sensors

MDPI

Article

A Kalman Filter for Multilinear Forms and Its Connection with Tensorial Adaptive Filters †

Laura-Maria Dogariu [1], Constantin Paleologu [1,*], Jacob Benesty [2], Cristian-Lucian Stanciu [1], Claudia-Cristina Oprea [1] and Silviu Ciochină [1]

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

[2] INRS-EMT, University of Quebec, Montreal, QC H5A 1K6, Canada; Jacob.Benesty@inrs.ca

* Correspondence: pale@comm.pub.ro

† This paper is an extended version of our paper published in Dogariu, L.-M.; Paleologu, C.; Benesty, J.; Oprea, C.; Ciochină, S. LMS Algorithms for Multilinear Forms. In Proceedings of the 2020 14th International Symposium on Electronics and Telecommunications (ISETC), Timişoara, Romania, 5–6 November 2020.

Abstract: The Kalman filter represents a very popular signal processing tool, with a wide range of applications within many fields. Following a Bayesian framework, the Kalman filter recursively provides an optimal estimate of a set of unknown variables based on a set of noisy observations. Therefore, it fits system identification problems very well. Nevertheless, such scenarios become more challenging (in terms of the convergence and accuracy of the solution) when the parameter space becomes larger. In this context, the identification of linearly separable systems can be efficiently addressed by exploiting tensor-based decomposition techniques. Such multilinear forms can be modeled as rank-1 tensors, while the final solution is obtained by solving and combining low-dimension system identification problems related to the individual components of the tensor. Recently, the identification of multilinear forms was addressed based on the Wiener filter and most well-known adaptive algorithms. In this work, we propose a tensorial Kalman filter tailored to the identification of multilinear forms. Furthermore, we also show the connection between the proposed algorithm and other tensor-based adaptive filters. Simulation results support the theoretical findings and show the appealing performance features of the proposed Kalman filter for multilinear forms.

Keywords: adaptive filters; Kalman filter; multilinear forms; system identification; tensor decomposition

check for updates

Citation: Dogariu, L.-M.; Paleologu, C.; Benesty, J.; Stanciu, C.-L.; Oprea, C.-C.; Ciochină, S. A Kalman Filter for Multilinear Forms and Its Connection with Tensorial Adaptive Filters. *Sensors* **2021**, *21*, 3555. https://doi.org/10.3390/s21103555

Academic Editor: Anastasios Doulamis

Received: 23 April 2021
Accepted: 17 May 2021
Published: 20 May 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 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 (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The identification of multilinear forms (or linearly separable systems) can be efficiently exploited in the framework of different applications, like channel equalization [1,2], nonlinear acoustic echo cancellation [3,4], source separation [5,6], array beamforming [7,8], and object recognition [9,10]. In these contexts, the basic approach relies on tensor decomposition and modeling techniques [11–14], since the multilinear forms can be modeled as rank-1 tensors. The main idea is to combine (i.e., "tensorize") the solutions to low-dimension problems, in order to efficiently solve a multidimensional system identification problem, which is usually characterized by a large parameter space. Such scenarios can appear in the framework of multichannel systems, e.g., those with a large number of sensors and actuators, such as in active noise control systems [15], adaptive beamforming [16], microphones arrays [17], etc.

The particular cases of bilinear and trilinear forms have been previously addressed in the literature from a system identification perspective. In [18], an iterative Wiener filter for bilinear forms was developed in the framework of a multiple-input/single-output (MISO) system. Compared to the conventional Wiener solution, the iterative version can obtain a good accuracy of the solution, even when a few data are available for the estimation of the

statistics. Furthermore, in [19], this solution was extended to the identification of trilinear forms, based on the decomposition of third-order tensors (of rank-1). Since there are inherent limitations to the Wiener filters (time-invariant framework, matrix inversion operation, etc.), a more appropriate and practical solution relies on adaptive filtering [20–22]. Consequently, several adaptive filters tailored to the identification of bilinear and trilinear forms have also been developed, following the main categories of algorithms. For example, the least-mean-square (LMS) and normalized LMS (NLMS) versions can be found in [23,24]. In addition, the recursive least-squares (RLS) algorithms for bilinear and trilinear forms were developed in [25,26], respectively. These algorithms have improved convergence features as compared to their LMS-based counterparts. Moreover, a Kalman filter tailored to the identification of bilinear forms was analyzed in [27,28].

These adaptive solutions are suitable in real-world scenarios, e.g., working in nonstationary environments and/or requiring real-time processing. Among them, the Kalman filter represents a very appealing choice [29–31]. As compared to other adaptive filters, where the system to be identified is considered to be deterministic in their derivations, the Kalman filter takes the "uncertainties" in the system into account, and is thus successfully employed in a wide range of applications, e.g., [32–37] and the references therein. Recently, in [36], an adaptive Kalman-filter-based variational Bayesian, which achieves a simultaneous estimation of the process noise covariance matrix and of the measurement noise covariance matrix, is presented, with applications in target tracking. In [37], the authors propose a multiple strong tracking adaptive square-root cubature Kalman filter, which can be used to improve the in-flight alignment, with applications in guided weaponry, unmanned automatic vehicles, and robots. The numerous different fields of applicability of the Kalman filter represent the main motivation behind the development presented in this paper, which targets the derivation of such a filter, tailored to the identification of multilinear forms.

Recently, an iterative Wiener filter was designed for multilinear forms [38], followed by the adaptive solutions based on the LMS and RLS algorithms [39,40]. The goal of this paper is twofold. First, it extends the work in [39], by proposing a Kalman filter for multilinear forms, with improved convergence features compared to the LMS-based counterparts. Second, it establishes the connection between the proposed Kalman filter and the tensor-based adaptive algorithms presented in [40], showing how these algorithms can be obtained as simplified versions of the Kalman filter for multilinear forms.

The rest of the paper is organized as follows. In Section 2, we present the system model behind the multilinear framework, which is formulated in the context of an MISO system identification problem. The proposed tensorial Kalman filter for the identification of such multilinear forms is derived in Section 3. Furthermore, in Section 4, we show how this algorithm is connected with the main tensor-based adaptive filters, which belong to the LMS and RLS families. Simulation results provided in Section 5 support the theoretical findings and indicate the good performance of the proposed Kalman filter for multilinear forms. Finally, in Section 6, the main conclusions are outlined, together with several perspectives for future works.

2. Multilinear Framework for MISO System Identification

Let us consider a real-valued MISO system with N individual channels, which are modeled as finite-impulse-response (FIR) filters of lengths L_n, $n = 1, 2, \ldots, N$. The impulse responses of these channels at the discrete-time index t are characterized by the vectors $\mathbf{h}_n(t) = [h_{n,1}(t)\ h_{n,2}(t)\ \cdots\ h_{n,L_n}(t)]^T$, $n = 1, 2, \ldots, N$, where the superscript ^{T} denotes the transpose operator. Furthermore, we assume that $\mathbf{h}_n(t)$, $n = 1, 2, \ldots, N$ are zero-mean random vectors, which follow a simplified first-order Markov model

$$\mathbf{h}_n(t) = \mathbf{h}_n(t-1) + \mathbf{w}_n(t), \tag{1}$$

where $\mathbf{w}_n(t)$, $n = 1, 2, \ldots, N$ are zero-mean white Gaussian noise vectors [uncorrelated with $\mathbf{h}_n(t-1)$], whose correlation matrices are $\mathbf{R}_{\mathbf{w}_n} = \sigma_{w_n}^2 \mathbf{I}_{L_n}$, $n = 1, 2, \ldots, N$, with $\mathbf{I}_{L_n}$

being an identity matrix of size $L_n \times L_n$. The variances $\sigma^2_{w_n}$, $n = 1, 2, \ldots, N$ capture the uncertainties in the corresponding channels $\mathbf{h}_n(t)$, $n = 1, 2, \ldots, N$.

The impulse responses of the channels can be grouped in a tensorial form, i.e., $\boldsymbol{\mathcal{H}}(t) \in \mathbb{R}^{L_1 \times L_2 \times \cdots \times L_N}$, such that

$$\boldsymbol{\mathcal{H}}(t) = \mathbf{h}_1(t) \circ \mathbf{h}_2(t) \circ \cdots \circ \mathbf{h}_N(t), \tag{2}$$

where $\circ$ denotes the outer product. Therefore, the elements of this rank-1 tensor are $(\boldsymbol{\mathcal{H}})_{l_1, l_2, \ldots, l_N}(t) = h_{1,l_1}(t) h_{2,l_2}(t) \cdots h_{N,l_N}(t)$. In addition, using the vectorization operation, $\mathrm{vec}(\cdot)$, we can write

$$\mathrm{vec}[\boldsymbol{\mathcal{H}}(t)] = \mathbf{h}_N(t) \otimes \mathbf{h}_{N-1}(t) \otimes \cdots \otimes \mathbf{h}_1(t), \tag{3}$$

where $\otimes$ denotes the Kronecker product.

The real-valued input signals that feed into the MISO system are described in the tensorial form $\boldsymbol{\mathcal{X}}(t) \in \mathbb{R}^{L_1 \times L_2 \times \cdots \times L_N}$, having the elements $(\boldsymbol{\mathcal{X}})_{l_1 l_2 \ldots l_N}(t) = x_{l_1 l_2 \ldots l_N}(t)$. Consequently, the output signal at the discrete-time index t results in

$$\begin{aligned} y(t) &= \boldsymbol{\mathcal{X}}(t) \times_1 \mathbf{h}_1^T(t) \times_2 \mathbf{h}_2^T(t) \times_3 \cdots \times_N \mathbf{h}_N^T(t) \\ &= \sum_{l_1=1}^{L_1} \sum_{l_2=1}^{L_2} \cdots \sum_{l_N=1}^{L_N} x_{l_1 l_2 \ldots l_N}(t) h_{1,l_1}(t) h_{2,l_2}(t) \cdots h_{N,l_N}(t) \\ &= \mathrm{vec}^T[\boldsymbol{\mathcal{H}}(t)] \mathrm{vec}[\boldsymbol{\mathcal{X}}(t)], \end{aligned} \tag{4}$$

where $\times_n$ denotes the mode-n product [5]. This represents a multilinear form, since it is a linear function of each of the vectors $\mathbf{h}_n(t)$, $n = 1, 2, \ldots, N$, considering that the other $N - 1$ components are fixed.

The last line in (4) represents a particular case of the MISO system identification problem, which resembles a single-input, single-output (SISO) scenario. Using the notation $\mathbf{x}(t) = \mathrm{vec}[\boldsymbol{\mathcal{X}}(t)]$ and $\mathbf{h}(t) = \mathrm{vec}[\boldsymbol{\mathcal{H}}(t)]$, the output signal from (4) becomes

$$y(t) = \mathbf{h}^T(t) \mathbf{x}(t), \tag{5}$$

where the vector $\mathbf{h}(t)$ represents the global impulse response of the system. Therefore, in this multilinear framework, the system identification problem can be formulated in two ways. First, this can be approached in terms of estimating the individual channels, $\mathbf{h}_n(t)$, $n = 1, 2, \ldots, N$. Alternatively, we can focus on the identification of the global impulse response, $\mathbf{h}(t)$, as in a conventional SISO system identification problem.

At this point, there are two important aspects that should be outlined. The global impulse response, $\mathbf{h}(t)$, is of length $L = \prod_{n=1}^{N} L_n$, but this results as a combination of $\sum_{n=1}^{N} L_n$ elements, which are the coefficients of the individual impulse responses, $\mathbf{h}_n(t)$, $n = 1, 2, \ldots, N$. Additionally, according to (3), we have

$$\begin{aligned} \mathbf{h}(t) &= \mathbf{h}_N(t) \otimes \mathbf{h}_{N-1}(t) \otimes \cdots \otimes \mathbf{h}_1(t) \\ &= \eta_N \mathbf{h}_N(t) \otimes \eta_{N-1} \mathbf{h}_{N-1}(t) \otimes \cdots \otimes \eta_1 \mathbf{h}_1(t), \end{aligned} \tag{6}$$

with $\eta_n \in \mathbb{R}$, $\eta_n \neq 0$ (for any $n = 1, 2, \ldots, N$), and $\prod_{n=1}^{N} \eta_n = 1$, so that the decomposition of $\mathbf{h}(t)$ is not unique. Consequently, from a system identification perspective, it would be more advantageous to approach the problem in terms of estimating the individual impulse responses $\mathbf{h}_n(t)$, $n = 1, 2, \ldots, N$, while the estimated global impulse response results similar to (6). On the other hand, the Kronecker product-based decomposition of $\mathbf{h}(t)$ does not lead to a unique set of estimates for the individual channels. However, there is no scaling ambiguity when identifying the global impulse response. In sum, the main idea behind the decomposition-based approach is to reformulate a high-dimension system identification problem as a combination of low-dimension solutions, which are "tensorized" together.

In realistic system identification scenarios, the output signal $y(t)$ is usually corrupted by an additive noise, $v(t)$, which is uncorrelated with the input signals. The variance of this noise signal is $\sigma_v^2 = E[v^2(t)]$, where $E[\cdot]$ denotes mathematical expectation. Thus, the reference (or desired) signal at the discrete-time index t results in

$$d(t) = y(t) + v(t). \tag{7}$$

The goal is to estimate the output of the system (or, equivalently, the impulse responses of the channels), given the reference and the input signals. In this context, the conventional Wiener filter provides the well-known solution [20–22]

$$\widehat{\mathbf{h}}_{\mathrm{W}} = \mathbf{R}^{-1}\mathbf{p}, \tag{8}$$

where $\widehat{\mathbf{h}}_{\mathrm{W}}$ is an estimate of the global impulse response, while $\mathbf{R} = E\left[\mathbf{x}(t)\mathbf{x}^T(t)\right]$ and $\mathbf{p} = E[d(t)\mathbf{x}(t)]$ represent the covariance matrix and the cross-correlation vector, respectively. Recently, an iterative Wiener filter [38] was developed in the framework of multilinear forms. It exploits the decomposition of the global impulse response, while the optimization criterion is applied, following a block coordinate descent approach, to the individual components [41]. As compared to the conventional Wiener filter from (8), the iterative version of multilinear forms leads to a superior performance, especially when a small amount of data are available for the estimation of $\mathbf{R}$ and $\mathbf{p}$. Nevertheless, both previously mentioned Wiener solutions present several limitations, such as the time-invariant framework, the matrix inversion operation, and the estimation of the statistics. These could make them unsuitable in real-world scenarios, e.g., working in nonstationary environments and/or requiring real-time processing.

In this context, a more appropriate approach is adaptive filtering. Since the LMS algorithm represents one of the simplest and most practical solutions, the LMS-based algorithms tailored to the identification of multilinear forms were developed in [39]. Furthermore, the RLS algorithm for multilinear forms was introduced in [40]. These versions are also referred to as tensor-based adaptive algorithms. They rely on the minimization of cost functions that depend on the error signal, i.e., the difference between the reference signal and the estimated output of the system. Nevertheless, in a realistic system identification framework, [i.e., in the presence of the system noise, according to (7)], the goal of the adaptive filter is not to make the error signal reach zero. The objective, instead, is to recover this system noise from the error signal of the adaptive filter, after it converges to the true solution. Consequently, it makes more sense to minimize the system misalignment, i.e., a measure of the difference between the true impulse response of the system and the estimated one. This is the optimization approach behind Kalman filtering. Moreover, the Kalman filter uses a specific parameter that captures the uncertainties in the system to be identified, as outlined in the discussion that follows (1), related to $\sigma_{w_n}^2$, $n = 1, 2, \ldots, N$. These parameters could act as control factors. On the other hand, the LMS and RLS adaptive filters do not depend on these uncertainties, since the impulse responses of the channels are considered as deterministic in the derivation of these algorithms. Therefore, the specific parameters of the Kalman filter would allow for better control.

3. Kalman Filter for the Identification of Multilinear Forms

Following (5), we can introduce the a priori error signal

$$\begin{aligned} e(t) &= d(t) - \widehat{y}(t) \\ &= d(t) - \widehat{\mathbf{h}}^T(t-1)\mathbf{x}(t), \end{aligned} \tag{9}$$

where $\widehat{y}(t)$ is the estimated output signal and $\widehat{\mathbf{h}}(t-1)$ represents an estimate of the global impulse response at the discrete-time index $t-1$. Since the global impulse response can be deconstructed based on (6), we can also consider that $\widehat{\mathbf{h}}(t)$ similarly results in a combination

of the estimated impulse responses of the channels, denoted by $\widehat{\mathbf{h}}_n(t),\ n = 1,2,\ldots,N$. Therefore, we have

$$\widehat{\mathbf{h}}(t) = \widehat{\mathbf{h}}_N(t) \otimes \widehat{\mathbf{h}}_{N-1}(t) \otimes \cdots \otimes \widehat{\mathbf{h}}_2(t) \otimes \widehat{\mathbf{h}}_1(t). \tag{10}$$

Alternatively, using the properties of the Kronecker product [42], (10) can be expressed in N equivalent ways, i.e.,

$$\begin{aligned}
\widehat{\mathbf{h}}(t) &= \left[\widehat{\mathbf{h}}_N(t) \otimes \widehat{\mathbf{h}}_{N-1}(t) \otimes \cdots \otimes \widehat{\mathbf{h}}_2(t) \otimes \mathbf{I}_{L_1}\right]\widehat{\mathbf{h}}_1(t) \\
&= \left[\widehat{\mathbf{h}}_N(t) \otimes \widehat{\mathbf{h}}_{N-1}(t) \otimes \cdots \otimes \mathbf{I}_{L_2} \otimes \widehat{\mathbf{h}}_1(t)\right]\widehat{\mathbf{h}}_2(t) \\
&\vdots \\
&= \left[\mathbf{I}_{L_N} \otimes \widehat{\mathbf{h}}_{N-1}(t) \otimes \cdots \otimes \widehat{\mathbf{h}}_2(t) \otimes \widehat{\mathbf{h}}_1(t)\right]\widehat{\mathbf{h}}_N(t).
\end{aligned}$$

Consequently, the a priori error signal from (9) can also be rewritten in N equivalent forms (targeting the individual filters), as follows

$$e_{\widehat{\mathbf{h}}_1}(t) = d(t) - \widehat{\mathbf{h}}_1^T(t-1)\mathbf{x}_{\widehat{\mathbf{h}}_2\widehat{\mathbf{h}}_3\ldots\widehat{\mathbf{h}}_N}(t), \tag{11}$$

$$e_{\widehat{\mathbf{h}}_2}(t) = d(t) - \widehat{\mathbf{h}}_2^T(t-1)\mathbf{x}_{\widehat{\mathbf{h}}_1\widehat{\mathbf{h}}_3\ldots\widehat{\mathbf{h}}_N}(t), \tag{12}$$

$$\vdots$$

$$e_{\widehat{\mathbf{h}}_N}(t) = d(t) - \widehat{\mathbf{h}}_N^T(t-1)\mathbf{x}_{\widehat{\mathbf{h}}_1\widehat{\mathbf{h}}_2\ldots\widehat{\mathbf{h}}_{N-1}}(t), \tag{13}$$

where

$$\mathbf{x}_{\widehat{\mathbf{h}}_2\widehat{\mathbf{h}}_3\ldots\widehat{\mathbf{h}}_N}(t) = \left[\widehat{\mathbf{h}}_N(t-1) \otimes \widehat{\mathbf{h}}_{N-1}(t-1) \otimes \cdots \otimes \widehat{\mathbf{h}}_2(t-1) \otimes \mathbf{I}_{L_1}\right]^T\mathbf{x}(t), \tag{14}$$

$$\mathbf{x}_{\widehat{\mathbf{h}}_1\widehat{\mathbf{h}}_3\ldots\widehat{\mathbf{h}}_N}(t) = \left[\widehat{\mathbf{h}}_N(t-1) \otimes \widehat{\mathbf{h}}_{N-1}(t-1) \otimes \cdots \otimes \mathbf{I}_{L_2} \otimes \widehat{\mathbf{h}}_1(t-1)\right]^T\mathbf{x}(t), \tag{15}$$

$$\vdots$$

$$\mathbf{x}_{\widehat{\mathbf{h}}_1\widehat{\mathbf{h}}_2\ldots\widehat{\mathbf{h}}_{N-1}}(t) = \left[\mathbf{I}_{L_N} \otimes \widehat{\mathbf{h}}_{N-1}(t-1) \otimes \cdots \otimes \widehat{\mathbf{h}}_2(t-1) \otimes \widehat{\mathbf{h}}_1(t-1)\right]^T\mathbf{x}(t). \tag{16}$$

In a similar manner, the reference signal from (7) can be expressed in N equivalent ways, which can be summarized as

$$d(n) = \mathbf{h}_n^T(t)\mathbf{x}_{\mathbf{h}_1\mathbf{h}_2\ldots\mathbf{h}_{n-1}\mathbf{h}_{n+1}\ldots\mathbf{h}_N}(t) + v(n), \tag{17}$$

where

$$\begin{aligned}
\mathbf{x}_{\mathbf{h}_1\mathbf{h}_2\ldots\mathbf{h}_{n-1}\mathbf{h}_{n+1}\ldots\mathbf{h}_N}(t) = [&\mathbf{h}_N(t-1) \otimes \mathbf{h}_{N-1}(t-1) \otimes \cdots \otimes \mathbf{h}_{n+1}(t-1) \otimes \mathbf{I}_{L_n} \\
&\otimes \mathbf{h}_{n-1}(t-1) \otimes \cdots \otimes \mathbf{h}_2(t-1) \otimes \mathbf{h}_1(t-1)]^T\mathbf{x}(t),
\end{aligned}$$

with $n = 1,2,\ldots,N$. For any value of $n = 1,2,\ldots,N$, expression (17) plays the role of an observation equation, while (1) represents a state equation. In the framework of multilinear forms, having N such pairs of state and observation equations, the objective is to find the optimal recursive estimator of $\mathbf{h}_n(t)$, for $n = 1,2,\ldots,N$, i.e., $\widehat{\mathbf{h}}_n(t)$. To this end, we will follow a multilinear optimization approach [41], by considering that $N-1$ impulse responses are fixed for all the previous time indices, while optimizing the remaining one at the current time index. In this case, within the optimization criterion of $\widehat{\mathbf{h}}_n(t),\ n = 1,2,\ldots,N$, we may assume that $\mathbf{x}_{\widehat{\mathbf{h}}_1\widehat{\mathbf{h}}_2\ldots\widehat{\mathbf{h}}_{n-1}\widehat{\mathbf{h}}_{n+1}\ldots\widehat{\mathbf{h}}_N}(t) \approx \mathbf{x}_{\mathbf{h}_1\mathbf{h}_2\ldots\mathbf{h}_{n-1}\mathbf{h}_{n+1}\ldots\mathbf{h}_N}(t)$.

Under these considerations and based on (17), we can introduce the a posteriori errors related to the N individual filters, which result in

$$\begin{aligned}\epsilon_{\widehat{\mathbf{h}}_n}(t) &= d(t) - \widehat{\mathbf{h}}_n^T(t)\mathbf{x}_{\widehat{\mathbf{h}}_1\widehat{\mathbf{h}}_2\ldots\widehat{\mathbf{h}}_{n-1}\widehat{\mathbf{h}}_{n+1}\ldots\widehat{\mathbf{h}}_N}(t)\\ &= \mathbf{h}_n^T(t)\mathbf{x}_{\mathbf{h}_1\mathbf{h}_2\ldots\mathbf{h}_{n-1}\mathbf{h}_{n+1}\ldots\mathbf{h}_N}(t) + v(t) - \widehat{\mathbf{h}}_n^T(t)\mathbf{x}_{\widehat{\mathbf{h}}_1\widehat{\mathbf{h}}_2\ldots\widehat{\mathbf{h}}_{n-1}\widehat{\mathbf{h}}_{n+1}\ldots\widehat{\mathbf{h}}_N}(t)\\ &\approx \left[\mathbf{h}_n^T(t) - \widehat{\mathbf{h}}_n^T(t)\right]\mathbf{x}_{\widehat{\mathbf{h}}_1\widehat{\mathbf{h}}_2\ldots\widehat{\mathbf{h}}_{n-1}\widehat{\mathbf{h}}_{n+1}\ldots\widehat{\mathbf{h}}_N}(t) + v(t)\\ &= \boldsymbol{\mu}_{\mathbf{h}_n}^T(t)\mathbf{x}_{\widehat{\mathbf{h}}_1\widehat{\mathbf{h}}_2\ldots\widehat{\mathbf{h}}_{n-1}\widehat{\mathbf{h}}_{n+1}\ldots\widehat{\mathbf{h}}_N}(t) + v(t),\end{aligned} \tag{18}$$

where $\boldsymbol{\mu}_{\mathbf{h}_n}(t) = \mathbf{h}_n(t) - \widehat{\mathbf{h}}_n(t)$ is the a posteriori misalignment (or the state estimation error) of the nth individual filter, with $n = 1, 2, \ldots, N$. Similarly, we can define the a priori misalignments $\mathbf{m}_{\mathbf{h}_n}(t) = \mathbf{h}_n(t) - \widehat{\mathbf{h}}_n(t-1)$, $n = 1, 2, \ldots, N$, so that [based on (1)]

$$\mathbf{m}_{\mathbf{h}_n}(t) = \boldsymbol{\mu}_{\mathbf{h}_n}(t-1) + \mathbf{w}_n(t) \tag{19}$$

and, consequently,

$$\mathbf{R}_{\mathbf{m}_{\mathbf{h}_n}}(t) = \mathbf{R}_{\boldsymbol{\mu}_{\mathbf{h}_n}}(t-1) + \sigma_{w_n}^2\mathbf{I}_{L_n}, \tag{20}$$

for $n = 1, 2, \ldots, N$, where $\mathbf{R}_{\mathbf{m}_{\mathbf{h}_n}}(t) = E\left[\mathbf{m}_{\mathbf{h}_n}(t)\mathbf{m}_{\mathbf{h}_n}^T(t)\right]$ and $\mathbf{R}_{\boldsymbol{\mu}_{\mathbf{h}_n}}(t) = E\left[\boldsymbol{\mu}_{\mathbf{h}_n}(t)\boldsymbol{\mu}_{\mathbf{h}_n}^T(t)\right]$ represent the correlation matrices of the a priori and a posteriori misalignments, respectively.

As explained in Section 2, according to (6), we can only identify the individual impulse responses up to some arbitrary scaling factors, η_n, $n = 1, 2, \ldots, N$. However, in terms of identifying the global impulse response (or, equivalently, the output signal), the group $\eta_1\mathbf{h}_1(t), \eta_2\mathbf{h}_2(t), \ldots, \eta_N\mathbf{h}_N(t)$ is equivalent to the group $\mathbf{h}_1(t), \mathbf{h}_2(t), \ldots, \mathbf{h}_N(t)$. Thus, in order to simplify the notation, the scaling factors do not appear explicitly in the misalignments.

In the context of the multilinear optimization strategy and the linear sequential Bayesian approach [43], the optimum estimates of the state vectors, $\widehat{\mathbf{h}}_n(t)$, $n = 1, 2, \ldots, N$, have the recursive forms

$$\begin{aligned}\widehat{\mathbf{h}}_1(t) &= \widehat{\mathbf{h}}_1(t-1) + \mathbf{k}_{\widehat{\mathbf{h}}_1}(t)\left[d(t) - \widehat{\mathbf{h}}_1^T(t-1)\mathbf{x}_{\widehat{\mathbf{h}}_2\widehat{\mathbf{h}}_3\ldots\widehat{\mathbf{h}}_N}(t)\right]\\ &= \widehat{\mathbf{h}}_1(t-1) + \mathbf{k}_{\widehat{\mathbf{h}}_1}(t)e(t),\end{aligned} \tag{21}$$

$$\begin{aligned}\widehat{\mathbf{h}}_2(t) &= \widehat{\mathbf{h}}_2(t-1) + \mathbf{k}_{\widehat{\mathbf{h}}_2}(t)\left[d(t) - \widehat{\mathbf{h}}_2^T(t-1)\mathbf{x}_{\widehat{\mathbf{h}}_1\widehat{\mathbf{h}}_3\ldots\widehat{\mathbf{h}}_N}(t)\right]\\ &= \widehat{\mathbf{h}}_2(t-1) + \mathbf{k}_{\widehat{\mathbf{h}}_2}(t)e(t),\end{aligned} \tag{22}$$

$$\vdots$$

$$\begin{aligned}\widehat{\mathbf{h}}_N(t) &= \widehat{\mathbf{h}}_N(t-1) + \mathbf{k}_{\widehat{\mathbf{h}}_N}(t)\left[d(t) - \widehat{\mathbf{h}}_N^T(t-1)\mathbf{x}_{\widehat{\mathbf{h}}_1\widehat{\mathbf{h}}_2\ldots\widehat{\mathbf{h}}_{N-1}}(t)\right]\\ &= \widehat{\mathbf{h}}_N(t-1) + \mathbf{k}_{\widehat{\mathbf{h}}_N}(t)e(t),\end{aligned} \tag{23}$$

where $\mathbf{k}_{\widehat{\mathbf{h}}_n}(t)$, $n = 1, 2, \ldots, N$ are the so-called Kalman gain vectors, and $e_{\widehat{\mathbf{h}}_1}(t) = e_{\widehat{\mathbf{h}}_2}(t) = \cdots = e_{\widehat{\mathbf{h}}_N}(t) = e(t)$ [based on (9) and (11)–(13)]. Next, the Kalman gain vectors are obtained by minimizing the cost functions:

$$J_{\widehat{\mathbf{h}}_2\widehat{\mathbf{h}}_3\ldots\widehat{\mathbf{h}}_N}\left[\mathbf{k}_{\widehat{\mathbf{h}}_1}(t)\right] = \frac{1}{L_1}\mathrm{tr}\left[\mathbf{R}_{\boldsymbol{\mu}_{\mathbf{h}_1}}(t)\right], \tag{24}$$

$$J_{\widehat{\mathbf{h}}_1\widehat{\mathbf{h}}_3\ldots\widehat{\mathbf{h}}_N}\left[\mathbf{k}_{\widehat{\mathbf{h}}_2}(t)\right] = \frac{1}{L_2}\mathrm{tr}\left[\mathbf{R}_{\boldsymbol{\mu}_{\mathbf{h}_2}}(t)\right], \tag{25}$$

$$\vdots$$

$$J_{\widehat{\mathbf{h}}_1\widehat{\mathbf{h}}_2\ldots\widehat{\mathbf{h}}_{N-1}}\left[\mathbf{k}_{\widehat{\mathbf{h}}_N}(t)\right] = \frac{1}{L_N}\mathrm{tr}\left[\mathbf{R}_{\boldsymbol{\mu}_{\mathbf{h}_N}}(t)\right], \tag{26}$$

where $\mathrm{tr}[\cdot]$ denotes the trace of a square matrix. This leads to

$$\mathbf{k}_{\widehat{\mathbf{h}}_1}(t) = \frac{\mathbf{R}_{\mathbf{m}_{\mathbf{h}_1}}(t)\mathbf{x}_{\widehat{\mathbf{h}}_2\widehat{\mathbf{h}}_3\ldots\widehat{\mathbf{h}}_N}(t)}{\mathbf{x}^T_{\widehat{\mathbf{h}}_2\widehat{\mathbf{h}}_3\ldots\widehat{\mathbf{h}}_N}(t)\mathbf{R}_{\mathbf{m}_{\mathbf{h}_1}}(t)\mathbf{x}_{\widehat{\mathbf{h}}_2\widehat{\mathbf{h}}_3\ldots\widehat{\mathbf{h}}_N}(t) + \sigma_v^2}, \tag{27}$$

$$\mathbf{k}_{\widehat{\mathbf{h}}_2}(t) = \frac{\mathbf{R}_{\mathbf{m}_{\mathbf{h}_2}}(t)\mathbf{x}_{\widehat{\mathbf{h}}_1\widehat{\mathbf{h}}_3\ldots\widehat{\mathbf{h}}_N}(t)}{\mathbf{x}^T_{\widehat{\mathbf{h}}_1\widehat{\mathbf{h}}_3\ldots\widehat{\mathbf{h}}_N}(t)\mathbf{R}_{\mathbf{m}_{\mathbf{h}_2}}(t)\mathbf{x}_{\widehat{\mathbf{h}}_1\widehat{\mathbf{h}}_3\ldots\widehat{\mathbf{h}}_N}(t) + \sigma_v^2}, \tag{28}$$

$$\vdots$$

$$\mathbf{k}_{\widehat{\mathbf{h}}_N}(t) = \frac{\mathbf{R}_{\mathbf{m}_{\mathbf{h}_N}}(t)\mathbf{x}_{\widehat{\mathbf{h}}_1\widehat{\mathbf{h}}_2\ldots\widehat{\mathbf{h}}_{N-1}}(t)}{\mathbf{x}^T_{\widehat{\mathbf{h}}_1\widehat{\mathbf{h}}_2\ldots\widehat{\mathbf{h}}_{N-1}}(t)\mathbf{R}_{\mathbf{m}_{\mathbf{h}_N}}(t)\mathbf{x}_{\widehat{\mathbf{h}}_1\widehat{\mathbf{h}}_2\ldots\widehat{\mathbf{h}}_{N-1}}(t) + \sigma_v^2}, \tag{29}$$

and

$$\mathbf{R}_{\boldsymbol{\mu}_{\mathbf{h}_1}}(t) = \left[\mathbf{I}_{L_1} - \mathbf{k}_{\widehat{\mathbf{h}}_1}(t)\mathbf{x}^T_{\widehat{\mathbf{h}}_2\widehat{\mathbf{h}}_3\ldots\widehat{\mathbf{h}}_N}(t)\right]\mathbf{R}_{\mathbf{m}_{\mathbf{h}_1}}(t), \tag{30}$$

$$\mathbf{R}_{\boldsymbol{\mu}_{\mathbf{h}_2}}(t) = \left[\mathbf{I}_{L_2} - \mathbf{k}_{\widehat{\mathbf{h}}_2}(t)\mathbf{x}^T_{\widehat{\mathbf{h}}_1\widehat{\mathbf{h}}_3\ldots\widehat{\mathbf{h}}_N}(t)\right]\mathbf{R}_{\mathbf{m}_{\mathbf{h}_2}}(t), \tag{31}$$

$$\vdots$$

$$\mathbf{R}_{\boldsymbol{\mu}_{\mathbf{h}_N}}(t) = \left[\mathbf{I}_{L_N} - \mathbf{k}_{\widehat{\mathbf{h}}_N}(t)\mathbf{x}^T_{\widehat{\mathbf{h}}_1\widehat{\mathbf{h}}_2\ldots\widehat{\mathbf{h}}_{N-1}}(t)\right]\mathbf{R}_{\mathbf{m}_{\mathbf{h}_N}}(t). \tag{32}$$

The resulting Kalman filter for multilinear forms is defined by the relations (11)–(13), (20), (27)–(32), followed by the updates (21)–(23), as summarized in Table 1. In order to remain consistent with [40], we will refer to this algorithm as the tensor-based Kalman filter (KF-T).

An estimation of the global impulse response can be obtained based on (10). Alternatively, the conventional Kalman filter can be used to find $\widehat{\mathbf{h}}(t)$, based on the observation (7) and a state equation for the global impulse response $\mathbf{h}(t)$ [similar to (1)]. In this case, the computational complexity of the conventional Kalman filter would be proportional to $\mathcal{O}(L^2)$, where $L = \prod_{n=1}^{N} L_n$ (i.e., the length of the global impulse response, as explained in Section 2). On the other hand, the KF-T algorithm combines the solutions of N Kalman-based filters of shorter lengths (i.e., L_n, $n = 1, 2, \ldots, N$), so that its computational cost is proportional to $\sum_{n=1}^{N} \mathcal{O}(L_n^2)$. Moreover, even if they are interconnected, these N individual filters can work in parallel, since the update of each filter at the discrete-time index t depends on the coefficients of all the other filters from the previous time index, i.e., $\widehat{\mathbf{h}}_n(t-1)$, $n = 1, 2, \ldots, N$.

Table 1. Tensor-based Kalman filter (KF-T) for multilinear forms.

Inputs:

$\mathbf{x}(t), d(t)$

Initialization:

$\widehat{\mathbf{h}}_1(0) = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}^T$

$\widehat{\mathbf{h}}_k(0) = \frac{1}{L_k}\begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix}^T,\ k = 2, \ldots, N$

$\mathbf{R}_{\boldsymbol{\mu}_{\mathbf{h}_n}}(0) = \varepsilon_n \mathbf{I}_{L_n},\ \varepsilon_n > 0,\ n = 1, 2, \ldots, N$

Parameters: σ_v^2 and $\sigma_{w_n}^2,\ n = 1, 2, \ldots, N$

For discrete-time index $t = 1, 2, \ldots$

For $n = 1, 2, \ldots, N$

Equations (14)–(16):

$$\mathbf{x}_{\widehat{\mathbf{h}}_1\widehat{\mathbf{h}}_2\ldots\widehat{\mathbf{h}}_{n-1}\widehat{\mathbf{h}}_{n+1}\ldots\widehat{\mathbf{h}}_N}(t) = \Big[\widehat{\mathbf{h}}_N(t-1) \otimes \widehat{\mathbf{h}}_{N-1}(t-1) \otimes \cdots \otimes \widehat{\mathbf{h}}_{n+1}(t-1) \otimes \mathbf{I}_{L_n} \otimes \widehat{\mathbf{h}}_{n-1}(t-1) \otimes \cdots \otimes \widehat{\mathbf{h}}_2(t-1) \otimes \widehat{\mathbf{h}}_1(t-1)\Big]^T \mathbf{x}(t)$$

Equations (11)–(13):

$$\widehat{y}(t) = \widehat{\mathbf{h}}_n^T(t-1)\mathbf{x}_{\widehat{\mathbf{h}}_1\widehat{\mathbf{h}}_2\ldots\widehat{\mathbf{h}}_{n-1}\widehat{\mathbf{h}}_{n+1}\ldots\widehat{\mathbf{h}}_N}(t)$$

$$e(t) = d(t) - \widehat{y}(t)$$

Equations (20):

$$\mathbf{R}_{\mathbf{m}_{\mathbf{h}_n}}(t) = \mathbf{R}_{\boldsymbol{\mu}_{\mathbf{h}_n}}(t-1) + \sigma_{w_n}^2 \mathbf{I}_{L_n}$$

Equations (27)–(29):

$$\mathbf{k}_{\widehat{\mathbf{h}}_n}(t) = \frac{\mathbf{R}_{\mathbf{m}_{\mathbf{h}_n}}(t)\mathbf{x}_{\widehat{\mathbf{h}}_1\widehat{\mathbf{h}}_2\ldots\widehat{\mathbf{h}}_{n-1}\widehat{\mathbf{h}}_{n+1}\ldots\widehat{\mathbf{h}}_N}(t)}{\mathbf{x}^T_{\widehat{\mathbf{h}}_1\widehat{\mathbf{h}}_2\ldots\widehat{\mathbf{h}}_{n-1}\widehat{\mathbf{h}}_{n+1}\ldots\widehat{\mathbf{h}}_N}(t)\mathbf{R}_{\mathbf{m}_{\mathbf{h}_n}}(t)\mathbf{x}_{\widehat{\mathbf{h}}_1\widehat{\mathbf{h}}_2\ldots\widehat{\mathbf{h}}_{n-1}\widehat{\mathbf{h}}_{n+1}\ldots\widehat{\mathbf{h}}_N}(t) + \sigma_v^2}$$

Equations (30)–(32):

$$\mathbf{R}_{\boldsymbol{\mu}_{\mathbf{h}_n}}(t) = \Big[\mathbf{I}_{L_n} - \mathbf{k}_{\widehat{\mathbf{h}}_n}(t)\mathbf{x}^T_{\widehat{\mathbf{h}}_1\widehat{\mathbf{h}}_2\ldots\widehat{\mathbf{h}}_{n-1}\widehat{\mathbf{h}}_{n+1}\ldots\widehat{\mathbf{h}}_N}(t)\Big]\mathbf{R}_{\mathbf{m}_{\mathbf{h}_n}}(t)$$

Equations (21)–(23):

$$\widehat{\mathbf{h}}_n(t) = \widehat{\mathbf{h}}_n(t-1) + \mathbf{k}_{\widehat{\mathbf{h}}_n}(t)e(t)$$

Equation (10):

$$\widehat{\mathbf{h}}(t) = \widehat{\mathbf{h}}_N(t) \otimes \widehat{\mathbf{h}}_{N-1}(t) \otimes \cdots \otimes \widehat{\mathbf{h}}_2(t) \otimes \widehat{\mathbf{h}}_1(t)$$

Outputs:

$\widehat{y}(t),\ \widehat{\mathbf{h}}(t),\ \widehat{\mathbf{h}}_n(t),\ n = 1, 2, \ldots, N.$

4. Connection with Tensor-Based Adaptive Filters

The KF-T algorithm developed in the previous section represents a generalization of the Kalman filter for bilinear forms (i.e., the particular case $N = 2$) presented in [27]. Nevertheless, it is also connected with other tensor-based adaptive algorithms, as will be shown in this section.

In [40], the tensor-based RLS (RLS-T) algorithm was introduced in the context of multilinear forms. At first, the RLS-T algorithm with the forgetting factors equal to one [44] has a striking resemblance to the KF-T using $\sigma_{w_n}^2 = 0,\ n = 1, 2, \ldots, N$. However, the KF-T does not rely on any matrix inversion, which is not the case for the RLS-T algorithm. Moreover, the RLS-T depends on the correlation matrices of the input signals, while the KF-T is related to the correlation matrices of the misalignments. This is a more proper approach in system identification scenarios, as was outlined at the end of Section 2. Additionally, the RLS-T algorithm does not depend on the variance of the additive noise (i.e., the parameter σ_v^2), or on the uncertainties in the system (i.e., the parameters $\sigma_{w_n}^2,\ n = 1, 2, \ldots, N$), since $\mathbf{h}_n(t),\ n = 1, 2, \ldots, N$ are considered as deterministic in its derivation. Nevertheless, these specific parameters of the KF-T allow for better control of the algorithm. For example, large

values of $\sigma_{w_n}^2$, $n = 1, 2, \ldots, N$ are suitable when the uncertainties in the system are high, in which case a good tracking behavior of the algorithm is needed; usually, the price is a lower accuracy, i.e., a higher misalignment. On the other hand, lower values of these parameters lead to improved accuracy, but reduce the tracking capability.

An interesting connection can be established between the KF-T and the tensor-based NLMS (NLMS-T) algorithm [39,40], which was also developed in the framework of multilinear forms. Let us consider that the KF-T has started to converge. Consequently, in the steady-state of the algorithm, we may also consider that $\mathbf{R}_{\mathbf{m}_{\mathbf{h}_n}}(t)$, $n = 1, 2, \ldots, N$ tend to become diagonal matrices. This approximation is reasonable, taking into account that the misalignment of the individual coefficients tend to become uncorrelated. Hence, we can use

$$\mathbf{R}_{\mathbf{m}_{\mathbf{h}_n}}(t) \approx \sigma_{\mathbf{m}_{\mathbf{h}_n}}^2(t)\mathbf{I}_{L_n}, \tag{33}$$

where $\sigma_{\mathbf{m}_{\mathbf{h}_n}}^2(t)$, $n = 1, 2, \ldots, N$ represent the elements from the main diagonal of the respective matrices. Furthermore, using the notation

$$\delta_{\mathbf{h}_n}(t) = \frac{\sigma_v^2}{\sigma_{\mathbf{m}_{\mathbf{h}_n}}^2(t)}, \; n = 1, 2, \ldots, N, \tag{34}$$

the expressions of the Kalman gain vectors from (27)–(29) simplify to

$$\widetilde{\mathbf{k}}_{\widehat{\mathbf{h}}_1}(t) = \frac{\mathbf{x}_{\widehat{\mathbf{h}}_2\widehat{\mathbf{h}}_3\ldots\widehat{\mathbf{h}}_N}(t)}{\mathbf{x}^T_{\widehat{\mathbf{h}}_2\widehat{\mathbf{h}}_3\ldots\widehat{\mathbf{h}}_N}(t)\mathbf{x}_{\widehat{\mathbf{h}}_2\widehat{\mathbf{h}}_3\ldots\widehat{\mathbf{h}}_N}(t) + \delta_{\mathbf{h}_1}(t)}, \tag{35}$$

$$\widetilde{\mathbf{k}}_{\widehat{\mathbf{h}}_2}(t) = \frac{\mathbf{x}_{\widehat{\mathbf{h}}_1\widehat{\mathbf{h}}_3\ldots\widehat{\mathbf{h}}_N}(t)}{\mathbf{x}^T_{\widehat{\mathbf{h}}_1\widehat{\mathbf{h}}_3\ldots\widehat{\mathbf{h}}_N}(t)\mathbf{x}_{\widehat{\mathbf{h}}_1\widehat{\mathbf{h}}_3\ldots\widehat{\mathbf{h}}_N}(t) + \delta_{\mathbf{h}_2}(t)}, \tag{36}$$

$$\vdots$$

$$\widetilde{\mathbf{k}}_{\widehat{\mathbf{h}}_N}(t) = \frac{\mathbf{x}_{\widehat{\mathbf{h}}_1\widehat{\mathbf{h}}_2\ldots\widehat{\mathbf{h}}_{N-1}}(t)}{\mathbf{x}^T_{\widehat{\mathbf{h}}_1\widehat{\mathbf{h}}_2\ldots\widehat{\mathbf{h}}_{N-1}}(t)\mathbf{x}_{\widehat{\mathbf{h}}_1\widehat{\mathbf{h}}_2\ldots\widehat{\mathbf{h}}_{N-1}}(t) + \delta_{\mathbf{h}_N}(t)}, \tag{37}$$

so that the updates (21)–(23) become

$$\begin{aligned}\widehat{\mathbf{h}}_1(t) &= \widehat{\mathbf{h}}_1(t-1) + \widetilde{\mathbf{k}}_{\widehat{\mathbf{h}}_1}(t)e(t) \\ &= \widehat{\mathbf{h}}_1(t-1) + \frac{\mathbf{x}_{\widehat{\mathbf{h}}_2\widehat{\mathbf{h}}_3\ldots\widehat{\mathbf{h}}_N}(t)e(t)}{\mathbf{x}^T_{\widehat{\mathbf{h}}_2\widehat{\mathbf{h}}_3\ldots\widehat{\mathbf{h}}_N}(t)\mathbf{x}_{\widehat{\mathbf{h}}_2\widehat{\mathbf{h}}_3\ldots\widehat{\mathbf{h}}_N}(t) + \delta_{\mathbf{h}_1}(t)},\end{aligned} \tag{38}$$

$$\begin{aligned}\widehat{\mathbf{h}}_2(t) &= \widehat{\mathbf{h}}_2(t-1) + \widetilde{\mathbf{k}}_{\widehat{\mathbf{h}}_2}(t)e(t) \\ &= \widehat{\mathbf{h}}_2(t-1) + \frac{\mathbf{x}_{\widehat{\mathbf{h}}_1\widehat{\mathbf{h}}_3\ldots\widehat{\mathbf{h}}_N}(t)e(t)}{\mathbf{x}^T_{\widehat{\mathbf{h}}_1\widehat{\mathbf{h}}_3\ldots\widehat{\mathbf{h}}_N}(t)\mathbf{x}_{\widehat{\mathbf{h}}_1\widehat{\mathbf{h}}_3\ldots\widehat{\mathbf{h}}_N}(t) + \delta_{\mathbf{h}_2}(t)},\end{aligned} \tag{39}$$

$$\vdots$$

$$\begin{aligned}\widehat{\mathbf{h}}_N(t) &= \widehat{\mathbf{h}}_N(t-1) + \widetilde{\mathbf{k}}_{\widehat{\mathbf{h}}_N}(t)e(t) \\ &= \widehat{\mathbf{h}}_N(t-1) + \frac{\mathbf{x}_{\widehat{\mathbf{h}}_1\widehat{\mathbf{h}}_2\ldots\widehat{\mathbf{h}}_{N-1}}(t)e(t)}{\mathbf{x}^T_{\widehat{\mathbf{h}}_1\widehat{\mathbf{h}}_2\ldots\widehat{\mathbf{h}}_{N-1}}(t)\mathbf{x}_{\widehat{\mathbf{h}}_1\widehat{\mathbf{h}}_2\ldots\widehat{\mathbf{h}}_{N-1}}(t) + \delta_{\mathbf{h}_N}(t)}.\end{aligned} \tag{40}$$

These updates are specific to a tensor-based variable-regularized NLMS algorithm tailored for multilinear forms, namely VR-NLMS-T. Such an algorithm is defined by the error signals from (11)–(13) and the updates (38)–(40). In this case, the control parameters are grouped into the variable regularization factors $\delta_{\mathbf{h}_n}(t)$, $n = 1, 2, \ldots, N$.

The VR-NLMS-T algorithm represents a simplified form of the KF-T. Nevertheless, due to the assumptions (33), the convergence features of the KF-T outperform the VR-NLMS-T algorithm. Starting from the VR-NLMS-T algorithm, we can show the connections with other tensor-based algorithms. For example, using constant values for the regularization parameters in the denominators of (38)–(40), i.e., $\delta_{\mathbf{h}_n}(t) = \delta_n,\ n = 1,2,\ldots,N$, while multiplying them with some normalized step-sizes $0 < \alpha_n \leq 1$ $(n = 1,2,\ldots,N)$ the nominators in (38)–(40), we obtain the NLMS-T algorithm [39,40]. Furthermore, replacing $\mathbf{x}^T_{\widehat{\mathbf{h}}_1\widehat{\mathbf{h}}_2\ldots\widehat{\mathbf{h}}_{n-1}\widehat{\mathbf{h}}_{n+1}\ldots\widehat{\mathbf{h}}_N}(t)\mathbf{x}_{\widehat{\mathbf{h}}_1\widehat{\mathbf{h}}_2\ldots\widehat{\mathbf{h}}_{n-1}\widehat{\mathbf{h}}_{n+1}\ldots\widehat{\mathbf{h}}_N}(t),\ n = 1,2,\ldots,N$ by $\sum_{n=1}^{N}\mathbf{x}^T_{\widehat{\mathbf{h}}_1\widehat{\mathbf{h}}_2\ldots\widehat{\mathbf{h}}_{n-1}\widehat{\mathbf{h}}_{n+1}\ldots\widehat{\mathbf{h}}_N}(t)\mathbf{x}_{\widehat{\mathbf{h}}_1\widehat{\mathbf{h}}_2\ldots\widehat{\mathbf{h}}_{n-1}\widehat{\mathbf{h}}_{n+1}\ldots\widehat{\mathbf{h}}_N}(t)$, the tensor LMS algorithm [23] is obtained.

5. Simulation Results

In this section, simulation results are provided in order to support the performance of the proposed KF-T and the main theoretical findings related to this algorithm. The goal of this analysis is fourfold, as follows. First, we evaluate the influence of the uncertainty parameters (i.e., $\sigma^2_{w_n},\ n = 1,2,\ldots,N$) on the performance of the KF-T. Second, we outline the connection between the proposed Kalman-based algorithm and its tensor-based counterparts, as discussed in Section 4. Third, we assess the performance of the KF-T as compared to the conventional Kalman filter. Finally, we analyze the behavior of the KF-T in a more general framework, for the identification of nonseparable systems.

The experiments are performed in the context of the multilinear framework presented in Section 2. The order of the system is $N = 4$ and the input signals that form the tensor $\boldsymbol{\mathcal{X}}(t)$ are AR(1) processes in most of the experiments. These inputs are obtained by filtering white Gaussian noises through an AR(1) model with a pole at 0.99. Such a high correlation degree (due to the pole close to 1) represents a challenge for most adaptive filters [20–22], in terms of their convergence features. In the last two experiments, we also used real-world speech sequences (corrupted by background noise) as input signals, which are also challenging due to their nonstationary character.

The MISO system used in simulations is characterized by four individual channels (i.e., $N = 4$), where their impulse responses are chosen as follows. The impulse response $\mathbf{h}_1(t)$ contains the first $L_1 = 16$ coefficients of the first network echo path from the ITU-T G168 Recommendation [45] (which is a standard for digital network echo cancellers). The impulse response $\mathbf{h}_2(t)$ is randomly generated (with Gaussian distribution), using the length $L_2 = 8$. The lengths of the other two impulse responses, i.e., $\mathbf{h}_3(t)$ and $\mathbf{h}_4(t)$, are set to $L_3 = 4$ and $L_4 = 2$, respectively. Their coefficients follow an exponential decay based on the rule $a_k^{l_k-1},\ l_k = 1,\ldots,L_k,\ k = 3,4$, using $a_3 = 0.8$ and $a_4 = 0.3$, respectively. The length of the global impulse response $\mathbf{h}(t)$ is $L = \prod_{n=1}^{N} L_n$, which, in our case, results in $L = 1024$. Such a length could be prohibiting in terms of implementation, especially for the Kalman-based and RLS-based adaptive filters. On the other hand, the tensor-based algorithms combine the solutions of N shorter filters of length $L_n,\ n = 1,2,\ldots,N$, which is much more advantageous, since, usually, $L_n \ll L$ (as in the current setup).

The output signal $y(n)$ from (4) is corrupted by an additive white Gaussian noise, $v(n)$; its variance is set to $\sigma_v^2 = 0.01$. We assume that this parameter is available in simulations. In practice, it can be estimated in several ways, e.g., [46,47]. Nevertheless, the influence of these different estimates on the performance of the proposed KF-T is beyond the scope of this paper. Finally, the reference signal results based on (7).

Two performance measures are used in simulations. First, the identification of the global impulse response is evaluated based on the normalized misalignment (NM), in dB, which is computed as

$$\mathrm{NM}\left[\mathbf{h}(t),\widehat{\mathbf{h}}(t)\right] = 10\log_{10}\left[\frac{\left\|\mathbf{h}(t)-\widehat{\mathbf{h}}(t)\right\|^2}{\|\mathbf{h}(t)\|^2}\right], \tag{41}$$

where $\|\cdot\|$ stands for the Euclidean norm. Second, since the identification of the individual impulse responses is influenced by the scaling factors [according to the discussion related to (6)], a proper performance measure is the normalized projection misalignment (NPM) [48], which is evaluated as

$$\mathrm{NPM}\left[\mathbf{h}_n(t), \widehat{\mathbf{h}}_n(t)\right] = 10\log_{10}\left\{1 - \left[\frac{\mathbf{h}_n^T(t)\widehat{\mathbf{h}}_n(t)}{\|\mathbf{h}_n(t)\|\left\|\widehat{\mathbf{h}}_n(t)\right\|}\right]^2\right\}, \tag{42}$$

for $n = 1, 2, \ldots, N$.

In the first set of experiments, we evaluate the impact of the uncertainty parameters on the performance of the proposed KF-T. The values of $\sigma^2_{w_n}$, $n = 1, 2, \ldots, N$ are subject to a compromise between the main performance criteria, i.e., fast convergence/tracking versus low misadjustment (i.e., good accuracy). The best accuracy of the solution is obtained for $\sigma^2_{w_n} = 0$, $n = 1, 2, \ldots, N$, i.e., when there are no uncertainties in the system. Such a setup is suitable in stationary environments. As can be seen in Figures 1 and 2 (in terms of the NM and NPM, respectively), the lower the values of $\sigma^2_{w_n}$, $n = 1, 2, \ldots, N$, the lower the misalignment of KF-T. On the other hand, larger values of these parameters improve the tracking capability of the algorithm, but achieve a higher misadjustment, reducing the accuracy of the solution. This behavior is supported in Figure 3, where an abrupt change in the system is considered in the middle of the experiments, by changing the sign of the coefficients of $\mathbf{h}_1(t)$. Since the initial convergence rate is not relevant in this case, and for a better visualization, we focused only on the tracking behavior in Figure 3. Despite having the lowest misalignment (i.e., the best accuracy), the KF-T using $\sigma^2_{w_n} = 0$, $n = 1, 2, \ldots, N$ has the slowest tracking capability. Nevertheless, larger values of $\sigma^2_{w_n}$, $n = 1, 2, \ldots, N$ lead to a significantly better performance in terms of tracking, while slightly sacrificing the accuracy of the solution (i.e., achieving a slightly higher misalignment level).

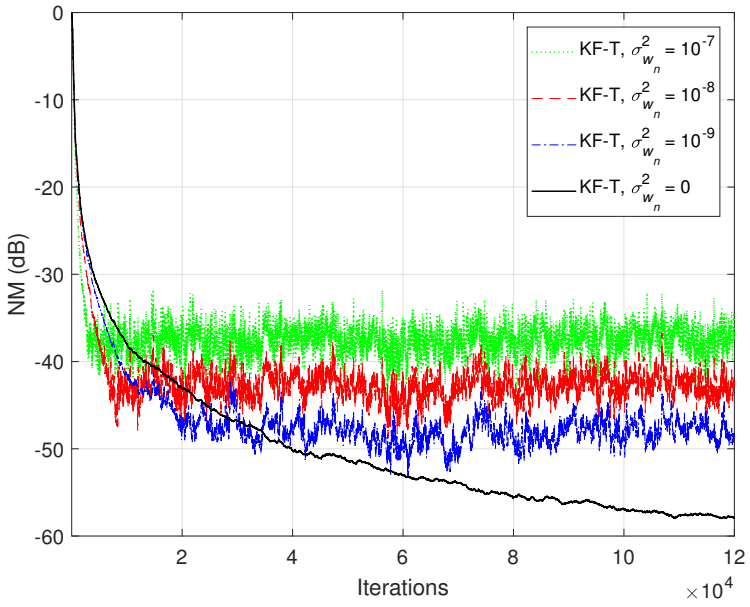


Figure 1. Normalized misalignment (NM) of the KF-T using different values of $\sigma^2_{w_n}$, $n = 1, 2, \ldots, N$, for the identification of the global impulse response, $\mathbf{h}(t)$.

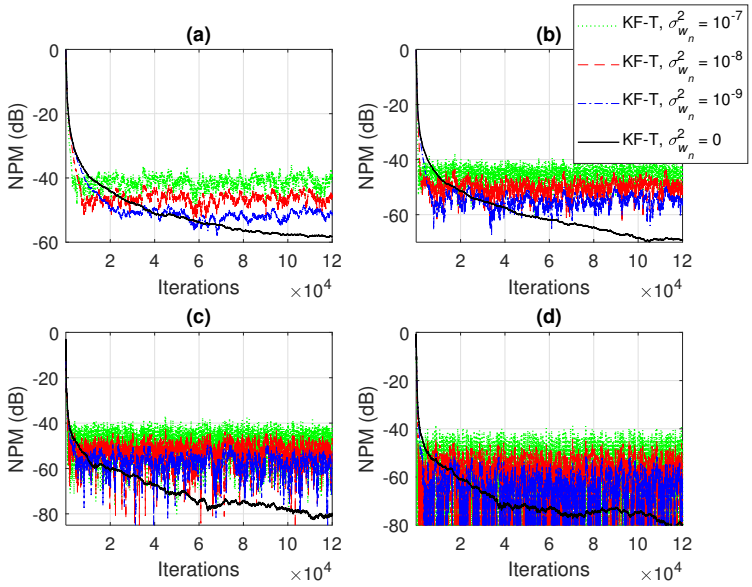


Figure 2. Normalized projection misalignment (NPM) of the KF-T using different values of $\sigma^2_{\bar{w}_n}$, $n = 1, 2, \ldots, N$, for the identification of the individual impulse responses, $\mathbf{h}_n(t)$, $n = 1, 2, \ldots, N$. (**a**) NPM $\left[\mathbf{h}_1(t), \widehat{\mathbf{h}}_1(t)\right]$, (**b**) NPM $\left[\mathbf{h}_2(t), \widehat{\mathbf{h}}_2(t)\right]$, (**c**) NPM $\left[\mathbf{h}_3(t), \widehat{\mathbf{h}}_3(t)\right]$, and (**d**) NPM $\left[\mathbf{h}_4(t), \widehat{\mathbf{h}}_4(t)\right]$.

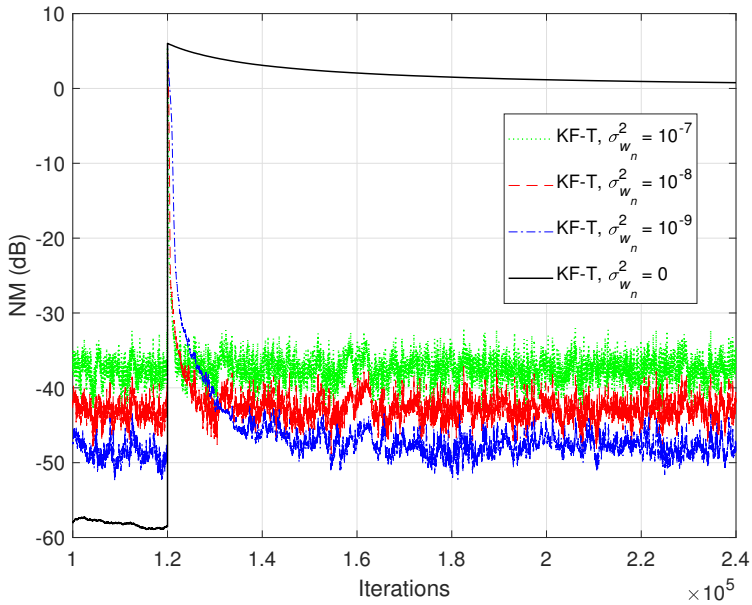


Figure 3. Normalized misalignment (NM) of the KF-T using different values of $\sigma^2_{\bar{w}_n}$, $n = 1, 2, \ldots, N$, for the identification of the global impulse response, $\mathbf{h}(t)$. The first individual impulse response changes from $\mathbf{h}_1(t)$ to $-\mathbf{h}_1(t)$ in the middle of simulation.

The connections between the proposed KF-T and other tensor-based adaptive algorithms were shown in Section 4. For example, the KF-T with $\sigma^2_{\bar{w}_n} = 0$, $n = 1, 2, \ldots, N$ resembles the RLS-T algorithm [40] using the forgetting factors $\lambda_n = 1$, $n = 1, 2, \ldots, N$. This aspect is supported in Figures 4 and 5, in terms of the NM and NPM, respectively. Both algorithms achieve similar initial convergence rates. However, the KF-T reaches a lower misalignment level and outperforms the RLS-T algorithm, thus supporting the discussion from Section 4. Moreover, in Figure 4, we also introduce comparisons with the solutions provided by the conventional and iterative Wiener filters (WFs) [38]. Both tensor-based algorithms outperform the conventional WF in terms of the accuracy of their solutions (i.e., lower misalignment). The KF-T converges (faster than the RLS-T) to the solution obtained

by the iterative WF for multilinear forms [38]. Nevertheless, as mentioned in Section 2 [in the discussion that follows (8)], the KF-T overcomes the inherent limitations of the iterative WF (e.g., time-invariant framework, estimation of the statistics, and the matrix inversion operation).

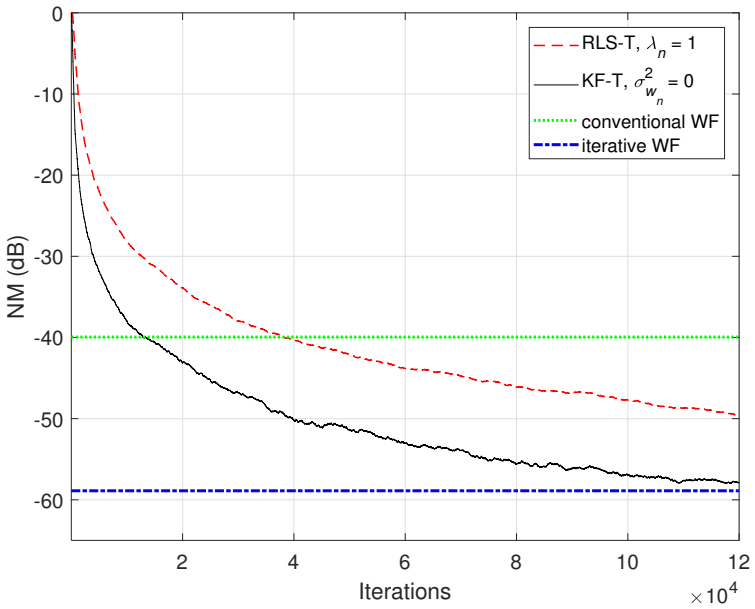


Figure 4. Normalized misalignment (NM) of the RLS-T algorithm using $\lambda_n = 1$, $n = 1, 2, \ldots, N$, the KF-T using $\sigma^2_{w_n} = 0$, $n = 1, 2, \ldots, N$, and the conventional and iterative WFs [38], for the identification of the global impulse response, $\mathbf{h}(t)$.

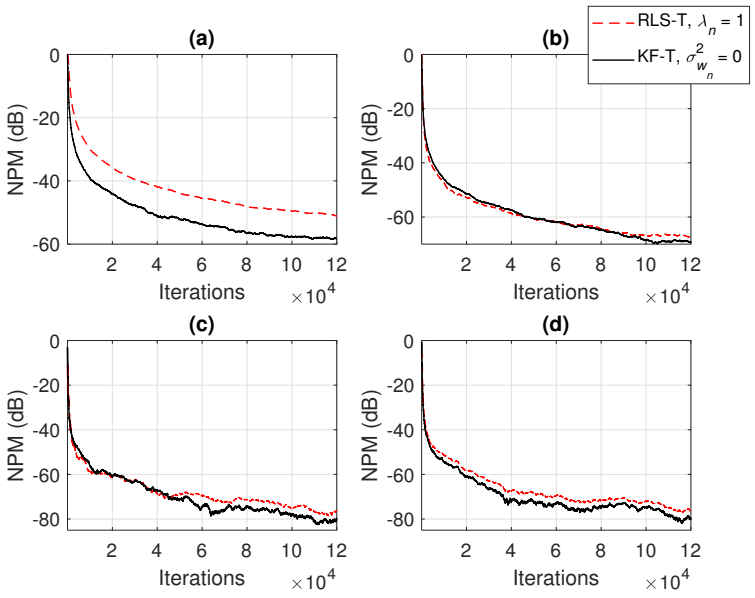


Figure 5. Normalized projection misalignment (NPM) of the RLS-T algorithm using $\lambda_n = 1$, $n = 1, 2, \ldots, N$ and the KF-T using $\sigma^2_{w_n} = 0$, $n = 1, 2, \ldots, N$, for the identification of the individual impulse responses, $\mathbf{h}_n(t)$, $n = 1, 2, \ldots, N$. (**a**) NPM$\left[\mathbf{h}_1(t), \widehat{\mathbf{h}}_1(t)\right]$, (**b**) NPM$\left[\mathbf{h}_2(t), \widehat{\mathbf{h}}_2(t)\right]$, (**c**) NPM$\left[\mathbf{h}_3(t), \widehat{\mathbf{h}}_3(t)\right]$, and (**d**) NPM$\left[\mathbf{h}_4(t), \widehat{\mathbf{h}}_4(t)\right]$.

Based on the assumption from (33), it was also shown in Section 4 that the KF-T behaves like a VR-NLMS-T algorithm, which is defined by the relations (34)–(40). This behavior is supported in Figures 6–8, when using different values of $\sigma^2_{w_n}$, $n = 1, 2, \ldots, N$.

The KF-T and the VR-NLMS-T reach the same misalignment level for the same values of the uncertainty parameters. On the other hand, due to the assumption in (33), the VR-NLMS-T algorithm experiences a slower convergence rate as compared to the KF-T. In fact, as outlined in Section 4, the VR-NLMS-T algorithm represents a simplified version of the KF-T, with a lower computational complexity (but paying a price in terms of the convergence features).

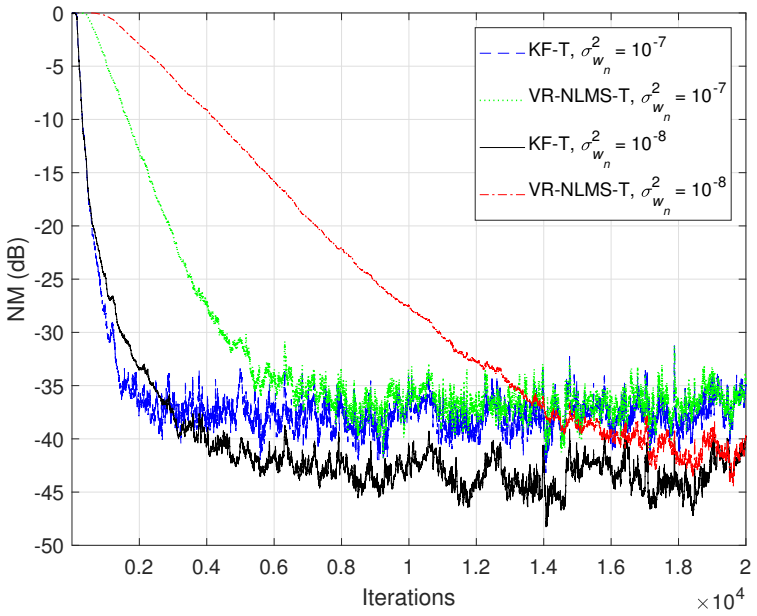


Figure 6. Normalized misalignment (NM) of the KF-T and VR-NLMS-T algorithm (using different values of $\sigma^2_{w_n}$, $n = 1, 2, \ldots, N$), for the identification of the global impulse response, $\mathbf{h}(t)$.

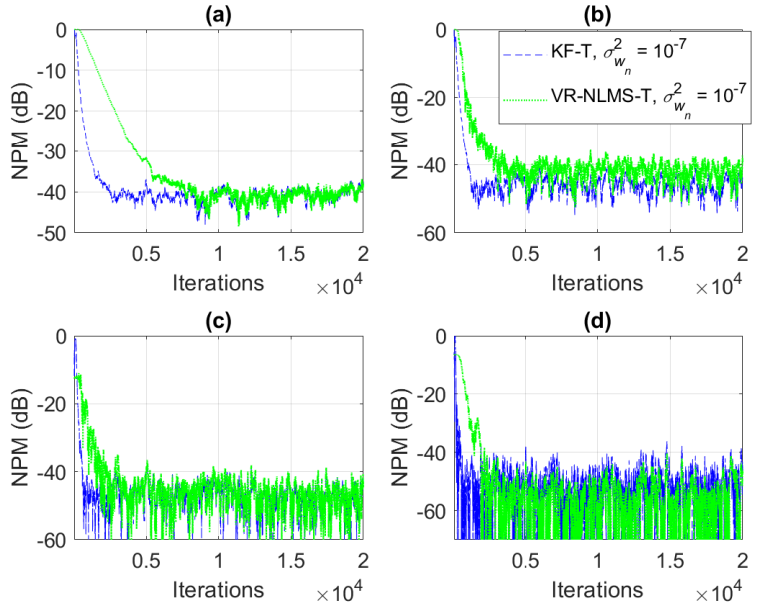


Figure 7. Normalized projection misalignment (NPM) of the KF-T and VR-NLMS-T algorithm using $\sigma^2_{w_n} = 10^{-7}$, $n = 1, 2, \ldots, N$, for the identification of the individual impulse responses, $\mathbf{h}_n(t)$, $n = 1, 2, \ldots, n$. (**a**) NPM$\left[\mathbf{h}_1(t), \widehat{\mathbf{h}}_1(t)\right]$, (**b**) NPM$\left[\mathbf{h}_2(t), \widehat{\mathbf{h}}_2(t)\right]$, (**c**) NPM$\left[\mathbf{h}_3(t), \widehat{\mathbf{h}}_3(t)\right]$, and (**d**) NPM$\left[\mathbf{h}_4(t), \widehat{\mathbf{h}}_4(t)\right]$.

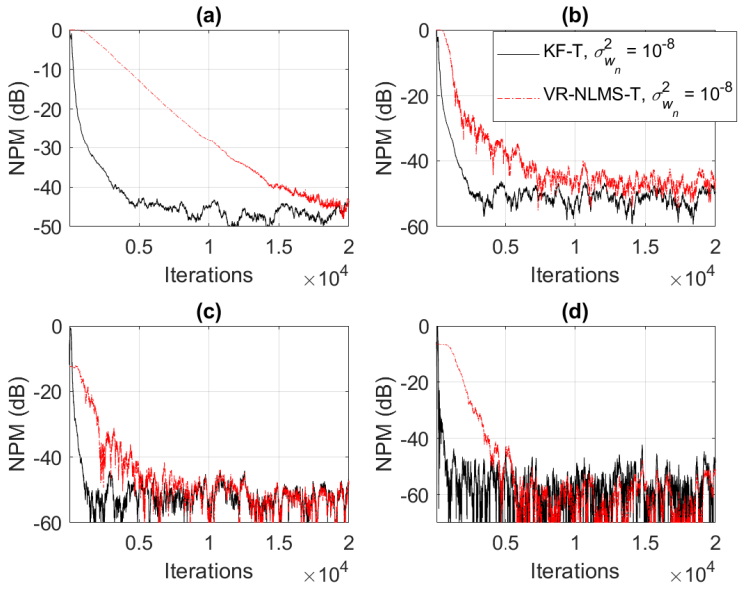


Figure 8. Normalized projection misalignment (NPM) of the KF-T and VR-NLMS-T algorithm using $\sigma^2_{w_n} = 10^{-8}$, $n = 1, 2, \ldots, N$, for the identification of the individual impulse responses, $\mathbf{h}_n(t)$, $n = 1, 2, \ldots, n$. (**a**) NPM$\left[\mathbf{h}_1(t), \widehat{\mathbf{h}}_1(t)\right]$, (**b**) NPM$\left[\mathbf{h}_2(t), \widehat{\mathbf{h}}_2(t)\right]$, (**c**) NPM$\left[\mathbf{h}_3(t), \widehat{\mathbf{h}}_3(t)\right]$, and (**d**) NPM$\left[\mathbf{h}_4(t), \widehat{\mathbf{h}}_4(t)\right]$.

In Figures 9 and 10, the NM and NPM performance of the proposed KF-T are compared to its tensor-based counterparts, i.e., the NLMS-T and the RLS-T algorithms [40]. The specific parameters of the KF-T are set to $\sigma^2_{w_n} = 10^{-9}$, $n = 1, 2, \ldots, N$, while the RLS-T algorithm uses the forgetting factors $\lambda_n = 0.999$, $n = 1, 2, \ldots, N$, in order to target a similar convergence behavior. The normalized step-sizes of the NLMS-T algorithm are set to $\alpha_n = 0.25$, $n = 1, 2, \ldots, N$, which represent the fastest convergence mode [39,40]. The NLMS-T and RLS-T algorithms reach similar misalignment levels, while the KF-T outperforms its counterparts in terms of both convergence rate and misalignment. As expected, the RLS-T is significantly faster (in terms of the convergence rate) as compared to the NLMS-T algorithm. Nevertheless, the KF-T provides an initial convergence rate that is slightly better as compared to the RLS-T algorithm, while also achieving a lower misalignment level (i.e., a better accuracy of the solution).

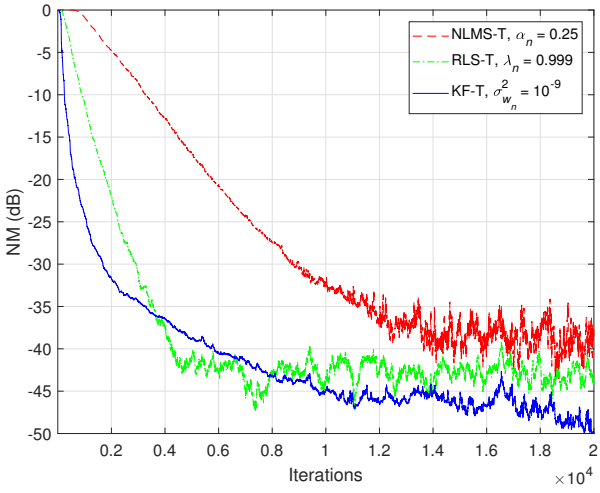


Figure 9. Normalized misalignment (NM) of the NLMS-T algorithm using $\alpha_n = 0.25$, $n = 1, 2, \ldots, N$, the RLS-T algorithm using $\lambda_n = 0.999$, $n = 1, 2, \ldots, N$, and the KF-T using $\sigma^2_{w_n} = 10^{-9}$, $n = 1, 2, \ldots, N$, for the identification of the global impulse response, $\mathbf{h}(t)$.

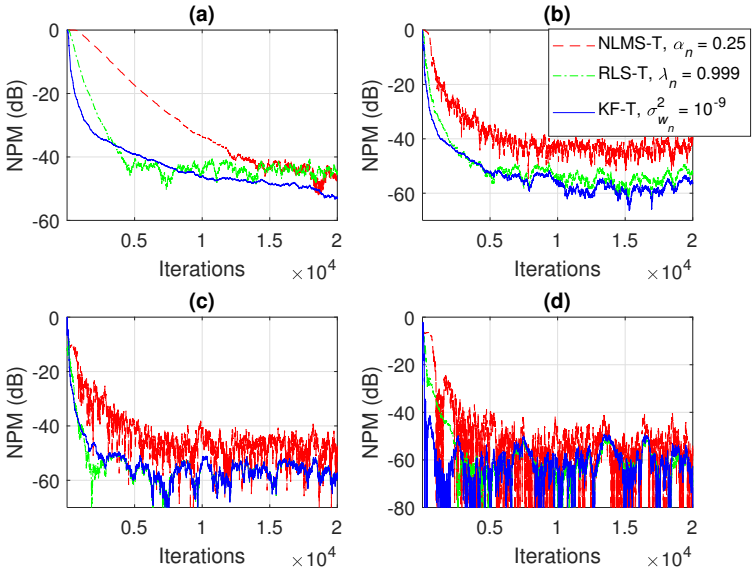


Figure 10. Normalized projection misalignment (NPM) of the NLMS-T algorithm using $\alpha_n = 0.25$, $n = 1, 2, \ldots, N$, the RLS-T algorithm using $\lambda_n = 0.999$, $n = 1, 2, \ldots, N$, and the KF-T using $\sigma^2_{w_n} = 10^{-9}$, $n = 1, 2, \ldots, N$, for the identification of the individual impulse responses, $\mathbf{h}_n(t)$, $n = 1, 2, \ldots, n$. (**a**) NPM$\left[\mathbf{h}_1(t), \widehat{\mathbf{h}}_1(t)\right]$, (**b**) NPM$\left[\mathbf{h}_2(t), \widehat{\mathbf{h}}_2(t)\right]$, (**c**) NPM$\left[\mathbf{h}_3(t), \widehat{\mathbf{h}}_3(t)\right]$, and (**d**) NPM$\left[\mathbf{h}_4(t), \widehat{\mathbf{h}}_4(t)\right]$.

The conventional Kalman filter (KF) can also be used for the identification of the global impulse response, $\mathbf{h}(t)$, as explained in the end of Section 3. In this case, there is a single adaptive filter (of length L) that has to be updated, while the overall computational complexity is proportional to $\mathcal{O}(L^2)$. Due to the large number of coefficients, dealing with such a long adaptive filter raises significant challenges in terms of the complexity, convergence, and accuracy of the solution. On the other hand, the proposed KF-T can obtain the estimate of the global impulse response by combining the solutions of much shorter adaptive filters of lengths L_n, $n = 1, 2, \ldots, N$, with $\prod_{n=1}^{N} L_n = L$. Therefore, the expected gain is twofold, in terms of both performance and complexity. This is supported in Figure 11, where the performance of the proposed KF-T is compared to the conventional KF. The specific parameters are set to target the best accuracy of the solution, i.e., $\sigma^2_{w_n} = 0$, $n = 1, 2, \ldots, N$ in case of the KF-T, while the conventional KF uses the same null value for its uncertainty parameter. The proposed KF-T outperforms the conventional KF in terms of accuracy, achieving a significantly lower misalignment level. Moreover, the complexity of the KF-T is proportional to $\sum_{n=1}^{N} \mathcal{O}(L_n^2)$, which is much more advantageous compared to the conventional KF, especially for $L_n \ll L$, with $n = 1, 2, \ldots, N$. For example, related to the experiment given in Figure 11, we could mention that the simulation time (using MATLAB R2018b) of the proposed KF-T was less than one minute, while the conventional KF took almost one hour to reach the final result. The experiment was performed on an Asus GL552VX device (Windows 10 OS), having an Intel Core i7-6700HQ CPU@2.60 GHz, with 4 Cores, 8 Logical Processors, and 16 GB of RAM.

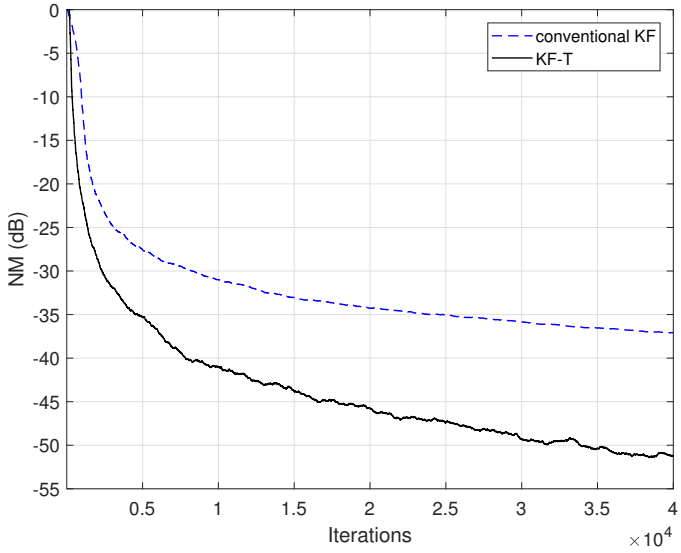


Figure 11. Normalized misalignment (NM) of the KF-T using $\sigma^2_{w_n} = 0,\ n = 1, 2, \ldots, N$ and the conventional KF using the same value of its uncertainty parameter, for the identification of the global impulse response, $\mathbf{h}(t)$.

In the last set of experiments, we focus on a more challenging scenario, when the global impulse response of the system is not separable. In this case, we target the identification of $\mathbf{h}(t) + \mathbf{u}(t)$, where $\mathbf{h}(t)$ is specified in the beginning of this section [i.e., $\mathbf{h}(t) = \mathbf{h}_4(t) \otimes \mathbf{h}_3(t) \otimes \mathbf{h}_2(t) \otimes \mathbf{h}_1(t)$], while $\mathbf{u}(t)$ is randomly generated, with Gaussian distribution, and its variance is set to $v\|\mathbf{h}(t)\|^2/L$ (using different values of v). Clearly, the higher the value of v, the more challenging the decomposition of the global impulse response. Since the KF-T is based on the decomposition in (10), it cannot model the noisy part $\mathbf{u}(t)$. Nevertheless, as can be seen in Figures 12 and 13 (in terms of the NM and NPM, respectively), the KF-T is able to achieve a reasonable attenuation of the misalignment (i.e., a good accuracy of the estimate) even for larger values of v. Consequently, the proposed KF-T has uses beyond the identification of rank-1 tensors, when the global impulse response contains a dominant separable (i.e., decomposable) part.

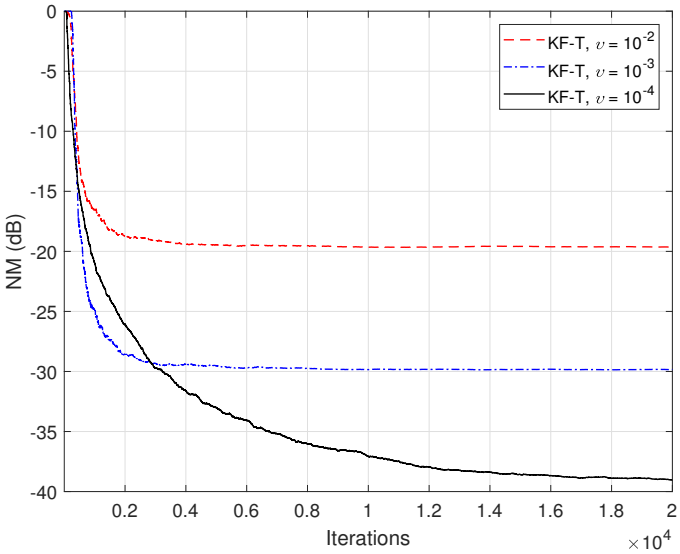


Figure 12. Normalized misalignment (NM) of the KF-T using $\sigma^2_{w_n} = 0,\ n = 1, 2, \ldots, N$ for the identification of the global impulse response, $\mathbf{h}(t) + \mathbf{u}(t)$, where $\mathbf{u}(t)$ is randomly generated (Gaussian distribution), with variance $v\|\mathbf{h}(t)\|^2/L$, using different values of v.

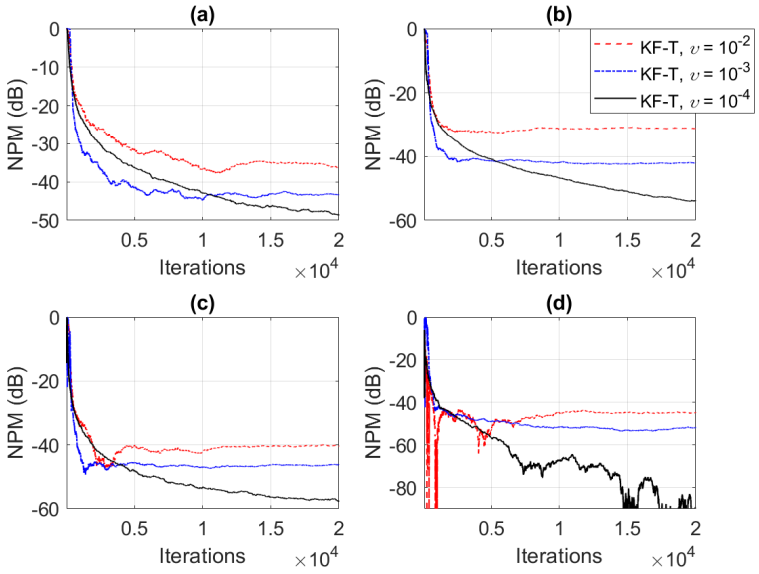


Figure 13. Normalized projection misalignment (NPM) of the KF-T using $\sigma^2_{w_n} = 0,\ n = 1,2,\ldots,N$ for the identification of the individual impulse responses, $\mathbf{h}_n(t),\ n = 1,2,\ldots,N$, which result from the decomposition of $\mathbf{h}(t) + \mathbf{u}(t)$, where $\mathbf{u}(t)$ is randomly generated (Gaussian distribution), with variance $v\|\mathbf{h}(t)\|^2/L$, using different values of v. (**a**) NPM$\left[\mathbf{h}_1(t), \widehat{\mathbf{h}}_1(t)\right]$, (**b**) NPM$\left[\mathbf{h}_2(t), \widehat{\mathbf{h}}_2(t)\right]$, (**c**) NPM$\left[\mathbf{h}_3(t), \widehat{\mathbf{h}}_3(t)\right]$, and (**d**) NPM$\left[\mathbf{h}_4(t), \widehat{\mathbf{h}}_4(t)\right]$.

The previous experiment is repeated in Figure 14, but using real-world speech sequences (corrupted by background noise) as input signals. Due to the nonstationary nature and highly correlated character of the speech signals, the performance of any adaptive algorithm is influenced in such a scenario. This is also the case for the KF-T, which pays with a slower convergence rate, as compared to the case analyzed in Figure 12. Nevertheless, this performance criterion can be improved by using higher values for the uncertainty parameters, $\sigma^2_{w_n},\ n = 1,2,\ldots,N$. This is further supported in Figure 15, where we can notice a higher convergence rate of the KF-T when increasing the values of $\sigma^2_{w_n}$, while causing a slight increase in the misalignment level.

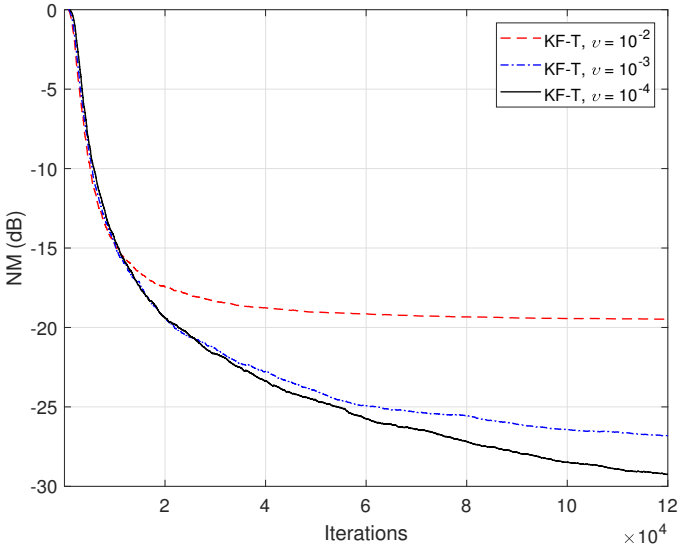


Figure 14. Normalized misalignment (NM) of the KF-T using $\sigma^2_{w_n} = 0,\ n = 1,2,\ldots,N$ for the identification of the global impulse response, $\mathbf{h}(t) + \mathbf{u}(t)$, where $\mathbf{u}(t)$ is randomly generated (Gaussian distribution), with variance $v\|\mathbf{h}(t)\|^2/L$, using different values of v. The input signals are speech sequences.

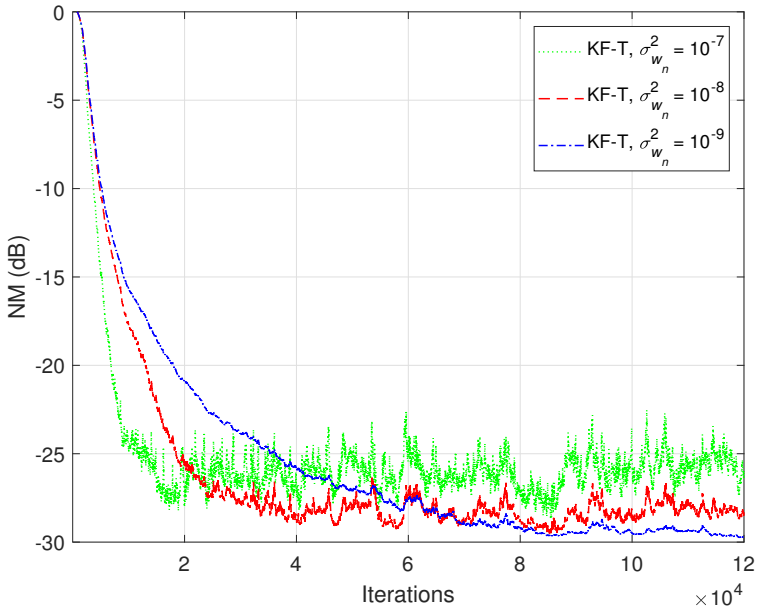


Figure 15. Normalized misalignment (NM) of the KF-T using different values of $\sigma^2_{w_n}$, $n = 1, 2, \ldots, N$, for the identification of the global impulse response, $\mathbf{h}(t) + \mathbf{u}(t)$, where $\mathbf{u}(t)$ is randomly generated (Gaussian distribution), with variance $v\|\mathbf{h}(t)\|^2/L$, using $v = 10^{-3}$. The input signals are speech sequences.

6. Conclusions

In this paper, we have presented a tensorial Kalman filter tailored to the identification of multilinear forms. The solution was developed in the framework of a MISO system, while the identification problem was reformulated based on linearly separable systems modeled as rank-1 tensors. We have also shown how the resulting KF-T algorithm is connected to the main categories of tensor-based adaptive filters, i.e., the NLMS-T and the RLS-T algorithms. In this context, the specific uncertainty parameters of the KF-T allow for better control of this, as compared to its counterparts. The simulation results indicated the good performance features of the KF-T and also supported the discussion related to its connection with other tensorial algorithms. Future works will focus on finding a more practical way to evaluate the uncertainty parameters, e.g., following a similar approach to variable adaptation factors, which act as time-dependent parameters. In addition, we aim to extend the decomposition-based technique to the identification of nonseparable systems (which could be modeled as higher rank tensors), in order to develop a more general and efficient version of the tensorial Kalman filter.

Author Contributions: Conceptualization, L.-M.D.; methodology, C.P.; validation, J.B.; software, C.-L.S.; investigation, C.-C.O.; formal analysis, S.C. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by two grants of the Romanian Ministry of Education and Research, CNCS—UEFISCDI, project number PN-III-P1-1.1-PD-2019-0340 and project number PN-III-P1-1.1-TE-2019-0529, within PNCDI III.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

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

References

1. Gesbert, D.; Duhamel, P. Robust blind joint data/channel estimation based on bilinear optimization. In Proceedings of the IEEE Workshop on Statistical Signal and Array Processing (WSSAP), Corfu, Greece, 24–26 June 1996; pp. 168–171.
2. Ribeiro, L.N.; Schwarz, S.; Rupp, M.; de Almeida, A.L.F.; Mota, J.C.M. A low-complexity equalizer for massive MIMO systems based on array separability. In Proceedings of the European Signal Processing Conference (EUSIPCO), Kos, Greece, 28 August–2 September 2017; pp. 2522–2526.
3. Stenger, A.; Kellermann, W. Adaptation of a memoryless preprocessor for nonlinear acoustic echo cancelling. *Signal Process.* **2000**, *80*, 1747–1760. [CrossRef]
4. Huang, Y.; Skoglund, J.; Luebs, A. Practically efficient nonlinear acoustic echo cancellers using cascaded block RLS and FLMS adaptive filters. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), New Orleans, LA, USA, 5–9 March 2017; pp. 596–600.
5. Cichocki, A.; Zdunek, R.; Pan, A.H.; Amari, S. *Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multiway Data Analysis and Blind Source Separation*; Wiley: Chichester, UK, 2009.
6. Boussé, M.; Debals, O.; Lathauwer, L.D. A tensor-based method for large-scale blind source separation using segmentation. *IEEE Trans. Signal Process.* **2017**, *65*, 346–358. [CrossRef]
7. Benesty, J.; Cohen, I.; Chen, J. *Array Processing–Kronecker Product Beamforming*; Springer: Cham, Switzerland, 2019.
8. Ribeiro, L.N.; de Almeida, A.L.F.; Mota, J.C.M. Separable linearly constrained minimum variance beamformers. *Signal Process.* **2019**, *158*, 15–25. [CrossRef]
9. Vasilescu, M.A.O.; Kim, E. Compositional hierarchical tensor factorization: Representing hierarchical intrinsic and extrinsic causal factors. In Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD), Anchorage, AK, USA, 4–8 August 2019; 9p.
10. Vasilescu, M.A.O.; Kim, E.; Zeng, X.S. CausalX: Causal eXplanations and block multilinear factor analysis. In Proceedings of the International Conference on Pattern Recognition (ICPR), Milan, Italy, 10–15 January 2020; 8p.
11. Vervliet, N.; Debals, O.; Sorber, L.; Lathauwer, L.D. Breaking the curse of dimensionality using decompositions of incomplete tensors: Tensor-based scientific computing in big data analysis. *IEEE Signal Process. Mag.* **2014**, *31*, 71–79. [CrossRef]
12. Cichocki, A.; Mandic, D.; Lathauwer, L.D.; Zhou, G.; Zhao, Q.; Caiafa, C.; Phan, A.H. Tensor decompositions for signal processing applications: From two-way to multiway component analysis. *IEEE Signal Process. Mag.* **2015**, *32*, 145–163. [CrossRef]
13. Sidiropoulos, N.; Lathauwer, L.D.; Fu, X.; Huang, K.; Papalexakis, E.; Faloutsos, C. Tensor decomposition for signal processing and machine learning. *IEEE Trans. Signal Process.* **2017**, *65*, 3551–3582. [CrossRef]
14. da Costa, M.N.; Favier, G.; Romano, J.M.T. Tensor modelling of MIMO communication systems with performance analysis and Kronecker receivers. *Signal Process.* **2018**, *145*, 304–316. [CrossRef]
15. Kuo, S.M.; Morgan, D.R. *Active Noise Control Systems: Algorithms and DSP Implementations*; Wiley: New York, NY, USA, 1996.
16. Li, J.; Stoica, P. *Robust Adaptive Beamforming*; Wiley: New York, NY, USA, 2006.
17. Benesty, J.; Chen, J.; Huang, Y. *Microphone Array Signal Processing*; Springer: Berlin/Heidelberg, Germany, 2008.
18. 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]
19. Dogariu, L.-M.; Ciochină, S.; Benesty, J.; Paleologu, C. System identification based on tensor decompositions: A trilinear approach. *Symmetry* **2019**, *11*, 556. [CrossRef]
20. Haykin, S. *Adaptive Filter Theory*, 4th ed.; Prentice-Hall: Upper Saddle River, NJ, USA, 2002.
21. Benesty, J.; Huang, Y. (Eds.) *Adaptive Signal Processing–Applications to Real-World Problems*; Springer: Berlin, Germany, 2003.
22. Diniz, P.S.R. *Adaptive Filtering: Algorithms and Practical Implementation*, 4th ed.; Springer: New York, NY, USA, 2013.
23. Rupp, M.; Schwarz, S. A tensor LMS algorithm. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), South Brisbane, Australia, 19–24 April 2015; pp. 3347–3351.
24. Paleologu, C.; Benesty, J.; Ciochină, S. Adaptive filtering for the identification of bilinear forms. *Digit. Signal Process.* **2018**, *75*, 153–167. [CrossRef]
25. Elisei-Iliescu, C.; Stanciu, C.; Paleologu, C.; Benesty, J.; Anghel, C.; Ciochină, S. Efficient recursive least-squares algorithms for the identification of bilinear forms. *Digit. Signal Process.* **2018**, *83*, 280–296. [CrossRef]
26. Elisei-Iliescu, C.; Dogariu, L.-M.; Paleologu, C.; Benesty, J.; Enescu, A.A.; Ciochină, S. A recursive least-squares algorithm for the identification of trilinear forms. *Algorithms* **2020**, *13*, 135. [CrossRef]
27. Dogariu, L.; Paleologu, C.; Ciochină, S.; Benesty, J.; Piantanida, P. Identification of bilinear forms with the Kalman filter. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, AB, Canada, 15–20 April 2018; pp. 4134–4138.
28. Dogariu, L.-M.; Ciochină, S.; Paleologu, C.; Benesty, J. A connection between the Kalman filter and an optimized LMS algorithm for bilinear forms. *Algorithms* **2018**, *11*, 211. [CrossRef]
29. Kalman, R.E. A new approach to linear filtering and prediction problems. *J. Basic Eng.* **1960**, *82*, 35–45. [CrossRef]
30. Sayed, A.H.; Kailath, T. A state-space approach to adaptive RLS filtering. *IEEE Signal Process. Mag.* **1994**, *11*, 18–60. [CrossRef]
31. Faragher, R. Understanding the basis of the Kalman filter via a simple and intuitive derivation. *IEEE Signal Process. Mag.* **2012**, *29*, 128–132. [CrossRef]

32. 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]
33. Vogt, H.; Enzner, G.; Sezgin, A. State-space adaptive nonlinear self-interference cancellation for full-duplex communication. *IEEE Trans. Signal Process.* **2019**, *67*, 2810–2825. [CrossRef]
34. Dogariu, L.-M.; Paleologu, C.; Benesty, J.; Ciochină, S. An efficient Kalman filter for the identification of low-rank systems. *Signal Process.* **2020**, *166*, 107239. [CrossRef]
35. Li, X.; Dong, H.; Han, S. Multiple linear regression with Kalman filter for predicting end prices of online auctions. In Proceedings of the IEEE DASC/PiCom/CBDCom/CyberSciTech, Calgary, AB, Canada, 17–22 August 2020; pp. 182–191.
36. Shan, C.; Zhou, W.; Yang, Y.; Jiang, Z. Multi-fading factor and updated monitoring strategy adaptive Kalman filter-based variational Bayesian. *Sensors* **2021**, *21*, 198. [CrossRef] [PubMed]
37. Zhang, H.; Zhang, X. Robust SCKF filtering method for MINS/GPS in-motion alignment. *Sensors* **2021**, *21*, 2597. [CrossRef]
38. Dogariu, L.-M.; Ciochină, S.; Paleologu, C.; Benesty, J.; Oprea, C. An iterative Wiener filter for the identification of multilinear forms. In Proceedings of the IEEE International Conference on Telecommunications and Signal Processing (TSP), Milan, Italy, 7–9 July 2020; pp. 193–197.
39. Dogariu, L.-M.; Paleologu, C.; Benesty, J.; Oprea, C.; Ciochină, S. LMS algorithms for multilinear forms. In Proceedings of the IEEE International Symposium on Electronics and Telecommunications (ISETC), Timişoara, Romania, 5–6 November 2020; 4p.
40. Dogariu, L.-M.; Stanciu, C.L.; Elisei-Iliescu, C.; Paleologu, C.; Benesty, J.; Ciochină, S. Tensor-based adaptive filtering algorithms. *Symmetry* **2021**, *13*, 481. [CrossRef]
41. Bertsekas, D.P. *Nonlinear Programming*, 2nd ed.; Athena Scientific: Belmont, MA, USA, 1999.
42. Loan, C.F.V. The ubiquitous Kronecker product. *J. Comput. Appl. Math.* **2000**, *123*, 85–100. [CrossRef]
43. Kay, S.M. *Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory*; Prentice Hall: Englewood Cliffs, NJ, USA, 1993.
44. Ciochină, S.; Paleologu, C.; Benesty, J.; Enescu, A.A. On the influence of the forgetting factor of the RLS adaptive filter in system identification. In Proceedings of the IEEE International Symposium on Signals, Circuits and Systems (ISSCS), Iaşi, Romania, 9–10 July 2009; pp. 205–208.
45. *Digital Network Echo Cancellers*; ITU-T Recommendations G.168; ITU: Geneva, Switzerland, 2002.
46. Iqbal, M.A.; Grant, S.L. Novel variable step size NLMS algorithm for echo cancellation. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Las Vegas, NV, USA, 31 March–4 April 2008; pp. 241–244.
47. Paleologu, C.; Ciochină, S.; Benesty, J. Double-talk robust VSSNLMS algorithm for under-modeling acoustic echo cancellation. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Las Vegas, NV, USA, 31 March–4 April 2008; pp. 245–248.
48. Morgan, D.R.; Benesty, J.; Sondhi, M.M. On the evaluation of estimated impulse responses. *IEEE Signal Process. Lett.* **1998**, *5*, 174–176. [CrossRef]