BF and regular KF using WGNs as input signals. The length of the global impulse response is $ML = 512$. The specific parameters are set to $\sigma_{\bar{w}_h}^2 = \sigma_{\bar{w}_g}^2 = \sigma_w^2 = 10^{-9}$.

Figure 4. Normalized misalignment of the KF-BF and regular KF using AR(1) processes as input signals. The length of the global impulse response is $ML = 512$. The specific parameters are set to $\sigma_{\bar{w}_h}^2 = \sigma_{\bar{w}_g}^2 = \sigma_w^2 = 10^{-9}$.

The previous experiment is repeated (for the same two types of inputs) in Figures 5 and 6, this time comparing the SKF-BF with the regular SKF [26]. As it can be observed, the simplified versions (SKF-BF and SKF) yield a slower convergence rate (especially in case of AR(1) inputs) as compared to the full versions (KF-BF and KF, respectively); however, the computational complexities for these simplified versions are much lower. As it was expected, the SKF-BF outperforms the regular SKF in terms of the convergence rate; the improvement is much more visible in the case of AR(1) inputs.

Figure 5. Normalized misalignment of the SKF-BF and regular SKF using WGNs as input signals. The length of the global impulse response is $ML = 512$. The specific parameters are set to $\sigma_{\bar{w}_h}^2 = \sigma_{\bar{w}_g}^2 = \sigma_w^2 = 10^{-9}$.

Figure 6. Normalized misalignment of the SKF-BF and regular SKF using AR(1) processes as input signals. The length of the global impulse response is $ML = 512$. The specific parameters are set to $\sigma_{\bar{w}_h}^2 = \sigma_{\bar{w}_g}^2 = \sigma_w^2 = 10^{-9}$.

Next, the performance of the SKF-BF is evaluated in Figures 7 and 8, but using the recursive estimation $\hat{\sigma}_{\bar{w}_h}^2(n)$ from (76) (not a constant value, as in the previous experiments). The spatial impulse response is assumed to be time invariant, so that we can set $\sigma_{\bar{w}_g}^2 = 0$. The regular SKF is considered for comparison, using a similar way to estimate its specific parameter, i.e., $\hat{\sigma}_w^2(n) = 1/(ML) \|\hat{\mathbf{f}}(n) - \hat{\mathbf{f}}(n-1)\|^2$ [26]. Because of the nature of the estimators (as it was explained in Section 6), the algorithms behave like variable step-size adaptive filters, achieving both low misalignment and fast convergence/tracking. Moreover, as we can notice from these two figures, the proposed SKF-BF still outperforms the regular SKF in terms of both performance criteria.

Figure 7. Normalized misalignment of the SKF-BF and regular SKF (for WGNs input signals), using the recursive estimates $\hat{\sigma}_{\bar{w}_h}^2(n)$ and $\hat{\sigma}_w^2(n)$, respectively; the SKF-BF uses $\sigma_{\bar{w}_g}^2 = 0$. The length of the global impulse responses is $ML = 512$.

Figure 8. Normalized misalignment of the SKF-BF and regular SKF (for AR(1) input signals), using the recursive estimates $\hat{\sigma}_{\bar{w}_h}^2(n)$ and $\hat{\sigma}_w^2(n)$, respectively; the SKF-BF uses $\sigma_{\bar{w}_g}^2 = 0$. The length of the global impulse responses is $ML = 512$.

As outlined in Section 5, there are strong similarities between the SKF-BF and OLMS-BF algorithms. In Figures 9 and 10, we compare the performances of these algorithms using two types of input signals, i.e., WGNs and AR(1) processes, respectively. Both algorithms use the recursive estimate $\hat{\sigma}_{\bar{w}_h}^2(n)$ (from (76)) and $\sigma_{\bar{w}_g}^2 = 0$. As we can notice, the SKF-BF and OLMS-BF algorithms behave quite similar, especially when the input signals are WGNs (Figure 9). When the input signals are AR(1) processes (Figure 10), the SKF-BF outperforms the OLMS-BF in terms of the initial convergence rate; however, it pays with a slower tracking reaction. Nevertheless, the overall performances of these algorithms are very similar, as supported by the comparison provided in Section 5.

Figure 9. Normalized misalignment of the SKF-BF and OLMS-BF algorithms using WGNs as input signals. Both algorithms use the recursive estimate $\hat{\sigma}_{\bar{w}_h}^2(n)$ and $\sigma_{\bar{w}_g}^2 = 0$. The length of the global impulse responses is $ML = 512$.

Figure 10. Normalized misalignment of the SKF-BF and OLMS-BF algorithms using AR(1) processes as input signals. Both algorithms use the recursive estimate $\hat{\sigma}_{\bar{w}_h}^2(n)$ and $\sigma_{\bar{w}_g}^2 = 0$. The length of the global impulse responses is $ML = 512$.

In the second set of experiments, the behavior of the OLMS-BF algorithm is analyzed, as compared to the NLMS-BF algorithm [12]. The NLMS-BF algorithm uses different values of its step-size parameters, $\alpha_{\hat{h}}$ and $\alpha_{\hat{g}}$. The performances are now evaluated in terms of both NPMs and NM, using both types of input signals as before (WGNs and AR(1) processes). The results are presented in Figures 11 and 12, using WGNs as inputs, and in Figures 13 and 14, where the input signals are AR(1) processes. It can be noticed that the proposed solution achieves similar convergence rate but a much lower misalignment level than the NLMS-BF algorithm with $\alpha_{\hat{h}} = \alpha_{\hat{g}} = 0.5$ (which provides the fastest convergence rate [12]). On the other hand, if we target a lower misalignment and set the step-sizes of the NLMS-BF to smaller values (i.e., $\alpha_{\hat{h}} = \alpha_{\hat{g}} = 0.1$ and $\alpha_{\hat{h}} = \alpha_{\hat{g}} = 0.01$), the convergence rate also

decreases. However, the OLMS-BF algorithm leads to a misalignment level similar to the NLMS-BF algorithm using the smallest step-sizes. In addition, when the input signals are AR(1) processes, the improvement offered by the OLMS-BF algorithm is even more apparent.

Figure 11. Normalized projection misalignment of the OLMS-BF and NLMS-BF (using different step-size parameters): **(Top)** identification of the temporal impulse response $\mathbf{h}(n)$, **(Bottom)** identification of the spatial impulse response $\mathbf{g}(n)$. The input signals are WGNs, $L = 64$, and $M = 8$.

Figure 12. Normalized misalignment of the OLMS-BF and NLMS-BF (using different step-size parameters). The input signals are WGNs and $ML = 512$.

Figure 13. Normalized projection misalignment of the OLMS-BF and NLMS-BF (using different step-size parameters): **(Top)** identification of the temporal impulse response $\mathbf{h}(n)$, **(Bottom)** identification of the spatial impulse response $\mathbf{g}(n)$. The input signals are AR(1) processes, $L = 64$, and $M = 8$.

Figure 14. Normalized misalignment of the OLMS-BF and NLMS-BF (using different step-size parameters). The input signals are AR(1) processes and $ML = 512$.

Next, the performance of the OLMS-BF algorithm is evaluated along with the JO-NLMS algorithm [29], which is applied for the identification of the global impulse response of length $ML = 512$. The results are presented in Figures 15 and 16, using WGNs and AR(1) input signals, respectively. As specified in Section 6, the JO-NLMS algorithm is the regular counterpart of the OLMS-BF in a classical (one-dimensional) system identification scenario. We can see that the proposed solution (tailored for bilinear forms, i.e., exploiting the two-dimensional decomposition) offers both faster convergence and tracking, as well as a lower misalignment, as compared to the JO-NLMS algorithm. The performance improvement is even more important in case of AR(1) input signals.

Figure 15. Normalized misalignment of the OLMS-BF and regular JO-NLMS algorithms. The input signals are WGNs and $ML = 512$.

Figure 16. Normalized misalignment of the OLMS-BF and regular JO-NLMS algorithms. The input signals are AR(1) processes and $ML = 512$.

Finally, to validate our approach, we assess the performance of the OLMS-BF algorithm when applying it in a context which is closer to a real scenario. The temporal impulse response $\mathbf{h}(n)$ is a real-world echo path of length $L = 512$. The spatial impulse response $\mathbf{g}(n)$, of length $M = 4$, is generated using an exponential decay with the elements $g_m = 0.5^m$, $m = 1, \dots, M$. Both impulse responses are then normalized such that $\|\mathbf{h}(n)\| = \|\mathbf{g}(n)\| = 1$. The input signal is an AR(1) process and we compare the behaviors of the OLMS-BF and NLMS-BF algorithms. The performance is illustrated in Figures 17 and 18. We can notice that the proposed solution slightly outperforms the fastest convergence rate of NLMS-BF, given by $\alpha_{\hat{\mathbf{h}}} = \alpha_{\hat{\mathbf{g}}} = 0.5$, but at the same time offering a much lower value of the misalignment. If, however, we use the NLMS-BF algorithm with the smaller

step-sizes (in order to obtain a better misalignment), the resulting convergence rate is much lower than the one of the OLMS-BF algorithm.

Figure 17. Normalized projection misalignment of the OLMS-BF and NLMS-BF (using different step-size parameters): **(Top)** identification of the temporal impulse response $\mathbf{h}(n)$, **(Bottom)** identification of the spatial impulse response $\mathbf{g}(n)$. The input signals are AR(1) processes, $L = 512$, and $M = 4$.

Figure 18. Normalized misalignment of the OLMS-BF and NLMS-BF (using different step-size parameters). The input signals are AR(1) processes and $ML = 2048$.

8. Discussion

In this paper, we have focused on the Kalman filter tailored for the identification of bilinear forms (KF-BF), together with its simplified version (SKF-BF). Also, we have developed an optimized version

of the LMS algorithm for bilinear forms, namely OLMS-BF. In addition, a comparison between the SKF-BF and OLMS-BF algorithms has been outlined, indicating strong similarities between these two solutions. In our framework, the bilinear term has been defined with respect to the impulse responses of the spatiotemporal model.

The SKF-BF provides a reduced computational complexity as compared to KF-BF; the downside is that it has a slower convergence rate, more visible for correlated inputs. On the other hand, the SKF-BF and OLMS-BF algorithms perform very similarly. Experimental results also indicate that the algorithms tailored for bilinear forms outperform their regular counterparts (in such two-dimensional system identification scenarios), in terms of both convergence rate and tracking, as well as the steady-state misalignment. Adding to that, the reduced computational amount provided by the use of two shorter adaptive filters instead of a single (much longer) one, led us to conclude that the proposed algorithms could represent appealing solutions for the identification of bilinear forms.

Author Contributions: Conceptualization, L.-M.D.; Formal analysis, S.C.; Software, C.P.; Methodology, J.B.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

We detail here the evaluation of the expectation terms from (53)–(56), which are required in the development of (51) and (52). First, based on (43), we can express the error as

$$\begin{aligned} e(n) &= \mathbf{c}_g^T(n-1)\mathbf{x}_{\hat{\mathbf{h}}}(n) + \bar{\mathbf{w}}_g^T(n)\mathbf{x}_{\hat{\mathbf{h}}}(n) + v(n) + v_g(n) \\ &= \mathbf{c}_g^T(n-1)\mathbf{x}_{\hat{\mathbf{h}}}(n) + \bar{\mathbf{w}}_g^T(n)\mathbf{x}_{\hat{\mathbf{h}}}(n) + v(n) + \mathbf{c}_h^T(n-1)\mathbf{x}_g(n) + \bar{\mathbf{w}}_h^T(n)\mathbf{x}_g(n). \end{aligned} \quad (\text{A1})$$

Using this relation, together with (14) and (16), and taking into account that $\mathbf{c}_g(n-1)$ and $\mathbf{x}_{\hat{\mathbf{h}}}(n)$ are uncorrelated, the term from (53) results in

$$\begin{aligned} A_g &= \mathbb{E} \left\{ \left[\mathbf{c}_g^T(n-1) + \bar{\mathbf{w}}_g^T(n) \right] \mathbf{x}_{\hat{\mathbf{h}}}(n) e(n) \right\} \\ &= \mathbb{E} \left\{ \left[\mathbf{c}_g^T(n-1) + \bar{\mathbf{w}}_g^T(n) \right] \mathbf{x}_{\hat{\mathbf{h}}}(n) \left[v(n) + \bar{\mathbf{w}}_g^T(n)\mathbf{x}_{\hat{\mathbf{h}}}(n) + v_g(n) + \mathbf{c}_g^T(n-1)\mathbf{x}_{\hat{\mathbf{h}}}(n) \right] \right\} \\ &= p_g(n) + \mathbb{E} \left[\mathbf{c}_g^T(n-1)\mathbf{x}_{\hat{\mathbf{h}}}(n)\mathbf{x}_{\hat{\mathbf{h}}}^T(n)\mathbf{c}_g(n-1) + \bar{\mathbf{w}}_g^T(n)\mathbf{x}_{\hat{\mathbf{h}}}(n)\mathbf{x}_{\hat{\mathbf{h}}}^T(n)\bar{\mathbf{w}}_g(n) \right] \\ &= p_g(n) + \text{tr} \left\{ \mathbb{E} \left[\mathbf{c}_g(n-1)\mathbf{c}_g^T(n-1) + \bar{\mathbf{w}}_g(n)\bar{\mathbf{w}}_g^T(n) \right] \mathbb{E} \left[\mathbf{x}_{\hat{\mathbf{h}}}(n)\mathbf{x}_{\hat{\mathbf{h}}}^T(n) \right] \right\} \\ &= p_g(n) + \left[m_g(n-1) + M\sigma_{\bar{w}_g}^2 \right] \mathbb{E} \left[\mathbf{x}_{\hat{\mathbf{h}}}(n)\mathbf{x}_{\hat{\mathbf{h}}}^T(n) \right], \end{aligned} \quad (\text{A2})$$

where $p_g(n)$ is given in (61) and we took into account that $\mathbf{c}_g(n-1)$ and $\mathbf{x}_{\hat{\mathbf{h}}}(n)$ are uncorrelated.

Next, we should concentrate on the last expectation term in (A2), which can be expressed as

$$\mathbb{E} \left[\mathbf{x}_{\hat{\mathbf{h}}}(n)\mathbf{x}_{\hat{\mathbf{h}}}^T(n) \right] = \mathbb{E} \left\{ \left[\mathbf{I}_M \otimes \hat{\mathbf{h}}(n-1) \right]^T \tilde{\mathbf{x}}(n)\tilde{\mathbf{x}}^T(n) \left[\mathbf{I}_M \otimes \hat{\mathbf{h}}(n-1) \right] \right\}. \quad (\text{A3})$$

The main diagonal terms of this matrix are $\mathbb{E} \left[\hat{\mathbf{h}}^T(n-1)\mathbf{x}_m(n)\mathbf{x}_m^T(n)\hat{\mathbf{h}}(n-1) \right]$, $m = 1, 2, \dots, M$. In the following, we consider the assumption that the input signals are independent and have the same power, while their covariance matrices are close to a diagonal matrix [27,28]. Consequently,

$$\begin{aligned} \mathbb{E} \left[\hat{\mathbf{h}}^T(n-1)\mathbf{x}_m(n)\mathbf{x}_m^T(n)\hat{\mathbf{h}}(n-1) \right] &= \text{tr} \left\{ \mathbb{E} \left[\mathbf{x}_m(n)\mathbf{x}_m^T(n)\hat{\mathbf{h}}(n-1)\hat{\mathbf{h}}^T(n-1) \right] \right\} \\ &= \text{tr} \left\{ \mathbb{E} \left[\mathbf{x}_m(n)\mathbf{x}_m^T(n) \right] \mathbb{E} \left[\hat{\mathbf{h}}(n-1)\hat{\mathbf{h}}^T(n-1) \right] \right\} \\ &= \sigma_x^2 \mathbb{E} \left[\left\| \hat{\mathbf{h}}(n-1) \right\|^2 \right]. \end{aligned} \quad (\text{A4})$$

Finally, using (A3) and (A4) in (A2), we obtain

$$A_{\mathbf{g}} = p_{\mathbf{g}}(n) + \sigma_{\mathbf{x}}^2 \mathbb{E} \left[\left\| \hat{\mathbf{h}}(n-1) \right\|^2 \right] \left[m_{\mathbf{g}}(n-1) + M\sigma_{\bar{w}_{\mathbf{g}}}^2 \right]. \quad (\text{A5})$$

In a similar manner, the corresponding term from (55) is derived as

$$A_{\mathbf{h}} = p_{\mathbf{h}}(n) + \sigma_{\mathbf{x}}^2 \mathbb{E} \left[\left\| \hat{\mathbf{g}}(n-1) \right\|^2 \right] \left[m_{\mathbf{h}}(n-1) + L\sigma_{\bar{w}_{\mathbf{h}}}^2 \right]. \quad (\text{A6})$$

The terms $p_{\mathbf{g}}(n)$ and $p_{\mathbf{h}}(n)$ are given in (61) and (62), respectively.

Further, we detail the evaluation of the expectation term from (54). To begin, let us focus on the product $\mathbf{x}_{\hat{\mathbf{h}}}^T(n)\mathbf{x}_{\hat{\mathbf{h}}}(n)$. Relying on the same considerations and assumptions from (A3) and (A4), we obtain

$$\begin{aligned} \mathbf{x}_{\hat{\mathbf{h}}}^T(n)\mathbf{x}_{\hat{\mathbf{h}}}(n) &= \tilde{\mathbf{x}}^T(n) \left[\mathbf{I}_M \otimes \hat{\mathbf{h}}(n-1) \right] \left[\mathbf{I}_M \otimes \hat{\mathbf{h}}(n-1) \right]^T \tilde{\mathbf{x}}(n) \\ &= \text{tr} \left\{ \left[\mathbf{I}_M \otimes \hat{\mathbf{h}}(n-1) \right]^T \tilde{\mathbf{x}}(n)\tilde{\mathbf{x}}^T(n) \left[\mathbf{I}_M \otimes \hat{\mathbf{h}}(n-1) \right] \right\} \\ &\approx M\sigma_{\mathbf{x}}^2 \left\| \hat{\mathbf{h}}(n-1) \right\|^2. \end{aligned} \quad (\text{A7})$$

Similarly,

$$\mathbf{x}_{\hat{\mathbf{g}}}^T(n)\mathbf{x}_{\hat{\mathbf{g}}}(n) \approx L\sigma_{\mathbf{x}}^2 \left\| \hat{\mathbf{g}}(n-1) \right\|^2. \quad (\text{A8})$$

Hence, considering some degree of stationarity of the input signals, (A7) can be seen as a deterministic quantity, yielding

$$B_{\mathbf{g}} = \mathbb{E} \left[\mathbf{x}_{\hat{\mathbf{h}}}^T(n)\mathbf{x}_{\hat{\mathbf{h}}}(n)e^2(n) \right] \approx M\sigma_{\mathbf{x}}^2 \mathbb{E} \left[\left\| \hat{\mathbf{h}}(n-1) \right\|^2 \right] \mathbb{E} \left[e^2(n) \right]. \quad (\text{A9})$$

Let us focus on the computation of the expectation term $\mathbb{E} \left[e^2(n) \right]$. Using (A1), we obtain

$$\begin{aligned} \mathbb{E} \left[e^2(n) \right] &= \sigma_v^2 + \mathbb{E} \left[\bar{\mathbf{w}}_{\mathbf{g}}^T(n)\mathbf{x}_{\hat{\mathbf{h}}}(n)\mathbf{x}_{\hat{\mathbf{h}}}^T(n)\bar{\mathbf{w}}_{\mathbf{g}}(n) \right] \\ &\quad + \mathbb{E} \left\{ \left[v_{\mathbf{g}}(n) + \mathbf{x}_{\hat{\mathbf{h}}}^T(n)\mathbf{c}_{\mathbf{g}}(n-1) \right] \left[v_{\mathbf{g}}(n) + \mathbf{c}_{\mathbf{g}}^T(n-1)\mathbf{x}_{\hat{\mathbf{h}}}(n) \right] \right\}. \end{aligned} \quad (\text{A10})$$

At this point, we need to evaluate the variance of $v_{\mathbf{g}}$, which can be developed as

$$\begin{aligned} \sigma_{v_{\mathbf{g}}}^2(n) &= \mathbb{E} \left[v_{\mathbf{g}}^2(n) \right] \\ &= \mathbb{E} \left\{ \left[\mathbf{c}_{\mathbf{h}}^T(n-1)\mathbf{x}_{\mathbf{g}}(n) + \bar{\mathbf{w}}_{\mathbf{h}}^T(n)\mathbf{x}_{\mathbf{g}}(n) \right] \left[\mathbf{x}_{\mathbf{g}}^T(n)\mathbf{c}_{\mathbf{h}}(n-1) + \mathbf{x}_{\mathbf{g}}^T(n)\bar{\mathbf{w}}_{\mathbf{h}}(n) \right] \right\} \\ &= \mathbb{E} \left[\mathbf{c}_{\mathbf{h}}^T(n-1)\mathbf{x}_{\mathbf{g}}(n)\mathbf{x}_{\mathbf{g}}^T(n)\mathbf{c}_{\mathbf{h}}(n-1) + \bar{\mathbf{w}}_{\mathbf{h}}^T(n)\mathbf{x}_{\mathbf{g}}(n)\mathbf{x}_{\mathbf{g}}^T(n)\bar{\mathbf{w}}_{\mathbf{h}}(n) \right] \\ &= \text{tr} \left\{ \mathbb{E} \left[\mathbf{c}_{\mathbf{h}}(n-1)\mathbf{c}_{\mathbf{h}}^T(n-1) \right] \mathbb{E} \left[\mathbf{x}_{\mathbf{g}}(n)\mathbf{x}_{\mathbf{g}}^T(n) \right] + \sigma_{\bar{w}_{\mathbf{h}}}^2 \mathbb{E} \left[\mathbf{x}_{\mathbf{g}}(n)\mathbf{x}_{\mathbf{g}}^T(n) \right] \right\} \\ &= \mathbb{E} \left[\left\| \mathbf{x}_{\mathbf{g}}(n) \right\|^2 \right] \left\{ L\sigma_{\bar{w}_{\mathbf{h}}}^2 + \mathbb{E} \left[\left\| \mathbf{c}_{\mathbf{h}}(n-1) \right\|^2 \right] \right\} \\ &= L\sigma_{\mathbf{x}}^2 \mathbb{E} \left[\left\| \mathbf{g}(n) \right\|^2 \right] \left[L\sigma_{\bar{w}_{\mathbf{h}}}^2 + m_{\mathbf{h}}(n-1) \right], \end{aligned} \quad (\text{A11})$$

where $\sigma_{\mathbf{x}}^2 = \mathbb{E} \left[\left\| \tilde{\mathbf{x}}(n) \right\|^2 \right]$. Therefore, (A10) results in

$$\mathbb{E} \left[e^2(n) \right] = \sigma_v^2 + \sigma_{\mathbf{x}}^2 \left[M\sigma_{\bar{w}_{\mathbf{g}}}^2 + m_{\mathbf{g}}(n-1) \right] \mathbb{E} \left[\left\| \hat{\mathbf{h}}(n-1) \right\|^2 \right] + \frac{1}{L}\sigma_{v_{\mathbf{g}}}^2(n) + 2p_{\mathbf{g}}(n), \quad (\text{A12})$$

thus obtaining

$$B_{\mathbf{g}} = M\sigma_{\mathbf{x}}^2 \mathbb{E} \left[\|\hat{\mathbf{h}}(n-1)\|^2 \right] \left\{ \sigma_v^2 + 2p_{\mathbf{g}}(n) + \frac{1}{L}\sigma_{v_{\mathbf{g}}}^2(n) + \sigma_{\mathbf{x}}^2 \mathbb{E} \left[\|\hat{\mathbf{h}}(n-1)\|^2 \right] \left[m_{\mathbf{g}}(n-1) + M\sigma_{\bar{w}_{\mathbf{g}}}^2 \right] \right\}. \quad (\text{A13})$$

For the corresponding term from (56), we use the dual expression for $e(n)$ (see (14)), which leads to

$$\mathbb{E} [e^2(n)] = \sigma_v^2 + \sigma_{\mathbf{x}}^2 \left[L\sigma_{\bar{w}_{\mathbf{h}}}^2 + m_{\mathbf{h}}(n-1) \right] \mathbb{E} [\|\hat{\mathbf{g}}(n-1)\|^2] + \frac{1}{M}\sigma_{v_{\mathbf{h}}}^2(n) + 2p_{\mathbf{h}}(n), \quad (\text{A14})$$

where (similar to (A11))

$$\begin{aligned} \sigma_{v_{\mathbf{h}}}^2(n) &= \mathbb{E} [v_{\mathbf{h}}^2(n)] \\ &= M\sigma_{\mathbf{x}}^2 \mathbb{E} [\|\mathbf{h}(n)\|^2] \left[M\sigma_{\bar{w}_{\mathbf{g}}}^2 + m_{\mathbf{g}}(n-1) \right]. \end{aligned} \quad (\text{A15})$$

Thus, we finally obtain

$$B_{\mathbf{h}} = L\sigma_{\mathbf{x}}^2 \mathbb{E} [\|\hat{\mathbf{g}}(n-1)\|^2] \left\{ \sigma_v^2 + 2p_{\mathbf{h}}(n) + \frac{1}{M}\sigma_{v_{\mathbf{h}}}^2(n) + \sigma_{\mathbf{x}}^2 \mathbb{E} [\|\hat{\mathbf{g}}(n-1)\|^2] \left[m_{\mathbf{h}}(n-1) + L\sigma_{\bar{w}_{\mathbf{h}}}^2 \right] \right\}. \quad (\text{A16})$$

Summarizing, we can use (A5), (A6), (A13), and (A16) in (51)–(52), in order to obtain the recursive relations from (63)–(64), which are further used in the development of the OLMS-BF algorithm.

Appendix B

We provide here (summarized in Table A1) the main parameters that were used throughout the paper, in order to facilitate the reading.

Table A1. The main parameters used throughout the paper.

Parameters	Significance
$\mathbf{X}(n)$	zero-mean multiple-input signal matrix
$\tilde{\mathbf{x}}(n) = \text{vec}[\mathbf{X}(n)]$	input signal vector
$y(n)$	output signal at time n (i.e., the bilinear form)
$d(n)$	zero-mean desired signal at time n
$v(n)$	zero-mean additive noise, of variance σ_v^2
$\mathbf{h}(n), \mathbf{g}(n)$	true impulse responses of the system, of lengths L and M , respectively
η	arbitrary scaling parameter
$\mathbf{f}(n)$	true spatiotemporal (i.e., global) impulse response of the system, of length ML
$\hat{\mathbf{h}}(n), \hat{\mathbf{g}}(n)$	estimated impulse responses of the system
$\hat{\mathbf{f}}(n) = \hat{\mathbf{g}}(n) \otimes \hat{\mathbf{h}}(n)$	estimated spatiotemporal impulse response of the system
$\mathbf{x}_{\mathbf{h}}(n), \mathbf{x}_{\mathbf{g}}(n)$	the input signals from Figures 1 and 2
$\hat{y}(n)$	estimated output signal at time n
$e(n)$	error signal between the desired and estimated signals at time n
$\bar{\mathbf{w}}_{\mathbf{h}}(n), \bar{\mathbf{w}}_{\mathbf{g}}(n)$	zero-mean white Gaussian noise vectors, with correlation matrices $\mathbf{R}_{\bar{\mathbf{w}}_{\mathbf{h}}}(n) = \sigma_{\bar{w}_{\mathbf{h}}}^2 \mathbf{I}_L$ and $\mathbf{R}_{\bar{\mathbf{w}}_{\mathbf{g}}}(n) = \sigma_{\bar{w}_{\mathbf{g}}}^2 \mathbf{I}_M$, respectively
$\mathbf{k}_{\mathbf{h}}(n), \mathbf{k}_{\mathbf{g}}(n)$	Kalman gain vectors
$\mathbf{c}_{\mathbf{h}}(n), \mathbf{c}_{\mathbf{g}}(n)$	a posteriori misalignments corresponding to the two impulse responses, with correlation matrices $\mathbf{R}_{\mathbf{c}_{\mathbf{h}}}(n)$ and $\mathbf{R}_{\mathbf{c}_{\mathbf{g}}}(n)$, respectively
$\mathbf{c}_{\mathbf{h}_a}(n), \mathbf{c}_{\mathbf{g}_a}(n)$	a priori misalignments corresponding to the two impulse responses, with correlation matrices $\mathbf{R}_{\mathbf{c}_{\mathbf{h}_a}}(n)$ and $\mathbf{R}_{\mathbf{c}_{\mathbf{g}_a}}(n)$, respectively
$m_{\mathbf{g}}(n) = \mathbb{E} [\ \mathbf{c}_{\mathbf{g}}(n)\ ^2],$ $m_{\mathbf{h}}(n) = \mathbb{E} [\ \mathbf{c}_{\mathbf{h}}(n)\ ^2]$	squared norms of the a posteriori misalignments
$\mu_{\hat{\mathbf{h}},o}, \mu_{\hat{\mathbf{g}},o}$	optimal step-size parameters corresponding to the two adaptive filters

References

1. Mohler, R.R.; Kolodziej, W.J. An overview of bilinear system theory and applications. *IEEE Trans. Syst. Man Cybern.* **1980**, *10*, 683–688.
2. Baik, H.K.; Mathews, V.J. Adaptive lattice bilinear filters. *IEEE Trans. Signal Process.* **1993**, *41*, 2033–2046. [\[CrossRef\]](#)
3. Han, S.; Kim, J.; Sung, K. Extended generalized total least squares method for the identification of bilinear systems. *IEEE Trans. Signal Process.* **1996**, *44*, 1015–1018.
4. Tsoulkas, V.; Koukoulas, P.; Kalouptsidis, N. Identification of input-output bilinear systems using cumulants. *IEEE Trans. Signal Process.* **2001**, *51*, 2753–2761. [\[CrossRef\]](#)
5. Dos Santos, P.L.; Ramos, J.A.; de Carvalho, J.L.M. Identification of bilinear systems with white noise inputs: an iterative deterministic-stochastic subspace approach. *IEEE Trans. Control Syst. Technol.* **2009**, *17*, 1145–1153. [\[CrossRef\]](#)
6. Forssén, U. Adaptive bilinear digital filters. *IEEE Trans. Circuits Syst. II Analog Digital Signal Process.* **1993**, *40*, 729–735. [\[CrossRef\]](#)
7. Hu, R.; Hassan, H.M. Echo cancellation in high speed data transmission systems using adaptive layered bilinear filters. *IEEE Trans. Commun.* **1994**, *42*, 655–663.
8. Zhu, Z.; Leung, H. Adaptive identification of nonlinear systems with application to chaotic communications. *IEEE Trans. Circuits Syst. I Fundam. Theory Appl.* **2000**, *47*, 1072–1080. [\[CrossRef\]](#)
9. Kuo, S.M.; Wu, H.-T. Nonlinear adaptive bilinear filters for active noise control systems. *IEEE Trans. Circuits Syst. I Regul. Pap.* **2005**, *52*, 617–624. [\[CrossRef\]](#)
10. Zhao, H.; Zeng, X.; He, Z. Low-complexity nonlinear adaptive filter based on a pipelined bilinear recurrent neural network. *IEEE Trans. Neural Netw.* **2011**, *22*, 1494–1507. [\[CrossRef\]](#)
11. Benesty, J.; Paleologu, C.; Ciochină, S. On the identification of bilinear forms with the Wiener filter. *IEEE Signal Process. Lett.* **2017**, *24*, 653–657. [\[CrossRef\]](#)
12. Paleologu, C.; Benesty, J.; Ciochină, S. Adaptive filtering for the identification of bilinear forms. *Digital Signal Process.* **2018**, *75*, 153–167. [\[CrossRef\]](#)
13. Dogariu, L.; Paleologu, C.; Ciochină, S.; Benesty, J.; Piantanida, P. Identification of bilinear forms with the Kalman filter. In Proceedings of the 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, AB, Canada, 15–20 April 2018; pp. 4134–4138.
14. Elisei-Iliescu, C.; Stanciu, C.; Paleologu, C.; Benesty, J.; Anghel, C.; Ciochină, S. Efficient recursive least-squares algorithms for the identification of bilinear forms. *Digital Signal Process.* **2018**, *83*, 280–296. [\[CrossRef\]](#)
15. Gesbert, D.; Duhamel, P. Robust blind joint data/channel estimation based on bilinear optimization. In Proceedings of the 8th Workshop on Statistical Signal and Array Processing, Corfu, Greece, 24–26 June 1996; pp. 168–171.
16. Stenger, A.; Kellermann, W.; Rabenstein, R. Adaptation of acoustic echo cancellers incorporating a memoryless nonlinearity. In Proceedings of the IEEE Workshop on Acoustic Echo and Noise Control (IWAENC'99), Pocono Manor, PA, USA, 27–29 September 1999; 4p.
17. Stenger, A.; Kellermann, W. Adaptation of a memoryless preprocessor for nonlinear acoustic echo cancelling. *Signal Process.* **2000**, *80*, 1747–1760. [\[CrossRef\]](#)
18. Huang, Y.; Skoglund, J.; Luebs, A. Practically efficient nonlinear acoustic echo cancellers using cascaded block RLS and FLMS adaptive filters. In Proceedings of the 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), New Orleans, LA, USA, 5–9 March 2017; pp. 596–600.
19. Kalman, R.E. A new approach to linear filtering and prediction problems. *J. Basic Eng.* **1960**, *82*, 35–45. [\[CrossRef\]](#)
20. Zarchan, P.; Musoff, H. *Fundamentals of Kalman Filtering: A Practical Approach*, 4th ed.; American Institute of Aeronautics and Astronautics, Incorporated: Reston, VA, USA, 2015.
21. Hsieh, C.-S. Robust two-stage Kalman filters for systems with unknown inputs. *IEEE Trans. Autom. Control* **2000**, *45*, 2374–2378. [\[CrossRef\]](#)
22. Chen, L.; Mercorelli, P.; Liu, S. A Kalman Estimator for Detecting Repetitive Disturbances. In Proceedings of the 2005, American Control Conference, Portland, OR, USA, 8–10 June 2005; pp. 1631–1636.

23. Mercorelli, P. A Motion-Sensorless Control for Intake Valves in Combustion Engines. *IEEE Trans. Ind. Electron.* **2016**, *64*, 3402–3412. [[CrossRef](#)]
24. Morgan, D.R.; Benesty, J.; Sondhi, M.M. On the evaluation of estimated impulse responses. *IEEE Signal Process. Lett.* **1998**, *5*, 174–176. [[CrossRef](#)]
25. Kay, S.M. *Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory*; Prentice Hall: Englewood Cliffs, NJ, USA, 1993.
26. Paleologu, C.; Benesty, J.; Ciochină, S. Study of the general Kalman filter for echo cancellation. *IEEE Trans. Audio Speech Lang. Process.* **2013**, *21*, 1539–1549. [[CrossRef](#)]
27. Haykin, S.; Widrow, B. (Eds.) *Least-Mean-Square Adaptive Filters*; Wiley: Hoboken, NJ, USA, 2003.
28. Sulyman, A.I.; Zerguine, A. Convergence and steady-state analysis of a variable step-size NLMS algorithm. *Signal Process.* **2003**, *83*, 1255–1273. [[CrossRef](#)]
29. Ciochină, S.; Paleologu, C.; Benesty, J. An optimized NLMS algorithm for system identification. *Signal Process.* **2016**, *118*, 115–121. [[CrossRef](#)]
30. Iqbal, M.A.; Grant, S.L. Novel variable step size NLMS algorithm for echo cancellation. In Proceedings of the 2008 IEEE International Conference on Acoustics, Speech and Signal Processing, Las Vegas, NV, USA, 31 March–4 April 2008; pp. 241–244.
31. Paleologu, C.; Ciochină, S.; Benesty, J. Double-talk robust VSS-NLMS algorithm for under-modeling acoustic echo cancellation. In Proceedings of the 2008 IEEE International Conference on Acoustics, Speech and Signal Processing, Las Vegas, NV, USA, 31 March–4 April 2008; pp. 245–248.

© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).