IDENTIFICATION OF BILINEAR FORMS WITH THE KALMAN FILTER

Laura Dogariu[⋆] *Constantin Paleologu*[⋆] *Silviu Ciochină*[⋆] *Jacob Benesty*[†] *Pablo Piantanida*[‡]

[⋆] University Politehnica of Bucharest, Romania; E-mail: ldogariu@comm.pub.ro
[†] INRS-EMT, University of Quebec, Montreal, Canada
[‡] L2S, CentraleSupélec - CNRS - Université Paris Sud, F-91192 Gif-sur-Yvette, France

ABSTRACT

In this paper, we develop the Kalman filter for the identification of bilinear forms. In this framework, the bilinear term is defined with respect to the impulse responses of a spatiotemporal model, which resembles a multiple-input/single-output system. Recently, the identification of such bilinear forms was addressed in terms of the Wiener filter and conventional adaptive algorithms, i.e., least-mean-square and recursive least-squares. In this work, apart from the derivation of the Kalman filter tailored for the identification of bilinear forms, a simplified (i.e., low complexity) version of the algorithm is also presented. Simulation results support the theoretical findings and indicate the good performance of the proposed solutions.

Index Terms— Adaptive filters, bilinear forms, Kalman filter, multiple-input/single-output (MISO) system, system identification.

1. INTRODUCTION

The bilinear systems have been previously studied in various contexts [1], among which the approximation of nonlinear systems. There are many applications deriving from here, e.g., system identification [2]– [5], digital filter design [6], echo cancellation [7], chaotic communications [8], active noise control [9], neural networks [10], etc. In these works, the bilinear term is defined in terms of an input-output relation (i.e., with respect to the data).

Recently, a different approach was presented in [11] by defining the bilinear term in the context of a multiple-input/single-output (MISO) system, with respect to the impulse responses of a spatiotemporal model. The identification of such bilinear forms with the Wiener filter has been addressed in [11], followed by adaptive solutions based on the least-mean-square (LMS), normalized LMS (NLMS), and recursive least-squares (RLS) algorithms [12]. Similar frameworks can be found in [13]– [17], in the context of different applications (most of them were not associated with bilinear forms).

In this paper, we focus on deriving a different type of algorithm, based on the Kalman filter [18], which will be tailored for the identification of such bilinear forms. Also, a simplified (i.e., low complexity) version of this algorithm is developed, which could be more suitable in practice.

The rest of this paper is organized as follows. Section 2 introduces the system model in the context of bilinear forms. In this framework, the Kalman filter is developed in Section 3, while the simplified version of the algorithm is derived in Section 4. Some practical considerations are provided in Section 5. Simulations performed in the context of system identification are presented in Section 6. Finally, Section 7 concludes this work.

This work was supported under the Grants SeaForest 86/2016 and SenSyStar 190/2017. P. Piantanida would like to acknowledge support for this work from the LIA of the CNRS on *Information, Learning and Control*.

2. SYSTEM MODEL

The signal model considered throughout the paper is given by [11]

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

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

We can rewrite the matrix $\mathbf{X}(t)$, of size $L \times M$, as a vector of length ML, by using the vectorization operation, i.e., $\text{vec}[\mathbf{X}(t)] = [\mathbf{x}_1^T(t) \quad \mathbf{x}_2^T(t) \quad \dots \quad \mathbf{x}_M^T(t)]^T = \widetilde{\mathbf{x}}(t)$. Thus, we can express the output signal as $y(t) = \mathbf{h}^T(t)\mathbf{X}(t)\mathbf{g}(t) = [\mathbf{g}(t) \otimes \mathbf{h}(t)]^T \widetilde{\mathbf{x}}(t) = \mathbf{f}^T(t)\widetilde{\mathbf{x}}(t)$, where $\otimes$ denotes the Kronecker product between the individual impulse responses and the vector $\mathbf{f}(t) = \mathbf{g}(t) \otimes \mathbf{h}(t)$, of length ML, represents the spatiotemporal (i.e., global) impulse response of the system. In this way, the signal model in (1) becomes

$$d(t) = \mathbf{f}^T(t)\widetilde{\mathbf{x}}(t) + v(t). \tag{2}$$

The difference with respect to the general case of a MISO system lies in the fact that in this bilinear context $\mathbf{f}(t)$ is formed with only $M + L$ different elements, despite that it is of length ML.

The purpose is to identify the two impulse responses $\mathbf{h}(t)$ and $\mathbf{g}(t)$, and, consequently, the spatiotemporal impulse response $\mathbf{f}(t)$. To this purpose, we can use two adaptive filters, $\widehat{\mathbf{h}}(t)$ and $\widehat{\mathbf{g}}(t)$, while the global impulse response can be evaluated as $\widehat{\mathbf{f}}(t) = \widehat{\mathbf{g}}(t) \otimes \widehat{\mathbf{h}}(t)$. Let $\eta \neq 0$ be a real-valued number. It is clear from (1) that $[\mathbf{h}(t)/\eta]^T \mathbf{X}(t) [\eta\mathbf{g}(t)] = \mathbf{h}^T(t)\mathbf{X}(t)\mathbf{g}(t) = y(t)$, so that the pair $\mathbf{h}(t)/\eta$ and $\eta\mathbf{g}(t)$ is equivalent to the pair $\mathbf{h}(t)$ and $\mathbf{g}(t)$ in the bilinear form. This implies that we can only identify $\widehat{\mathbf{h}}(t)$ and $\widehat{\mathbf{g}}(t)$ up to a scaling factor. A similar discussion can be found in [13] and [17] in the context of blind identification/equalization and nonlinear acoustic echo cancellation, respectively. However, since $\mathbf{f}(t) = \mathbf{g}(t) \otimes \mathbf{h}(t) = [\eta\mathbf{g}(t)] \otimes [\mathbf{h}(t)/\eta]$, 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, as defined in [19], and for the identification of the global filter $\mathbf{f}(t)$ we shall use the normalized misalignment, which is defined as $\|\mathbf{f}(t) - \widehat{\mathbf{f}}(t)\|^2/\|\mathbf{f}(t)\|^2$, where $\|\cdot\|$ denotes the Euclidean norm.

In [11], this system identification problem has been addressed in terms of the Wiener filter. Hence, it was considered that the impulse responses that have to be identified are time-invariant systems (which is a basic assumption in the context of the Wiener filter). However, in practice, these systems could be variable in time. Consequently, in this paper, we approach the system identification problem in terms of the Kalman filter. In this context, the signal model from (1) can be considered as the observation equation, while the system impulse responses can be modeled as state equations. Therefore, we consider that $\mathbf{h}(t)$ and $\mathbf{g}(t)$ are zero-mean random vectors, which follow a simplified first-order Markov model, i.e.,

$$\mathbf{h}(t) = \mathbf{h}(t-1) + \mathbf{w_h}(t), \tag{3}$$

$$\mathbf{g}(t) = \mathbf{g}(t-1) + \mathbf{w_g}(t), \tag{4}$$

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

3. KALMAN FILTER FOR BILINEAR FORMS

In the section, we address the previously described system identification problem based on the Kalman filter. Given the two adaptive filters $\widehat{\mathbf{h}}(t)$ and $\widehat{\mathbf{g}}(t)$, the estimated signal is given by $\widehat{y}(t) = \widehat{\mathbf{h}}^T(t-1)\mathbf{X}(t)\widehat{\mathbf{g}}(t-1)$. As a result, the a priori error signal between the desired and estimated signals can be defined as

$$\begin{aligned} e(t) &= d(t) - \widehat{y}(t) = d(t) - \widehat{\mathbf{h}}^T(t-1)\mathbf{X}(t)\widehat{\mathbf{g}}(t-1) \\ &= d(t) - \left[\widehat{\mathbf{g}}(t-1) \otimes \widehat{\mathbf{h}}(t-1)\right]^T \widetilde{\mathbf{x}}(t) = d(t) - \widehat{\mathbf{f}}^T(t-1)\widetilde{\mathbf{x}}(t) \\ &= d(t) - \widehat{\mathbf{h}}^T(t-1)\widetilde{\mathbf{x}}_{\widehat{\mathbf{g}}}(t) = d(t) - \widehat{\mathbf{g}}^T(t-1)\widetilde{\mathbf{x}}_{\widehat{\mathbf{h}}}(t), \end{aligned} \tag{5}$$

where $\widetilde{\mathbf{x}}_{\widehat{\mathbf{g}}}(t) = [\widehat{\mathbf{g}}(t-1) \otimes \mathbf{I}_L]^T\widetilde{\mathbf{x}}(t)$ and $\widetilde{\mathbf{x}}_{\widehat{\mathbf{h}}}(t) = [\mathbf{I}_M \otimes \widehat{\mathbf{h}}(t-1)]^T\widetilde{\mathbf{x}}(t)$.

In the context of the linear sequential Bayesian approach, the optimal estimates of the state vectors have the forms [20]:

$$\widehat{\mathbf{h}}(t) = \widehat{\mathbf{h}}(t-1) + \mathbf{k_h}(t)e(t), \tag{6}$$

$$\widehat{\mathbf{g}}(t) = \widehat{\mathbf{g}}(t-1) + \mathbf{k_g}(t)e(t), \tag{7}$$

where $\mathbf{k_h}(t)$ and $\mathbf{k_g}(t)$ are the Kalman gain vectors. Next, let us define the a posteriori misalignments (which represent the state estimation errors) related to the temporal and spatial impulse responses as $\boldsymbol{\mu}_\mathbf{h}(t) = \mathbf{h}(t)/\eta - \widehat{\mathbf{h}}(t)$ and $\boldsymbol{\mu}_\mathbf{g}(t) = \eta\mathbf{g}(t) - \widehat{\mathbf{g}}(t)$, for which their correlation matrices are $\mathbf{R}_{\boldsymbol{\mu}_\mathbf{h}}(t) = E[\boldsymbol{\mu}_\mathbf{h}(t)\boldsymbol{\mu}^T_\mathbf{h}(t)]$ and $\mathbf{R}_{\boldsymbol{\mu}_\mathbf{g}}(t) = E[\boldsymbol{\mu}_\mathbf{g}(t)\boldsymbol{\mu}^T_\mathbf{g}(t)]$, respectively. As mentioned in Section 2, we can only identify the impulse responses up to this arbitrary scaling factor η; however, the pair $\mathbf{h}(t)/\eta$ and $\eta\mathbf{g}(t)$ is equivalent to the pair $\mathbf{h}$ and $\mathbf{g}$ in the bilinear form. Also, we can define the a priori misalignments related to the two impulse responses:

$$\mathbf{m_h}(t) = \frac{1}{\eta}\mathbf{h}(t) - \widehat{\mathbf{h}}(t-1) = \boldsymbol{\mu}_\mathbf{h}(t-1) + \frac{1}{\eta}\mathbf{w_h}(t), \tag{8}$$

$$\mathbf{m_g}(t) = \eta\mathbf{g}(t) - \widehat{\mathbf{g}}(t-1) = \boldsymbol{\mu}_\mathbf{g}(t-1) + \eta\mathbf{w_g}(t), \tag{9}$$

whose correlation matrices are $\mathbf{R}_{\mathbf{m_h}}(t) = E\left[\mathbf{m_h}(t)\mathbf{m}^T_\mathbf{h}(t)\right]$ and $\mathbf{R}_{\mathbf{m_g}}(t) = E\left[\mathbf{m_g}(t)\mathbf{m}^T_\mathbf{g}(t)\right]$, respectively.

For the sake of simplicity of the coming development, let us multiply (3) and (4) by $1/\eta$ and η, respectively, and introduce the notation $\overline{\mathbf{w}}_\mathbf{h}(t) = \mathbf{w_h}(t)/\eta$ and $\overline{\mathbf{w}}_\mathbf{g}(t) = \eta\mathbf{w_g}(t)$; these also represent zero-mean white Gaussian noise vectors, with correlation matrices $\mathbf{R}_{\overline{\mathbf{w}}_\mathbf{h}}(t) = \sigma^2_{\overline{w}_\mathbf{h}}\mathbf{I}_L$ and $\mathbf{R}_{\overline{\mathbf{w}}_\mathbf{g}}(t) = \sigma^2_{\overline{w}_\mathbf{g}}\mathbf{I}_M$, respectively. Clearly, we have $\sigma^2_{\overline{w}_\mathbf{h}} = \sigma^2_{w_\mathbf{h}}/\eta^2$ and $\sigma^2_{\overline{w}_\mathbf{g}} = \eta^2\sigma^2_{w_\mathbf{g}}$. Consequently, using this notation and developing in (8) and (9), we get

$$\mathbf{R}_{\mathbf{m_h}}(t) = \mathbf{R}_{\boldsymbol{\mu}_\mathbf{h}}(t-1) + \sigma^2_{\overline{w}_\mathbf{h}}\mathbf{I}_L, \tag{10}$$

$$\mathbf{R}_{\mathbf{m_g}}(t) = \mathbf{R}_{\boldsymbol{\mu}_\mathbf{g}}(t-1) + \sigma^2_{\overline{w}_\mathbf{g}}\mathbf{I}_M. \tag{11}$$

The Kalman gain vectors are obtained by minimizing the criterions $J_\mathbf{h}(t) = (1/L)\mathrm{tr}\left[\mathbf{R}_{\boldsymbol{\mu}_\mathbf{h}}(t)\right]$ and $J_\mathbf{g}(t) = (1/M)\mathrm{tr}\left[\mathbf{R}_{\boldsymbol{\mu}_\mathbf{g}}(t)\right]$ with respect to $\mathbf{k_h}(t)$ and $\mathbf{k_g}(t)$, respectively, where $\mathrm{tr}[\cdot]$ denotes the trace of a square matrix. From these minimizations, we find that

$$\mathbf{k_h}(t) = \frac{\mathbf{R}_{\mathbf{m_h}}(t)\widetilde{\mathbf{x}}_{\widehat{\mathbf{g}}}(t)}{\widetilde{\mathbf{x}}^T_{\widehat{\mathbf{g}}}(t)\mathbf{R}_{\mathbf{m_h}}(t)\widetilde{\mathbf{x}}_{\widehat{\mathbf{g}}}(t) + \sigma^2_v}, \tag{12}$$

$$\mathbf{k_g}(t) = \frac{\mathbf{R}_{\mathbf{m_g}}(t)\widetilde{\mathbf{x}}_{\widehat{\mathbf{h}}}(t)}{\widetilde{\mathbf{x}}^T_{\widehat{\mathbf{h}}}(t)\mathbf{R}_{\mathbf{m_g}}(t)\widetilde{\mathbf{x}}_{\widehat{\mathbf{h}}}(t) + \sigma^2_v}, \tag{13}$$

and

$$\mathbf{R}_{\boldsymbol{\mu}_\mathbf{h}}(t) = \left[\mathbf{I}_L - \mathbf{k_h}(t)\widetilde{\mathbf{x}}^T_{\widehat{\mathbf{g}}}(t)\right]\mathbf{R}_{\mathbf{m_h}}(t), \tag{14}$$

$$\mathbf{R}_{\boldsymbol{\mu}_\mathbf{g}}(t) = \left[\mathbf{I}_M - \mathbf{k_g}(t)\widetilde{\mathbf{x}}^T_{\widehat{\mathbf{h}}}(t)\right]\mathbf{R}_{\mathbf{m_g}}(t). \tag{15}$$

Summarizing, the equations that define the Kalman filter for bilinear forms (namely KF-BF) are given in Table 1.

Table 1: Kalman filter for bilinear forms (KF-BF).

Initialization:
$\widehat{\mathbf{h}}(0) = [1 \quad 0 \quad \ldots \quad 0]^T, \ \widehat{\mathbf{g}}(0) = (1/M)[1 \quad 1 \quad \ldots \quad 1]^T$
$\mathbf{R}_{\boldsymbol{\mu}_\mathbf{h}}(0) = \epsilon\mathbf{I}_L, \ \mathbf{R}_{\boldsymbol{\mu}_\mathbf{g}}(0) = \epsilon\mathbf{I}_M, \ \epsilon =$ small positive constant
Parameters: $\sigma^2_{\overline{w}_\mathbf{h}}, \ \sigma^2_{\overline{w}_\mathbf{g}}, \ \sigma^2_v$ known or estimated
Algorithm:
$\mathbf{R}_{\mathbf{m_h}}(t) = \mathbf{R}_{\boldsymbol{\mu}_\mathbf{h}}(t-1) + \sigma^2_{\overline{w}_\mathbf{h}}\mathbf{I}_L$
$\mathbf{R}_{\mathbf{m_g}}(t) = \mathbf{R}_{\boldsymbol{\mu}_\mathbf{g}}(t-1) + \sigma^2_{\overline{w}_\mathbf{g}}\mathbf{I}_M$
$\mathbf{k_h}(t) = \mathbf{R}_{\mathbf{m_h}}(t)\widetilde{\mathbf{x}}_{\widehat{\mathbf{g}}}(t)\left[\widetilde{\mathbf{x}}^T_{\widehat{\mathbf{g}}}(t)\mathbf{R}_{\mathbf{m_h}}(t)\widetilde{\mathbf{x}}_{\widehat{\mathbf{g}}}(t) + \sigma^2_v\right]^{-1}$
$\mathbf{k_g}(t) = \mathbf{R}_{\mathbf{m_g}}(t)\widetilde{\mathbf{x}}_{\widehat{\mathbf{h}}}(t)\left[\widetilde{\mathbf{x}}^T_{\widehat{\mathbf{h}}}(t)\mathbf{R}_{\mathbf{m_g}}(t)\widetilde{\mathbf{x}}_{\widehat{\mathbf{h}}}(t) + \sigma^2_v\right]^{-1}$
$e(t) = d(t) - \widetilde{\mathbf{x}}^T_{\widehat{\mathbf{g}}}(t)\widehat{\mathbf{h}}(t-1) = d(t) - \widetilde{\mathbf{x}}^T_{\widehat{\mathbf{h}}}(t)\widehat{\mathbf{g}}(t-1)$
$\widehat{\mathbf{h}}(t) = \widehat{\mathbf{h}}(t-1) + \mathbf{k_h}(t)e(t)$
$\widehat{\mathbf{g}}(t) = \widehat{\mathbf{g}}(t-1) + \mathbf{k_g}(t)e(t)$
$\mathbf{R}_{\boldsymbol{\mu}_\mathbf{h}}(t) = \left[\mathbf{I}_L - \mathbf{k_h}(t)\widetilde{\mathbf{x}}^T_{\widehat{\mathbf{g}}}(t)\right]\mathbf{R}_{\mathbf{m_h}}(t)$
$\mathbf{R}_{\boldsymbol{\mu}_\mathbf{g}}(t) = \left[\mathbf{I}_M - \mathbf{k_g}(t)\widetilde{\mathbf{x}}^T_{\widehat{\mathbf{h}}}(t)\right]\mathbf{R}_{\mathbf{m_g}}(t)$

4. SIMPLIFIED KALMAN FILTER FOR BILINEAR FORMS

In order to reduce the computational complexity of the KF-BF, a simplified version of this algorithm is derived in this section. The idea of this simplified algorithm is inspired by the work developed in [21] and [22], in the context of echo cancellation. To begin, let us assume that the KF-BF has converged to its steady-state. In this case, $\mathbf{R}_{\mathbf{m}_\mathbf{h}}(t)$ and $\mathbf{R}_{\mathbf{m}_\mathbf{g}}(t)$ tend to become diagonal matrices with all the elements on the main diagonal equal to small positive numbers, $\sigma^2_{\mathbf{m}_\mathbf{h}}(t)$ and $\sigma^2_{\mathbf{m}_\mathbf{g}}(t)$, respectively. Consequently, we can use the approximations $\mathbf{R}_{\mathbf{m}_\mathbf{h}}(t) \approx \sigma^2_{\mathbf{m}_\mathbf{h}}(t)\mathbf{I}_L$ and $\mathbf{R}_{\mathbf{m}_\mathbf{g}}(t) \approx \sigma^2_{\mathbf{m}_\mathbf{g}}(t)\mathbf{I}_M$.

In this way, the Kalman gain vectors for both temporal and spatial impulse responses simplify to

$$\mathbf{k}_\mathbf{h}(t) = \frac{\widetilde{\mathbf{x}}_{\widehat{\mathbf{g}}}(t)}{\widetilde{\mathbf{x}}^T_{\widehat{\mathbf{g}}}(t)\widetilde{\mathbf{x}}_{\widehat{\mathbf{g}}}(t) + \delta_\mathbf{h}(t)}, \quad (16)$$

$$\mathbf{k}_\mathbf{g}(t) = \frac{\widetilde{\mathbf{x}}_{\widehat{\mathbf{h}}}(t)}{\widetilde{\mathbf{x}}^T_{\widehat{\mathbf{h}}}(t)\widetilde{\mathbf{x}}_{\widehat{\mathbf{h}}}(t) + \delta_\mathbf{g}(t)}, \quad (17)$$

where $\delta_\mathbf{h}(t) = \sigma^2_v/\sigma^2_{\mathbf{m}_\mathbf{h}}(t)$ and $\delta_\mathbf{g}(t) = \sigma^2_v/\sigma^2_{\mathbf{m}_\mathbf{g}}(t)$ can be seen as variable regularization parameters. Then, the Kalman vectors from (16) and (17) are used in the updates (6) and (7), respectively.

Next, another simplification can be performed, by assuming that the matrices that appear in the update of $\mathbf{R}_{\boldsymbol{\mu}_\mathbf{h}}(t)$ and $\mathbf{R}_{\boldsymbol{\mu}_\mathbf{g}}(t)$ can be approximated as

$$\mathbf{I}_L - \mathbf{k}_\mathbf{h}(t)\widetilde{\mathbf{x}}^T_{\widehat{\mathbf{g}}}(t) \approx \left[1 - \frac{1}{L}\mathbf{k}^T_\mathbf{h}(t)\widetilde{\mathbf{x}}_{\widehat{\mathbf{g}}}(t)\right]\mathbf{I}_L, \quad (18)$$

$$\mathbf{I}_M - \mathbf{k}_\mathbf{g}(t)\widetilde{\mathbf{x}}^T_{\widehat{\mathbf{h}}}(t) \approx \left[1 - \frac{1}{M}\mathbf{k}^T_\mathbf{g}(t)\widetilde{\mathbf{x}}_{\widehat{\mathbf{h}}}(t)\right]\mathbf{I}_M. \quad (19)$$

These approximations are based on the fact that, as the filters start to converge, the misalignments of the individual coefficients tend to become uncorrelated; consequently, the matrices $\mathbf{R}_{\boldsymbol{\mu}_\mathbf{h}}(t)$ and $\mathbf{R}_{\boldsymbol{\mu}_\mathbf{g}}(t)$ tend to become diagonal. Therefore, using the notation $\mathbf{R}_{\mathbf{m}_\mathbf{h}}(t) \approx \sigma^2_{\mathbf{m}_\mathbf{h}}(t)\mathbf{I}_L = r_{\mathbf{m}_\mathbf{h}}(t)\mathbf{I}_L$ and $\mathbf{R}_{\mathbf{m}_\mathbf{g}}(t) \approx \sigma^2_{\mathbf{m}_\mathbf{g}}(t)\mathbf{I}_M = r_{\mathbf{m}_\mathbf{g}}(t)\mathbf{I}_M$, together with $\mathbf{R}_{\boldsymbol{\mu}_\mathbf{h}}(t) \approx \sigma^2_{\boldsymbol{\mu}_\mathbf{h}}(t)\mathbf{I}_L = r_{\boldsymbol{\mu}_\mathbf{h}}(t)\mathbf{I}_L$ and $\mathbf{R}_{\boldsymbol{\mu}_\mathbf{g}}(t) \approx \sigma^2_{\boldsymbol{\mu}_\mathbf{g}}(t)\mathbf{I}_M = r_{\boldsymbol{\mu}_\mathbf{g}}(t)\mathbf{I}_M$, the simplified Kalman filter for bilinear forms (namely SKF-BF) is summarized in Table 2. In terms of complexity, this simplified version is much more advantageous as compared to the KF-BF, which involves matrix operations.

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

Initialization:
$\widehat{\mathbf{h}}(0) = [1 \quad 0 \quad \ldots \quad 0]^T$, $\widehat{\mathbf{g}}(0) = (1/M)[1 \quad 1 \quad \ldots \quad 1]^T$
$r_{\boldsymbol{\mu}_\mathbf{h}}(0) = r_{\boldsymbol{\mu}_\mathbf{g}}(0) = \epsilon$ (small positive constant)
Parameters: $\sigma^2_{\overline{w}_\mathbf{h}}$, $\sigma^2_{\overline{w}_\mathbf{g}}$, σ^2_v known or estimated
Algorithm:
$r_{\mathbf{m}_\mathbf{h}}(t) = r_{\boldsymbol{\mu}_\mathbf{h}}(t-1) + \sigma^2_{\overline{w}_\mathbf{h}}(t)$
$r_{\mathbf{m}_\mathbf{g}}(t) = r_{\boldsymbol{\mu}_\mathbf{g}}(t-1) + \sigma^2_{\overline{w}_\mathbf{g}}(t)$
$\delta_\mathbf{h}(t) = \sigma^2_v/r_{\mathbf{m}_\mathbf{h}}(t)$
$\delta_\mathbf{g}(t) = \sigma^2_v/r_{\mathbf{m}_\mathbf{g}}(t)$
$e(t) = d(t) - \widetilde{\mathbf{x}}^T_{\widehat{\mathbf{g}}}(t)\widehat{\mathbf{h}}(t-1) = d(t) - \widetilde{\mathbf{x}}^T_{\widehat{\mathbf{h}}}(t)\widehat{\mathbf{g}}(t-1)$
$\widehat{\mathbf{h}}(t) = \widehat{\mathbf{h}}(t-1) + \widetilde{\mathbf{x}}_{\widehat{\mathbf{g}}}(t)e(t)\left[\widetilde{\mathbf{x}}^T_{\widehat{\mathbf{g}}}(t)\widetilde{\mathbf{x}}_{\widehat{\mathbf{g}}}(t) + \delta_\mathbf{h}(t)\right]^{-1}$
$\widehat{\mathbf{g}}(t) = \widehat{\mathbf{g}}(t-1) + \widetilde{\mathbf{x}}_{\widehat{\mathbf{h}}}(t)e(t)\left[\widetilde{\mathbf{x}}^T_{\widehat{\mathbf{h}}}(t)\widetilde{\mathbf{x}}_{\widehat{\mathbf{h}}}(t) + \delta_\mathbf{g}(t)\right]^{-1}$
$r_{\boldsymbol{\mu}_\mathbf{h}}(t) = \left\{1 - \frac{1}{L}\widetilde{\mathbf{x}}^T_{\widehat{\mathbf{g}}}(t)\widetilde{\mathbf{x}}_{\widehat{\mathbf{g}}}(t)\left[\widetilde{\mathbf{x}}^T_{\widehat{\mathbf{g}}}(t)\widetilde{\mathbf{x}}_{\widehat{\mathbf{g}}}(t) + \delta_\mathbf{h}(t)\right]^{-1}\right\} r_{\mathbf{m}_\mathbf{h}}(t)$
$r_{\boldsymbol{\mu}_\mathbf{g}}(t) = \left\{1 - \frac{1}{M}\widetilde{\mathbf{x}}^T_{\widehat{\mathbf{h}}}(t)\widetilde{\mathbf{x}}_{\widehat{\mathbf{h}}}(t)\left[\widetilde{\mathbf{x}}^T_{\widehat{\mathbf{h}}}(t)\widetilde{\mathbf{x}}_{\widehat{\mathbf{h}}}(t) + \delta_\mathbf{g}(t)\right]^{-1}\right\} r_{\mathbf{m}_\mathbf{g}}(t)$

5. PRACTICAL CONSIDERATIONS

The previously developed KF-BF and SKF-BF are designed to identify the individual impulse responses of the bilinear form, while the global impulse response can be obtained based on the Kronecker product between them. Alternatively, we may use the regular Kalman filter to directly identify the spatiotemporal impulse response, based on the observation equation (2) and considering the state equation $\mathbf{f}(t) = \mathbf{f}(t-1) + \mathbf{w}(t)$, where $\mathbf{w}(t)$ is a zero-mean white Gaussian noise signal vector. The correlation matrix of $\mathbf{w}(t)$ is $\mathbf{R}_\mathbf{w}(t) = \sigma^2_w\mathbf{I}_{ML}$, where $\mathbf{I}_{ML}$ is the $ML \times ML$ identity matrix and the variance σ^2_w captures the uncertainties in $\mathbf{f}(t)$.

In this context, following the approach presented in [21], it is straightforward to derive the regular Kalman filter (KF) and the simplified KF (SKF), which are able to identify the global impulse response using a single adaptive filter $\widehat{\mathbf{f}}(t)$; for further details, please see Sections VI and VII in [21]. Nevertheless, we should note that the solution based on the regular KF and SKF involves an adaptive filter of length ML, while the new KF-BF and SKF-BF use two shorter filters of lengths L and M, respectively. Consequently, a faster converge rate/tracking is expected for the proposed algorithms as compared to the conventional approaches.

Next, several considerations should be made related to the specific parameters that need to be set within the algorithms. First, the noise power σ^2_v is required within the Kalman gain vectors. In practice, this parameter can be estimated in different ways; for example, some simple and efficient methods for this purpose can be found in [23] or [24]. We should note that different other estimators can be used for the noise power; the analysis of their influence on the algorithms' performance is beyond the scope of this paper.

Second, the parameters related to the uncertainties in the unknown systems should be set or estimated, i.e., $\sigma^2_{\overline{w}_\mathbf{h}}$ and $\sigma^2_{\overline{w}_\mathbf{g}}$. Small values of these parameters imply a good misalignment but a poor tracking, while large values (i.e., the uncertainties in the unknown systems are high) imply a good tracking but a high misalignment. In other words, there is always a compromise between good tracking and low misalignment. In practice, if some a priori information is available (about the systems we need to identify), it could be taken into account in order to set these parameters. For example, if the spatial impulse response is assumed to be time-invariant, we could set $\sigma^2_{\overline{w}_\mathbf{g}} = 0$, while tuning only the parameter related to the temporal impulse response. In order to evaluate this parameter, let us rewrite (3) as $\overline{\mathbf{w}}_\mathbf{h}(t) = [\mathbf{h}(t) - \mathbf{h}(t-1)]/\eta$. Next, using the ℓ_2 norm in both sides of this equation, together with the approximation $\|\overline{\mathbf{w}}_\mathbf{h}(t)\|^2_2 \approx L\sigma^2_{\overline{w}_\mathbf{h}}$ (which is valid when $L \gg 1$), and replacing $\mathbf{h}(t)/\eta$ by its estimate $\widehat{\mathbf{h}}(t)$, we can evaluate

$$\widehat{\sigma}^2_{\overline{w}_\mathbf{h}}(t) = \frac{1}{L}\left\|\widehat{\mathbf{h}}(t) - \widehat{\mathbf{h}}(t-1)\right\|^2_2. \quad (20)$$

As we can notice, the estimation from (20) is designed to achieve a proper compromise between good tracking and low misalignment.

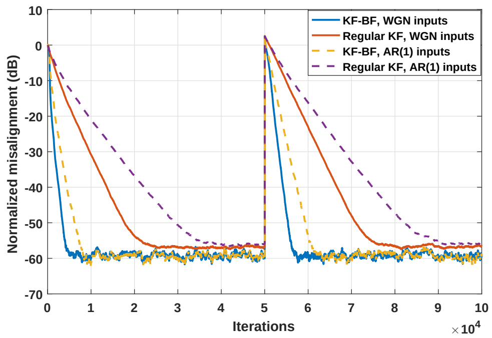

Fig. 1: Normalized misalignment of the KF-BF and regular KF for different types of input signals. The length of the global impulse response is $ML = 512$. The other parameters are set to $\sigma_v^2 = 0.01$, $\sigma_{w_\mathbf{h}}^2 = \sigma_{w_\mathbf{g}}^2 = \sigma_w^2 = 10^{-9}$, and $\epsilon = 10^{-5}$.

When the algorithm starts to converge or when there is an abrupt change of the system, the difference between $\widehat{\mathbf{h}}(t)$ and $\widehat{\mathbf{h}}(t-1)$ is significant, so that the parameter $\widehat{\sigma}_{w_\mathbf{h}}^2(t)$ takes large values, thus providing fast convergence and tracking. On the other hand, when the algorithm starts to converge to its steady-state, the difference between $\widehat{\mathbf{h}}(t)$ and $\widehat{\mathbf{h}}(t-1)$ reduces, thus leading to small values of $\widehat{\sigma}_{w_\mathbf{h}}^2(t)$ and, consequently, to a low misalignment.

6. SIMULATION RESULTS

Simulations are performed in the context of system identification, in order to outline the performance of the proposed KF-BF and SKF-BF algorithms, as compared to their regular counterparts, i.e., KF and SKF (introduced in the beginning of Section 5). In our experiments, the temporal impulse response is randomly generated (with Gaussian distribution) and its length is set to $L = 64$. The coefficients of the spatial impulse response (of length $M = 8$) are also randomly generated. Consequently, the length of the spatiotemporal (i.e., global) impulse response is $ML = 8 \times 64 = 512$. In order to evaluate the tracking capabilities of the algorithms, an abrupt change of the temporal impulse response is simulated in the middle of all the experiments (by generating a new random vector of length $L = 64$, with Gaussian distribution). The input signals $x_m(t)$, $m = 1, 2, \ldots, M$ are either white Gaussian noises (WGNs) or AR(1) processes [each one of them is generated by filtering a white Gaussian noise through a first-order system $1/\left(1 - 0.8z^{-1}\right)$]. The additive noise $v(t)$ is white and Gaussian, with the variance $\sigma_v^2 = 0.01$; we assume that this parameter is available in all the simulations. The performance measure is the normalized misalignment (in dB), to evaluate the identification of the global impulse response.

In Fig. 1, the proposed KF-BF is compared to the regular KF. The parameters of the algorithms are set to $\sigma_{w_\mathbf{h}}^2 = \sigma_{w_\mathbf{g}}^2 = \sigma_w^2 = 10^{-9}$. Also, the initialization constant is set to $\epsilon = 10^{-5}$ for all the algorithms. It can be noticed from the figure that the proposed KF-BF achieves a faster convergence rate as compared to the regular KF, for both types of input signals, having also a better tracking capability. The gain is more apparent in case of the AR(1) inputs.

The previous experiment is repeated in Fig. 2, comparing the SKF-BF with its regular counterpart SKF. As we can notice, the SKF-BF and SKF have a slower convergence rate [especially in case of AR(1) inputs] as compared to the KF-BF and KF, respectively; however, the computational complexities of the simplified versions are much lower. As expected, the SKF-BF outperforms the SKF in terms of convergence rate.

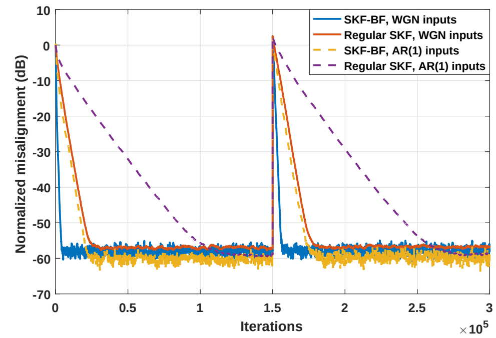

Fig. 2: Normalized misalignment of the SKF-BF and regular SKF for different types of input signals. Other conditions same as in Fig. 1.

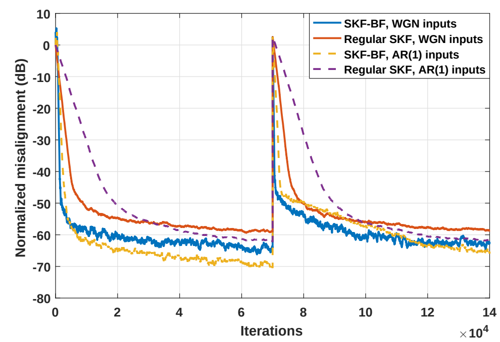

Fig. 3: Normalized misalignment of the SKF-BF and regular SKF (for different types of input signals), using the recursive estimates $\widehat{\sigma}_{w_\mathbf{h}}^2(t)$ and $\widehat{\sigma}_w^2(t)$, respectively. The length of the global impulse responses is $ML = 512$. The other parameters are set to $\sigma_v^2 = 0.01$, $\sigma_{w_\mathbf{g}}^2 = 0$, and $\epsilon = 10^{-5}$.

In Fig. 3, the performance of the SKF-BF is evaluated, using the recursive estimation $\widehat{\sigma}_{w_\mathbf{h}}^2(t)$ from (20), instead of a constant value as in the previous experiments. Also, the spatial impulse response is assumed to be time invariant, so that we can set $\sigma_{w_\mathbf{g}}^2 = 0$. For comparison, the regular SKF is involved, using a similar way to estimate its specific parameter, i.e., $\widehat{\sigma}_w^2(t) = \left\|\widehat{\mathbf{f}}(t) - \widehat{\mathbf{f}}(t-1)\right\|_2^2/(ML)$. Due to the nature of these estimators (as explained in the end of Section 5), the algorithms perform now similar to the variable step-size adaptive filters, achieving both low misalignment and fast tracking. However, as we can notice in Fig. 3, the proposed SKF-BF still outperforms the regular SKF in terms of both performance criteria.

7. CONCLUSIONS

In this paper, we have presented the Kalman filter tailored for the identification of bilinear forms (namely KF-BF), where the bilinear term has been defined with respect to the impulse responses of the spatiotemporal model. Also, a simplified version of this algorithm has been derived, i.e., the SKF-BF, which offers a reduced computational complexity; the price to pay is a slower convergence rate, especially for correlated inputs. Simulation results indicate that the proposed algorithms could represent appealing solutions for such bilinear system identification problems.

8. REFERENCES

[1] R. R. Mohler and W. J. Kolodziej, "An overview of bilinear system theory and applications," *IEEE Trans. Systems, Man, Cybernetics*, vol. SMC-10, pp. 683–688, Oct. 1980.

[2] H. K. Baik and V. J. Mathews, "Adaptive lattice bilinear filters," *IEEE Trans. Signal Process.*, vol. 41, pp. 2033–2046, June 1993.

[3] S. Han, J. Kim, and K. Sung, "Extended generalized total least squares method for the identification of bilinear systems," *IEEE Trans. Signal Process.*, vol. 44, pp. 1015–1018, Apr. 1996.

[4] V. Tsoulkas, P. Koukoulas, and N. Kalouptsidis, "Identification of input-output bilinear systems using cumulants," *IEEE Trans. Signal Process.*, vol. 51, pp. 2753–2761, Nov. 2001.

[5] 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.

[6] U. Forssén, "Adaptive bilinear digital filters," *IEEE Trans. Circuits Systems-II: Analog and Digital Signal Process.*, vol. 40, pp. 729–735, Nov. 1993.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] C. Paleologu, J. Benesty, and S. Ciochină, "Adaptive filtering for the identification of bilinear forms," *Digital Signal Processing*, to appear.

[13] D. Gesbert and P. Duhamel, "Robust blind joint data/channel estimation based on bilinear optimization," in *Proc. IEEE WSSAP*, 1996, pp. 168–171.

[14] A. Stenger, W. Kellermann, R. Rabenstein, "Adaptation of acoustic echo cancellers incorporating a memoryless nonlinearity," in *Proc. IEEE IWAENC*, 1999 (4 pages).

[15] A. Stenger and W. Kellermann, "Adaptation of a memoryless preprocessor for nonlinear acoustic echo cancelling," *Signal Process.*, vol. 80, pp. 1747–1760, Sept. 2000.

[16] M. Rupp and S. Schwarz, "A tensor LMS algorithm," in *Proc. IEEE ICASSP*, 2015, pp. 3347–3351.

[17] Y. Huang, J. Skoglund, and A. Luebs, "Practically efficient nonlinear acoustic echo cancellers using cascaded block RLS and FLMS adaptive filters," in *Proc. IEEE ICASSP*, 2017, pp. 596–570.

[18] R. E. Kalman, "A new approach to linear filtering and prediction problems," *J. Basic Engineering*, vol. 82, pp. 35–45, Mar. 1960.

[19] 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.

[20] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory*. Englewood Cliffs, NJ: Prentice Hall, 1993.

[21] C. Paleologu, J. Benesty, and S. Ciochină, "Study of the general Kalman filter for echo cancellation," *IEEE Trans. Audio, Speech, Language Process.*, vol 21, pp. 1539–1549, Aug. 2013.

[22] C. Paleologu, J. Benesty, and S. Ciochină, "Study of the optimal and simplified Kalman filters for echo cancellation," *Proc. IEEE ICASSP*, 2013, pp. 580–584.

[23] M. A. Iqbal and S. L. Grant, "Novel variable step size NLMS algorithm for echo cancellation," in *Proc. IEEE ICASSP*, 2008, pp. 241–244.

[24] C. Paleologu, S. Ciochină, and J. Benesty, "Double-talk robust VSS-NLMS algorithm for under-modeling acoustic echo cancellation," in *Proc. IEEE ICASSP*, 2008, pp. 245–248.