A Decomposition-Based Kalman Filter for the Identification of Acoustic Impulse Responses

Laura-Maria Dogariu
Dept. Telecommunications
University Politehnica of Bucharest
Bucharest, Romania
ldogariu@comm.pub.ro

Constantin Paleologu
Dept. Telecommunications
University Politehnica of Bucharest
Bucharest, Romania
pale@comm.pub.ro

Jacob Benesty
INRS-EMT
University of Quebec
Montreal, Canada
Jacob.Benesty@inrs.ca

Cristian-Lucian Stanciu
Dept. Telecommunications
University Politehnica of Bucharest
Bucharest, Romania
cristian@comm.pub.ro

Silviu Ciochină
Dept. Telecommunications
University Politehnica of Bucharest
Bucharest, Romania
silviu@comm.pub.ro

***Abstract*—The identification of acoustic impulse responses usually involves long length adaptive filters, with hundreds or even thousands of coefficients. This issue raises significant challenges in terms of both the computational complexity and convergence features. Recently, a decomposition-based solution using the Kronecker product and low-rank approximations was proposed in this context, by exploiting the intrinsic nature of the acoustic impulse responses. These systems are characterized by early reflections and late reverberation, each of these components having different characteristics that should be considered. In this paper, we propose a Kalman filter following this approach, which outperforms the previously developed solution based on the recursive least-squares algorithm. Simulations performed in the framework of acoustic echo cancellation support the performance features of the proposed algorithm.**

***Index Terms*—Acoustic impulse response, acoustic echo cancellation, Kalman filter, Kronecker product decomposition, system identification.**

I. Introduction

System identification problems are frequently addressed with adaptive filtering algorithms [1]. In this context, the identification of acoustic impulse responses is an important task in the framework of many applications related to the acoustic environment [2]. Nevertheless, the long length of the impulse response and its time-varying nature impose specific challenges for the adaptive filter.

A recent approach targets the identification of low-rank systems based on the nearest Kronecker product decomposition of the impulse response, in conjunction with bilinear forms [3]. As a result, a long length impulse response can be modeled using a combination of shorter filters, thus gaining in terms of both performance and complexity. Due to these features, the decomposition-based approach has been successfully applied in the context of different applications related to the acoustic environment, e.g., see [4]– [13] and the references therein.

This work was supported by two grants of the Ministry of Research, Innovation and Digitization, CNCS–UEFISCDI, projects PN-III-P4-PCE-2021-0438 and PN-III-P4-PCE-2021-0780, within PNCDI III.

More recently, in [14], this approach was further exploited by taking advantage of the specific structure of acoustic impulse responses, which are characterized by early reflections and late reverberation. These components of the impulse response own different characteristics, so that they can be processed separately in terms of their low-rank features. As a consequence, the iterative Wiener filter and recursive least-squares (RLS) algorithm developed in [14] outperform their counterparts from [3] and [4], respectively, which do not consider the intrinsic characteristics of the acoustic impulse responses. In this paper, we further develop a Kalman filter following this very recent decomposition-based approach. The proposed solution has a computational complexity similar to the RLS algorithm, but it performs better in terms of the convergence features.

In the following, Section II introduces the system model that involves the Kronecker product decomposition, while the proposed solution based on the Kalman filter is developed in Section III. Simulation results are presented in Section IV, in the framework of acoustic echo cancellation. Finally, the conclusions are outlined in Section V.

II. System Model Using the Kronecker Product

Let us consider the framework of a linear acoustic single-input single-output (SISO) system, with real-valued signals. In this context, the reference signal at the discrete-time index n is given by

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

where $\mathbf{h}(n)$ is the unknown acoustic impulse response of length L, the superscript ^{T} denotes the transpose operator, the column vector $\mathbf{x}(n) = \begin{bmatrix} x(n) & x(n-1) & \cdots & x(n-L+1) \end{bmatrix}^T$ contains the most recent L time samples of the zero-mean input signal $x(n)$, and $v(n)$ is the zero-mean additive noise [which is uncorrelated to $x(n)$], with variance $\sigma_v^2 = E\left[v^2(n)\right]$ and $E[\cdot]$

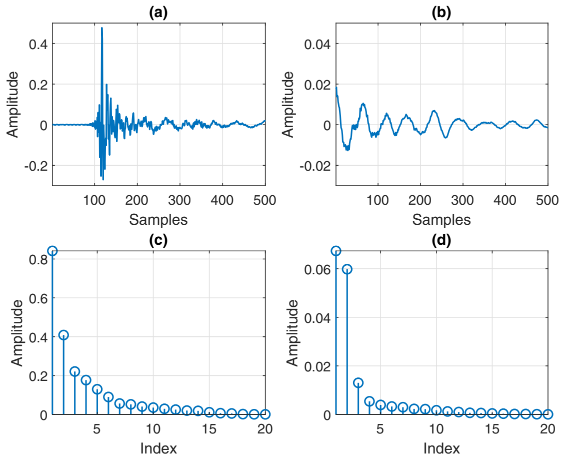

Fig. 1. The components of an acoustic impulse response (of length $L = 1000$) and the singular values of their corresponding matrices: (a) $\mathbf{h}_{\rm er}(n)$, (b) $\mathbf{h}_{\rm lr}(n)$, (c) singular values of $\mathbf{H}_{\rm er}(n)$, and (d) singular values of $\mathbf{H}_{\rm lr}(n)$. The decomposition setup is $L_{\rm er} = L_{\rm lr} = L/2$, with $L_{{\rm er},1} = L_{{\rm lr},1} = 25$ and $L_{{\rm er},2} = L_{{\rm lr},2} = 20$.

denoting mathematical expectation. In this context, the main objective is to estimate the L coefficients of the acoustic impulse response $\mathbf{h}(n)$.

As shown in [14], $\mathbf{h}(n)$ can be decomposed as

$$\mathbf{h}(n) = \begin{bmatrix} \mathbf{h}_{\rm er}^T(n) & \mathbf{h}_{\rm lr}^T(n) \end{bmatrix}^T, \tag{2}$$

where $\mathbf{h}_{\rm er}(n)$, of length $L_{\rm er}$, contains early reflections (plus the direct path) of the impulse response, and $\mathbf{h}_{\rm lr}(n)$, of length $L_{\rm lr}$, contains elements of late reverberation of $\mathbf{h}(n)$, with $L = L_{\rm er} + L_{\rm lr}$. These two components of the impulse response are of very different nature, as illustrated in Figs. 1(a) and (b), where a typical acoustic echo path is considered, with $L_{\rm er} = L_{\rm lr}$[1] and $L = 1000$. However, both components have low rank and can be processed separately. In order to simplify the notation, let $\star \in \{{\rm er}, {\rm lr}\}$. Also, without loss of generality, let us assume that $L_\star = L_{\star,1}L_{\star,2}$, with $L_{\star,1} \geq L_{\star,2}$. Thus, each component can be decomposed as $\mathbf{h}_\star(n) = \sum_{l=1}^{L_{\star,2}} \mathbf{h}_{\star,2}^l(n) \otimes \mathbf{h}_{\star,1}^l(n)$ [3], where $\otimes$ is the Kronecker product [15], and $\mathbf{h}_{\star,1}^l(n)$ and $\mathbf{h}_{\star,2}^l(n)$ are short impulse responses of lengths $L_{\star,1}$ and $L_{\star,2}$, respectively. Alternatively, we can express

$$\mathbf{h}_\star(n) = {\rm vec}\left[\mathbf{H}_{\star,1}(n)\mathbf{H}_{\star,2}^T(n)\right] = {\rm vec}\left[\mathbf{H}_\star(n)\right], \tag{3}$$

where ${\rm vec}(\cdot)$ is the vectorization operation (i.e., converting a matrix into a column vector, by stacking the columns of the matrix), $\mathbf{H}_{\star,1}(n) = \begin{bmatrix} \mathbf{h}_{\star,1}^1(n) & \mathbf{h}_{\star,1}^2(n) & \cdots & \mathbf{h}_{\star,1}^{L_{\star,2}}(n) \end{bmatrix}$ and $\mathbf{H}_{\star,2}(n) = \begin{bmatrix} \mathbf{h}_{\star,2}^1(n) & \mathbf{h}_{\star,2}^2(n) & \cdots & \mathbf{h}_{\star,2}^{L_{\star,2}}(n) \end{bmatrix}$ are matrices of sizes $L_{\star,1} \times L_{\star,2}$ and $L_{\star,2} \times L_{\star,2}$, respectively, and $\mathbf{H}_\star(n) = \mathbf{H}_{\star,1}(n)\mathbf{H}_{\star,2}^T(n)$ is the equivalent matrix (of size $L_{\star,1} \times L_{\star,2}$) representation of $\mathbf{h}_\star$. In case of the component impulse responses shown in Figs. 1(a) and (b), the singular values of their corresponding matrices $\mathbf{H}_\star$ are represented in Figs. 1(c) and (d), respectively, using $L_{{\rm er},1} = L_{{\rm lr},1} = 25$ and $L_{{\rm er},2} = L_{{\rm lr},2} = 20$. The faster these singular values decrease to zero, the farther the column rank of the corresponding matrix is from its maximum value. Hence, if ${\rm rank}\left[\mathbf{H}_\star(n)\right] = P_\star < L_{\star,2}$, the components of the impulse response can be decomposed as $\mathbf{h}_\star(n) = \sum_{p=1}^{P_\star} \mathbf{h}_{\star,2}^p(n) \otimes \mathbf{h}_{\star,1}^p(n)$.

Based on the previous considerations, it is more reasonable to process the estimation of $\mathbf{h}_{\rm er}(n)$ and $\mathbf{h}_{\rm lr}(n)$ differently as they may have different ranks. First, we can rewrite the components of the impulse response as

$$\begin{aligned}\mathbf{h}_\star(n) &= \sum_{p=1}^{P_\star} \left[\mathbf{h}_{\star,2}^p(n) \otimes \mathbf{I}_{L_{\star,1}}(n)\right] \mathbf{h}_{\star,1}^p(n) \\ &= \sum_{p=1}^{P_\star} \mathbf{H}_{\mathbf{h}_{\star,2}^p}(n)\mathbf{h}_{\star,1}^p(n) = \underline{\mathbf{H}}_{\star,2}(n)\underline{\mathbf{h}}_{\star,1}(n),\end{aligned} \tag{4}$$

with $\mathbf{I}_{L_{\star,1}}$ the $L_{\star,1} \times L_{\star,1}$ identity matrix, $\mathbf{H}_{\mathbf{h}_{\star,2}^p}(n) = \mathbf{h}_{\star,2}^p(n) \otimes \mathbf{I}_{L_{\star,1}}$, $\underline{\mathbf{H}}_{\star,2}(n) = \begin{bmatrix} \mathbf{H}_{\mathbf{h}_{\star,2}^1}(n) & \mathbf{H}_{\mathbf{h}_{\star,2}^2}(n) & \cdots & \mathbf{H}_{\mathbf{h}_{\star,2}^{P_\star}}(n) \end{bmatrix}$ has the size $L_\star \times P_\star L_{\star,1}$, and the column vector $\underline{\mathbf{h}}_{\star,1}(n) = \begin{bmatrix} \left(\mathbf{h}_{\star,1}^1\right)^T(n) & \left(\mathbf{h}_{\star,1}^2\right)^T(n) & \cdots & \left(\mathbf{h}_{\star,1}^{P_\star}\right)^T(n) \end{bmatrix}^T$ has the length $P_\star L_{\star,1}$. Alternatively, we can rewrite

$$\begin{aligned}\mathbf{h}_\star(n) &= \sum_{p=1}^{P_\star} \left[\mathbf{I}_{L_{\star,2}} \otimes \mathbf{h}_{\star,1}^p(n)\right] \mathbf{h}_{\star,2}^p(n) \\ &= \sum_{p=1}^{P_\star} \mathbf{H}_{\mathbf{h}_{\star,1}^p}(n)\mathbf{h}_{\star,2}^p(n) = \underline{\mathbf{H}}_{\star,1}(n)\underline{\mathbf{h}}_{\star,2}(n),\end{aligned} \tag{5}$$

where $\mathbf{I}_{L_{\star,2}}$ is the $L_{\star,2} \times L_{\star,2}$ identity matrix, $\mathbf{H}_{\mathbf{h}_{\star,1}^p}(n) = \mathbf{I}_{L_{\star,2}} \otimes \mathbf{h}_{\star,1}^p(n)$, $\underline{\mathbf{H}}_{\star,1}(n) = \begin{bmatrix} \mathbf{H}_{\mathbf{h}_{\star,1}^1}(n) & \mathbf{H}_{\mathbf{h}_{\star,1}^2}(n) & \cdots & \mathbf{H}_{\mathbf{h}_{\star,1}^{P_\star}}(n) \end{bmatrix}$ is a matrix of size $L_\star \times P_\star L_{\star,2}$, and the column vector $\underline{\mathbf{h}}_{\star,2}(n) = \begin{bmatrix} \left(\mathbf{h}_{\star,2}^1\right)^T(n) & \left(\mathbf{h}_{\star,2}^2\right)^T(n) & \cdots & \left(\mathbf{h}_{\star,2}^{P_\star}\right)^T(n) \end{bmatrix}^T$ has $P_\star L_{\star,2}$ elements. Next, the input signal vector can also be decomposed as $\mathbf{x}(n) = \begin{bmatrix} \mathbf{x}_{\rm er}^T(n) & \mathbf{x}_{\rm lr}^T(n) \end{bmatrix}^T$, where $\mathbf{x}_{\rm er}(n)$ and $\mathbf{x}_{\rm lr}(n)$ are the parts of the input signal related to early reflections and late reverberation, respectively. Thus, the reference signal in (1) can be written in two equivalent forms:

$$\begin{aligned} d(n) &= \underline{\overline{\mathbf{h}}}_{:1}^T(n) \begin{bmatrix} \underline{\mathbf{H}}_{{\rm er},2}^T(n)\mathbf{x}_{\rm er}(n) \\ \underline{\mathbf{H}}_{{\rm lr},2}^T(n)\mathbf{x}_{\rm lr}(n) \end{bmatrix} + v(n) & (6) \\ &= \underline{\overline{\mathbf{h}}}_{:2}^T(n) \begin{bmatrix} \underline{\mathbf{H}}_{{\rm er},1}^T(n)\mathbf{x}_{\rm er}(n) \\ \underline{\mathbf{H}}_{{\rm lr},1}^T(n)\mathbf{x}_{\rm lr}(n) \end{bmatrix} + v(n), & (7) \end{aligned}$$

where $\underline{\overline{\mathbf{h}}}_{:1}(n) = \begin{bmatrix} \underline{\mathbf{h}}_{{\rm er},1}^T(n) & \underline{\mathbf{h}}_{{\rm lr},1}^T(n) \end{bmatrix}^T$ has $P_{\rm er}L_{{\rm er},1} + P_{\rm lr}L_{{\rm lr},1}$ coefficients and $\underline{\overline{\mathbf{h}}}_{:2}(n) = \begin{bmatrix} \underline{\mathbf{h}}_{{\rm er},2}^T(n) & \underline{\mathbf{h}}_{{\rm lr},2}^T(n) \end{bmatrix}^T$ has $P_{\rm er}L_{{\rm er},2} + P_{\rm lr}L_{{\rm lr},2}$ coefficients. It can be noticed that the original SISO system identification problem that targets the estimation of L coefficients [of $\mathbf{h}(n)$] can be solved based

[1]Choosing $L_{\rm er} = L_{\rm lr} = L/2$ is the simplest and likely the most practical setup when no a priori information about the room's characteristics or the acoustic impulse response are available.

on a combination of two filters, $\underline{\overline{\mathbf{h}}}_{:1}(n)$ and $\underline{\overline{\mathbf{h}}}_{:2}(n)$, according to (6) and (7). For the common values of the decomposition parameters (i.e., $L_{\star,1}$, $L_{\star,2}$, and $P_\star$), the total number of coefficients to be estimated can be much smaller than L.

III. Decomposition-Based Kalman Filter

In order to develop a Kalman filter based on the previous approach, let us consider that the two impulse responses from (6) and (7) are zero-mean random vectors, which follow the simplified first-order Markov models:

$$\underline{\overline{\mathbf{h}}}_{:1}(n) = \underline{\overline{\mathbf{h}}}_{:1}(n-1) + \mathbf{w}_1(n), \tag{8}$$

$$\underline{\overline{\mathbf{h}}}_{:2}(n) = \underline{\overline{\mathbf{h}}}_{:2}(n-1) + \mathbf{w}_2(n), \tag{9}$$

where $\mathbf{w}_1(n)$ and $\mathbf{w}_2(n)$ are zero-mean white Gaussian noise vectors, which are uncorrelated to $\underline{\overline{\mathbf{h}}}_{:1}(n-1)$ and $\underline{\overline{\mathbf{h}}}_{:2}(n-1)$, respectively. The previous relations from (8) and (9) represent the state equations, while (6) and (7) are the observation equations. In this "bilinear" framework, the optimal estimates of $\underline{\overline{\mathbf{h}}}_{:1}(n)$ and $\underline{\overline{\mathbf{h}}}_{:2}(n)$ recursively result based on the linear sequential Bayesian approach [16] as

$$\widehat{\underline{\overline{\mathbf{h}}}}_{:1}(n) = \widehat{\underline{\overline{\mathbf{h}}}}_{:1}(n-1) + \mathbf{k}_1(n)e(n), \tag{10}$$

$$\widehat{\underline{\overline{\mathbf{h}}}}_{:2}(n) = \widehat{\underline{\overline{\mathbf{h}}}}_{:2}(n-1) + \mathbf{k}_2(n)e(n), \tag{11}$$

where $\mathbf{k}_1(n)$ and $\mathbf{k}_2(n)$ are the Kalman gain vectors [17], while $e(n)$ denotes the a priori estimation error, which can be evaluated based on (6) and (7). Also, in a more compact form,

$$e(n) = d(n) - \widehat{\underline{\overline{\mathbf{h}}}}_{:1}^T(n-1)\overline{\mathbf{x}}_{:2}(n) \tag{12}$$

$$= d(n) - \widehat{\underline{\overline{\mathbf{h}}}}_{:2}^T(n-1)\overline{\mathbf{x}}_{:1}(n), \tag{13}$$

where

$$\overline{\mathbf{x}}_{:2}(n) = \begin{bmatrix} \widehat{\underline{\mathbf{H}}}_{\mathrm{er},2}^T(n-1)\mathbf{x}_{\mathrm{er}}(n) \\ \widehat{\underline{\mathbf{H}}}_{\mathrm{lr},2}^T(n-1)\mathbf{x}_{\mathrm{lr}}(n) \end{bmatrix},$$

$$\overline{\mathbf{x}}_{:1}(n) = \begin{bmatrix} \widehat{\underline{\mathbf{H}}}_{\mathrm{er},1}^T(n-1)\mathbf{x}_{\mathrm{er}}(n) \\ \widehat{\underline{\mathbf{H}}}_{\mathrm{lr},1}^T(n-1)\mathbf{x}_{\mathrm{lr}}(n) \end{bmatrix},$$

and $\widehat{\underline{\mathbf{H}}}_{\star,2}^T(n)$ and $\widehat{\underline{\mathbf{H}}}_{\star,1}^T(n)$ are constructed in a similar way to $\underline{\mathbf{H}}_{\star,2}(n)$ and $\underline{\mathbf{H}}_{\star,1}(n)$, respectively [according to (4) and (5)], using the estimates from (10) and (11).

Next, we define the a priori and a posteriori misalignments associated to the optimal estimates of the state vectors:

$$\mathbf{m}_1(n) = \underline{\overline{\mathbf{h}}}_{:1}(n) - \widehat{\underline{\overline{\mathbf{h}}}}_{:1}(n-1),\ \mathbf{m}_2(n) = \underline{\overline{\mathbf{h}}}_{:2}(n) - \widehat{\underline{\overline{\mathbf{h}}}}_{:2}(n-1),$$

$$\widetilde{\mathbf{m}}_1(n) = \underline{\overline{\mathbf{h}}}_{:1}(n) - \widehat{\underline{\overline{\mathbf{h}}}}_{:1}(n),\ \widetilde{\mathbf{m}}_2(n) = \underline{\overline{\mathbf{h}}}_{:2}(n) - \widehat{\underline{\overline{\mathbf{h}}}}_{:2}(n).$$

Based on the state equations, it can be noticed that $\mathbf{m}_1(n) = \widetilde{\mathbf{m}}_1(n-1) + \mathbf{w}_1(n)$ and $\mathbf{m}_2(n) = \widetilde{\mathbf{m}}_2(n-1) + \mathbf{w}_2(n)$. Thus, in terms of the associated correlation matrices, we obtain

$$\mathbf{R}_{\mathbf{m}_1}(n) = \mathbf{R}_{\widetilde{\mathbf{m}}_1}(n-1) + \mathbf{R}_{\mathbf{w}_1}(n), \tag{14}$$

$$\mathbf{R}_{\mathbf{m}_2}(n) = \mathbf{R}_{\widetilde{\mathbf{m}}_2}(n-1) + \mathbf{R}_{\mathbf{w}_2}(n), \tag{15}$$

where

$$\mathbf{R}_{\mathbf{m}_1}(n) = E\left[\mathbf{m}_1(n)\mathbf{m}_1^T(n)\right],\ \mathbf{R}_{\widetilde{\mathbf{m}}_1}(n) = E\left[\widetilde{\mathbf{m}}_1(n)\widetilde{\mathbf{m}}_1^T(n)\right],$$

$$\mathbf{R}_{\mathbf{m}_2}(n) = E\left[\mathbf{m}_2(n)\mathbf{m}_2^T(n)\right],\ \mathbf{R}_{\widetilde{\mathbf{m}}_2}(n) = E\left[\widetilde{\mathbf{m}}_2(n)\widetilde{\mathbf{m}}_2^T(n)\right],$$

$$\mathbf{R}_{\mathbf{w}_1}(n) = E\left[\mathbf{w}_1(n)\mathbf{w}_1^T(n)\right] = \sigma_{w_1}^2(n)\mathbf{I}_{P_{\mathrm{er}}L_{\mathrm{er},1}+P_{\mathrm{lr}}L_{\mathrm{lr},1}},$$

$$\mathbf{R}_{\mathbf{w}_2}(n) = E\left[\mathbf{w}_2(n)\mathbf{w}_2^T(n)\right] = \sigma_{w_2}^2(n)\mathbf{I}_{P_{\mathrm{er}}L_{\mathrm{er},2}+P_{\mathrm{lr}}L_{\mathrm{lr},2}},$$

while the variances $\sigma_{w_1}^2(n)$ and $\sigma_{w_2}^2(n)$ that multiply the identity matrices of the corresponding sizes (indicated in subscripts) capture the uncertainties in the state vectors.

The Kalman gain vectors required in the updates (10) and (11) are obtained by minimizing the optimization criteria:

$$\mathcal{J}_1(n) = \left(P_{\mathrm{er}}L_{\mathrm{er},1} + P_{\mathrm{lr}}L_{\mathrm{lr},1}\right)^{-1} \mathrm{tr}\left[\mathbf{R}_{\widetilde{\mathbf{m}}_1}(n)\right], \tag{16}$$

$$\mathcal{J}_2(n) = \left(P_{\mathrm{er}}L_{\mathrm{er},2} + P_{\mathrm{lr}}L_{\mathrm{lr},2}\right)^{-1} \mathrm{tr}\left[\mathbf{R}_{\widetilde{\mathbf{m}}_2}(n)\right], \tag{17}$$

where $\mathrm{tr}[\cdot]$ denotes the trace of a square matrix. The development of (16) and (17) is based on a bilinear optimization strategy [18], [19], considering that one of the systems is fixed within the optimization criterion of the other one. Similar strategies were adopted in [3]– [5] and [14]. Following this approach, the Kalman gain vectors result in

$$\mathbf{k}_1(n) = \mathbf{R}_{\mathbf{m}_1}(n)\overline{\mathbf{x}}_{:2}(n)\left[\overline{\mathbf{x}}_{:2}^T(n)\mathbf{R}_{\mathbf{m}_1}(n)\overline{\mathbf{x}}_{:2}(n) + \sigma_v^2\right]^{-1}, \tag{18}$$

$$\mathbf{k}_2(n) = \mathbf{R}_{\mathbf{m}_2}(n)\overline{\mathbf{x}}_{:1}(n)\left[\overline{\mathbf{x}}_{:1}^T(n)\mathbf{R}_{\mathbf{m}_2}(n)\overline{\mathbf{x}}_{:1}(n) + \sigma_v^2\right]^{-1}, \tag{19}$$

while the matrices $\mathbf{R}_{\widetilde{\mathbf{m}}_1}(n)$ and $\mathbf{R}_{\widetilde{\mathbf{m}}_2}(n)$ are obtained as

$$\mathbf{R}_{\widetilde{\mathbf{m}}_1}(n) = \mathbf{R}_{\mathbf{m}_1}(n) - \mathbf{k}_1(n)\overline{\mathbf{x}}_{:2}^T(n)\mathbf{R}_{\mathbf{m}_1}(n), \tag{20}$$

$$\mathbf{R}_{\widetilde{\mathbf{m}}_2}(n) = \mathbf{R}_{\mathbf{m}_2}(n) - \mathbf{k}_2(n)\overline{\mathbf{x}}_{:1}^T(n)\mathbf{R}_{\mathbf{m}_2}(n). \tag{21}$$

Finally, the coefficients of $\widehat{\underline{\overline{\mathbf{h}}}}_{:1}(n)$ and $\widehat{\underline{\overline{\mathbf{h}}}}_{:2}(n)$ can be combined based on (2) and (4)–(7), in order to obtain an estimate of the global impulse response, $\widehat{\mathbf{h}}(n)$.

The resulting Kalman filter based on the Kronecker product decompositions, namely KF-KPD, is defined by equations (10)–(15) and (18)–(21). There are some parameters that have to be set or estimated within the KF-KPD. The first one is σ_v^2, which represents the variance of the additive noise. For example, in many applications related to the acoustic environment, the power of the background noise can be estimated during silence periods [2]. Other practical methods for estimating σ_v^2 can be found in [20]. The other parameters that should be set are $\sigma_{w_1}^2(n)$ and $\sigma_{w_2}^2(n)$, and they are related to the uncertainties in the system. The choice of these parameters leads to a compromise between the tracking capabilities and the accuracy of the estimates. Small values of $\sigma_{w_1}^2(n)$ and $\sigma_{w_2}^2(n)$ lead to a good accuracy (i.e., low misalignment), but reducing the tracking capabilities of the filter. On the other hand, increasing the values of these parameters improves the tracking behavior, while paying with a less accurate solution (i.e., high misalignment). A very simple method to evaluate these parameters is inspired from [21] and it is related to the uncertainties in the global filter. Therefore, we can set $\sigma_{w_1}^2(n) = \sigma_{w_2}^2(n) = (1/L)\left\|\widehat{\mathbf{h}}(n) - \widehat{\mathbf{h}}(n-1)\right\|_2^2$, where $\|\cdot\|_2$

denotes the Euclidean norm. In this case, when the KF-KPD starts to converge or when there is an abrupt change of the system to be identified, the difference between the estimates at consecutive time indexes is large, and so are the values of the uncertainties parameters. Contrary, when the algorithm is converging to its steady-state, this difference is reducing, together with the values of $\sigma_{w_1}^2(n)$ and $\sigma_{w_2}^2(n)$. An alternative (more elaborated) method involves individual control factors [22], by considering independent fluctuations for each coefficient [using different values on the diagonal of the matrices $\mathbf{R}_{\mathbf{w}_1}(n)$ and $\mathbf{R}_{\mathbf{w}_2}(n)$]. However, the influence of these methods on the performance of the KF-KPD is beyond the scope of this paper.

In terms of the computational complexity, the conventional Kalman filter involves a computational amount proportional to $\mathcal{O}(L^2)$. This could be prohibitively expensive for most of the practical applications that require the identification of acoustic impulse responses, due to the very large values of L. As shown before, the proposed KF-KPD combines the coefficients of two shorter filters of lengths $P_{\text{er}}L_{\text{er},1}+P_{\text{lr}}L_{\text{lr},1}$ and $P_{\text{er}}L_{\text{er},2}+P_{\text{lr}}L_{\text{lr},2}$, where $L=L_{\text{er},1}L_{\text{er},2}+L_{\text{lr},1}L_{\text{lr},2}$, while $P_{\text{er}}<L_{\text{er},2}$ and $P_{\text{lr}} \ll L_{\text{lr},2}$ [14]. Even if the computational complexity of the KF-KPD is still proportional to the square of the filters' lengths, it is more advantageous as compared to the conventional Kalman filter. In addition, improved performance is expected for the KF-KPD, in terms of the convergence and accuracy features. The proposed algorithm performs similar to the previously developed Kalman filter based on the nearest Kronecker product decomposition (KF-NKP) [5]. However, the KF-NKP performs the decomposition using two filters of length PL_1 and PL_2, with $L=L_1L_2$ and $P<L_2$, thus involving $\mathcal{O}\left[(PL_1)^2+(PL_2)^2\right]$ operations. On the other hand, the proposed KF-KPD has the complexity order $\mathcal{O}\left[(P_{\text{er}}L_{\text{er},1}+P_{\text{lr}}L_{\text{lr},1})^2+(P_{\text{er}}L_{\text{er},2}+P_{\text{lr}}L_{\text{lr},2})^2\right]$, so that is more advantageous in terms of the computation complexity, for the common setup of its parameters.

IV. Simulation Results

Simulations are performed in the context of acoustic echo cancellation. Two measured acoustic impulse responses are used in the experiments, with $L=1000$. They have different sparseness degrees [23], which are evaluated using the measure $\xi_{12}\left[\mathbf{h}(n)\right] = L/(L-\sqrt{L})\left\{1-\|\mathbf{h}(n)\|_1/[\sqrt{L}\|\mathbf{h}(n)\|_2]\right\}$, where $\|\cdot\|_1$ stands for the ℓ_1 norm (i.e., the sum of the absolute values of the coefficients). Clearly, $0 \leq \xi_{12}\left[\mathbf{h}(n)\right] \leq 1$. A smaller value of this measure is related to a denser impulse response, so that its corresponding matrix is closer to full-rank, while a larger value of the sparseness measure implies a sparser impulse response and a lower rank of the corresponding matrix. The performance measure used in all the experiments is the normalized misalignment (in dB), which is evaluated as $20\log_{10}\left[\left\|\mathbf{h}(n)-\widehat{\mathbf{h}}(n)\right\|_2/\left\|\mathbf{h}(n)\right\|_2\right]$.

In the following, the two acoustic impulse responses are referred as $\mathbf{h}_\alpha$ and $\mathbf{h}_\beta$; the second one was used for the example provided in Fig. 1. Their sparseness measures are $\xi_{12}\left(\mathbf{h}_\alpha\right)=0.7384$ and $\xi_{12}\left(\mathbf{h}_\beta\right)=0.6846$. Two input signals are used in simulations, with a sampling rate of 8 kHz. The first one is an autoregressive (AR) process obtained by filtering a white Gaussian noise through a first-order AR model [AR(1)] with a pole at 0.9. The second one is a recorded speech signal. The background noise $v(n)$ is white and Gaussian, with a signal-to-noise ratio (SNR) of 20 dB. The SNR is evaluated based on (1) as $\text{SNR}=\sigma_y^2/\sigma_v^2$, where σ_y^2 denotes the variance of $y(n)$. We consider that σ_v^2 is available in the experiments.

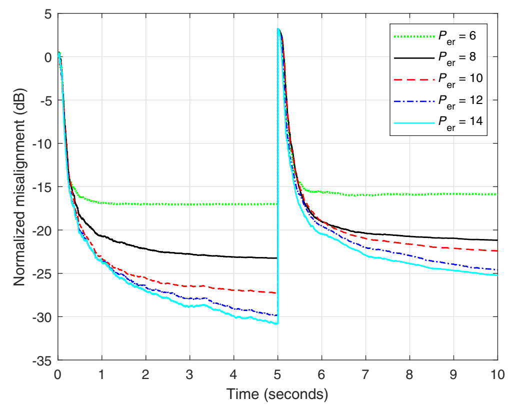

Fig. 2. Normalized misalignment of the KF-KPD using $P_{\text{lr}}=2$ and different values of P_{er}. The input signal in an AR(1) process and the echo path changes after 5 seconds (from $\mathbf{h}_\alpha$ to $\mathbf{h}_\beta$).

The decomposition setup for the KF-KPD is the same used in Fig. 1, i.e., $L_{\text{er},1}=L_{\text{lr},1}=25$ and $L_{\text{er},2}=L_{\text{lr},2}=20$. Based on the findings from [14], a value of $P_{\text{lr}} \ll L_{\text{lr},2}$ is recommended, which is also justified in Fig. 1(d). Consequently, we set $P_{\text{lr}}=2$ in all the following experiments. First, in Fig. 2, the performance of the KF-KPD is evaluated for different values of $P_{\text{er}}<L_{\text{er},2}$. The input signal is an AR(1) process and the echo path changes after 5 seconds (from $\mathbf{h}_\alpha$ to $\mathbf{h}_\beta$). As we can notice, the performance is improved when increasing the value of P_{er}, but up to a limit that is related to the approximated rank of $\mathbf{H}_{\text{er}}(n)$, as indicated in Fig. 1(c). Also, it can be noticed that a better performance is obtained for the identification of the sparser impulse response, since the approximated rank of its corresponding matrix is lower in this case. Nevertheless, for both impulse responses, a value of $P_{\text{er}}=8$ (which is reasonably lower than $L_{\text{er},2}$) leads to a good attenuation of the misalignment.

A similar setup is used in Fig. 3, where the KF-KPD is compared to the recently developed RLS-KPD algorithm [14] using different values of the forgetting factors, which are chosen as $\lambda_1 = 1-1/\left[K\left(P_{\text{er}}L_{\text{er},1}+P_{\text{lr}}L_{\text{lr},1}\right)\right]$ and $\lambda_2 = 1-1/\left[K\left(P_{\text{er}}L_{\text{er},2}+P_{\text{lr}}L_{\text{lr},2}\right)\right]$, with $K>1$, $P_{\text{er}}=8$, and $P_{\text{lr}}=2$. As we can notice, increasing the forgetting factors of the RLS-KPD algorithm (i.e., using larger values of K) improves the misalignment, but sacrifices the tracking capability. On the other hand, the proposed KF-KPD achieves a better compromise between the performance criteria.

Finally, in Fig. 4, the conventional Kalman filter (KF) is

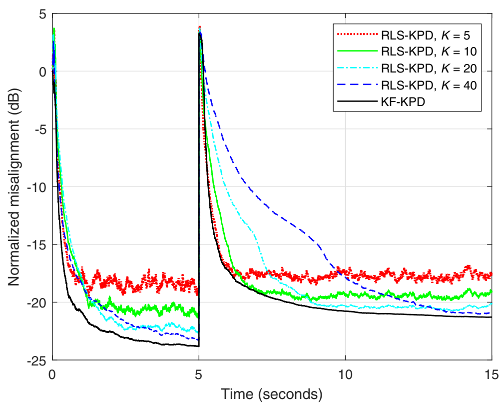

Fig. 3. Normalized misalignment of the RLS-KPD algorithm [14] using different values of the forgetting factors (i.e., different values of K) and the KF-KPD. Both algorithms use the same decomposition setup, with $P_{\text{er}} = 8$ and $P_{\text{lr}} = 2$. The input signal in an AR(1) process and the echo path changes after 5 seconds (from $\mathbf{h}_\alpha$ to $\mathbf{h}_\beta$).

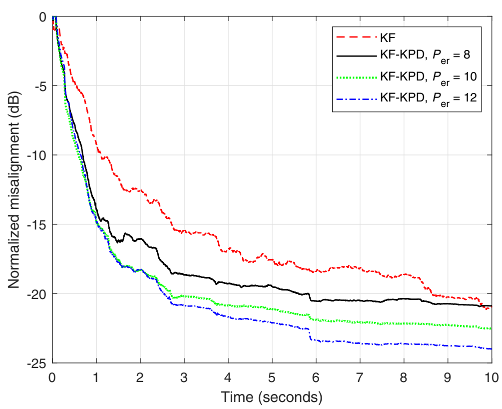

Fig. 4. Normalized misalignment of the conventional KF and the KF-KPD using $P_{\text{lr}} = 2$ and different values of P_{er}, for the identification of the impulse response $\mathbf{h}_\alpha$. The input signal is speech.

compared to the KF-KPD using $P_{\text{lr}} = 2$ and different values of P_{er}. In this experiment, the input signal is speech and the impulse response $\mathbf{h}_\alpha$ is considered. Since the proposed KF-KPD combines the solution of two shorter filters, it outperforms the conventional KF in terms of the convergence features. This gain can be noticed even for $P_{\text{er}} = 8$, while the performance of the KF-KPD is improving for larger values of this decomposition parameter.

V. Conclusions

In this paper, we have developed a Kalman filter for the identification of acoustic impulse responses, which uses Kronecker product decompositions and low-rank approximations. In this framework, the different characteristics of the two main components of the acoustic impulse response (i.e., early reflections and late reverberation) are exploited. The resulting KF-KPD outperforms the previously developed RLS-KPD [14] in terms of the main convergence criteria.

References

[1] P. S. R. Diniz, *Adaptive Filtering: Algorithms and Practical Implementation*. Fourth Edition, New York, NY, USA: Springer-Verlag, 2013.

[2] E. Hänsler and G. Schmidt, *Acoustic Echo and Noise Control–A Practical Approach*. Hoboken, NJ, USA: Wiley, 2004.

[3] C. Paleologu, J. Benesty, and S. Ciochină, "Linear system identification based on a Kronecker product decomposition," *IEEE/ACM Trans. Audio, Speech, Language Processing*, vol. 26, pp. 1793–1808, Oct. 2018.

[4] C. Elisei-Iliescu, C. Paleologu, J. Benesty, C. Stanciu, C. Anghel, and S. Ciochină, "Recursive least-squares algorithms for the identification of low-rank systems," *IEEE/ACM Trans. Audio, Speech, Language Processing*, vol. 27, pp. 903–918, May 2019.

[5] L.-M. Dogariu, C. Paleologu, J. Benesty, and S. Ciochină, "An efficient Kalman filter for the identification of low-rank systems," *Signal Processing*, vol. 166, id. 107239, 9 pages, Jan. 2020.

[6] W. Yang, G. Huang, J. Chen, J. Benesty, I. Cohen, and W. Kellermann, "Robust dereverberation with Kronecker product based multichannel linear prediction," *IEEE Signal Processing Lett.*, vol. 28, pp. 101–105, Jan. 2021.

[7] G. Itzhak, J. Benesty, and I. Cohen, "On the design of differential Kronecker product beamformers," *IEEE/ACM Trans. Audio, Speech, Language Processing*, vol. 29, pp. 1397–1410, Mar. 2021.

[8] E. V. Kuhn, C. A. Pitz, M. V. Matsuo, K. J. Bakri, R. Seara, and J. Benesty, "A Kronecker product CLMS algorithm for adaptive beamforming," *Digital Signal Processing*, vol. 111, id. 102968, Apr. 2021.

[9] S. S. Bhattacharjee and N. V. George, "Fast and efficient acoustic feedback cancellation based on low rank approximation," *Signal Processing*, vol. 182, id. 107984, May 2021.

[10] S. S. Bhattacharjee and N. V. George, "Nearest Kronecker product decomposition based linear-in-the-parameters nonlinear filters," *IEEE/ACM Trans. Audio, Speech, Language Processing*, vol. 29, pp. 2111–2122, May 2021.

[11] X. Wang, J. Benesty, J. Chen, G. Huang, and I. Cohen, "Beamforming with cube microphone arrays via Kronecker product decompositions," *IEEE/ACM Trans. Audio, Speech, Language Processing*, vol. 29, pp. 1774–1784, May 2021.

[12] G. Huang, J. Benesty, I. Cohen, and J. Chen, "Kronecker product multichannel linear filtering for adaptive weighted prediction error-based speech dereverberation," *IEEE/ACM Trans. Audio, Speech, Language Processing*, vol. 30, pp. 1277–1289, Mar. 2022.

[13] S. Vadhvana, S. K. Yadav, S. S. Bhattacharjee, and N. V. George, "An improved constrained LMS algorithm for fast adaptive beamforming based on a low rank approximation," *IEEE Trans. Circuits Systems II: Express Briefs*, vol. 69, pp. 3605–3609, Aug. 2022.

[14] L.-M. Dogariu, J. Benesty, C. Paleologu, and S. Ciochină, "Identification of room acoustic impulse responses via Kronecker product decompositions," *IEEE/ACM Trans. Audio, Speech, Language Processing*, vol. 30, pp. 2828–2841, Sept. 2022.

[15] C. F. Van Loan, "The ubiquitous Kronecker product," *J. Computational Applied Mathematics*, vol. 123, pp. 85–100, 2000.

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

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

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

[19] J. Benesty, C. Paleologu, and S. Ciochină, "On the identification of bilinear forms with the Wiener filter," *IEEE Signal Processing Lett.*, vol. 24, pp. 653–657, May 2017.

[20] C. Paleologu, J. Benesty, S. L. Grant, and C. Osterwise, "Variable step-size NLMS algorithms designed for echo cancellation," in *Proc. IEEE Asilomar*, 2009, pp. 633–637.

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

[22] C. Paleologu, J. Benesty, S. Ciochină, and S. L. Grant, "A Kalman filter with individual control factors for echo cancellation," in *Proc. IEEE ICASSP*, 2014, pp. 6015–6019.

[23] P. O. Hoyer, "Non-negative matrix factorization with sparseness constraints," *J. Machine Learning Research*, vol. 49, pp. 1208–1215, June 2001.