A RECURSIVE LEAST-SQUARES ALGORITHM BASED ON THE NEAREST KRONECKER PRODUCT DECOMPOSITION

Camelia Elisei-Iliescu[†], *Constantin Paleologu*[†], *Jacob Benesty*[‡], *and Silviu Ciochină*[†]

[†] University Politehnica of Bucharest, Romania, e-mail: {pale, silviu}@comm.pub.ro
[‡] INRS-EMT, University of Quebec, Montreal, Canada, e-mail: benesty@emt.inrs.ca

ABSTRACT

The recursive least-squares (RLS) adaptive filter is an appealing choice in system identification problems, mainly due to its fast convergence rate. However, this algorithm is computationally very complex, which may make it useless for the identification of high length impulse responses, like in echo cancellation. In this paper, we focus on a new approach to improve the efficiency of the RLS algorithm. The basic idea is to exploit the impulse response decomposition based on the nearest Kronecker product and low-rank approximation. Thus, a high-dimension system identification problem is reformulated in terms of low-dimension problems, which are tensorized together. Simulations performed in the context of echo cancellation indicate the good performance of the RLS algorithm based on this approach.

Index Terms— Adaptive filter, echo cancellation, low-rank approximation, nearest Kronecker product, recursive least-squares.

1. INTRODUCTION

In many system identification problems, the basic approach is to model the unknown system using an adaptive filter [1]. However, in several applications, the unknown system can be modeled as a finite impulse response filter with a large number of coefficients, which raises additional challenges for the adaptive filter used for this purpose. In this context, the echo cancellation application [2], [3] represents one of the most popular examples, which involves long length adaptive filters (e.g., hundreds or even thousands of coefficients).

Due to its fast convergence rate, the recursive least-squares (RLS) algorithm could represent an appealing choice for system identification problems. However, its main limitation is related to the high computational complexity, which becomes prohibited in case of long length adaptive filters. Several methods to reduce the computational amount of the RLS algorithm can be found in the literature, which lead to the so-called fast or efficient versions, e.g., see [4]–[7] and the references therein. Nevertheless, the problem of dealing with long length adaptive filters is still a critical issue in system identification contexts like echo cancellation.

Recently, in [8], we presented a system identification approach based on an optimal low-rank approximation that involves the Kronecker product decomposition [9] of the impulse response. The main idea was to exploit this impulse response decomposition, so that a high-dimension system identification problem could be reformulated in terms of low-dimension problems, which are tensorized together. This approach fits very well for the identification of typical echo paths (which are usually low-rank systems). In [8], we addressed the system identification problem based on an iterative Wiener filter. However, due to its inherent limitations, the Wiener filter is not a practical solution in real-world echo cancellation scenarios.

In this paper, we propose an RLS algorithm based on the nearest Kronecker product decomposition together with low-rank approximation. These techniques were previously exploited in different frameworks, which are mainly related to tensor decomposition and modelling [10]–[21]. However, the adaptive algorithms developed in this context, e.g., [22]–[27], did not involve low-rank approximations. Thus, they face a major limitation related to the particular (quasi-periodic) form of the impulse response to be identified. While this form resembles (to some extend) a channel with echoes, it is far away from the very different natures of realistic echo paths. The proposed RLS algorithm is designed to overcome this limitation, being applicable for the identification of low-rank systems, like echo paths.

The paper is organized as follows. Section 2 summarizes the framework of our approach [8]. The proposed RLS algorithm based on the nearest Kronecker product decomposition is derived in Section 3. Simulation results (in the context of echo cancellation) are provided in Section 4. Finally, Section 5 concludes this work.

2. LINEAR SYSTEM IDENTIFICATION BASED ON A KRONECKER PRODUCT DECOMPOSITION

Let $\mathbf{h}$ be a real-valued (unknown) impulse response of length $L = L_1 L_2$. Without loss of generality, we may assume that $L_1 \geq L_2$. Hence, the impulse response can be decomposed as

$$\mathbf{h} = \begin{bmatrix} \mathbf{s}_1^T & \mathbf{s}_2^T & \dots & \mathbf{s}_{L_2}^T \end{bmatrix}^T, \tag{1}$$

where $\mathbf{s}_l,\ l = 1, 2, \dots, L_2$ are L_2 short impulse responses of length L_1 each and the superscript ^{T} is the transpose operator. In [8], it was shown how well $\mathbf{h}$ can be approximated by $\mathbf{h}_2 \otimes \mathbf{h}_1$, where $\mathbf{h}_1$ and $\mathbf{h}_2$ are two impulse responses of lengths L_1 and L_2, respectively, and $\otimes$ is the Kronecker product. To this purpose, let us define the normalized misalignment as $\mathcal{M}(\mathbf{h}_1, \mathbf{h}_2) = \|\mathbf{h} - \mathbf{h}_2 \otimes \mathbf{h}_1\|_2 / \|\mathbf{h}\|_2$ [9], where $\|\cdot\|_2$ denotes the ℓ_2 norm. Also, we can reorganize the components of $\mathbf{h}$ into a matrix of size $L_1 \times L_2$:

$$\mathbf{H} = \begin{bmatrix} \mathbf{s}_1 & \mathbf{s}_2 & \dots & \mathbf{s}_{L_2} \end{bmatrix}, \tag{2}$$

so that we can write $\mathcal{M}(\mathbf{h}_1, \mathbf{h}_2) = \left\|\mathbf{H} - \mathbf{h}_1 \mathbf{h}_2^T\right\|_\mathrm{F} / \|\mathbf{H}\|_\mathrm{F}$, where $\|\cdot\|_\mathrm{F}$ denotes the Frobenius norm. To find the optimal values of $\mathbf{h}_1$ and $\mathbf{h}_2$, we need to minimize $\mathcal{M}(\mathbf{h}_1, \mathbf{h}_2)$. But minimizing this quantity is equivalent to finding the nearest rank-1 matrix to $\mathbf{H}$ [9], which is related to its singular value decomposition (SVD) [28].

Thus, using the SVD, this matrix can be factorized as $\mathbf{H} = \mathbf{U}_1 \boldsymbol{\Sigma} \mathbf{U}_2^T = \sum_{l=1}^{L_2} \sigma_l \mathbf{u}_{1,l} \mathbf{u}_{2,l}^T$, where $\mathbf{U}_1$ and $\mathbf{U}_2$ are two orthogonal matrices of sizes $L_1 \times L_1$ and $L_2 \times L_2$, respectively, and $\boldsymbol{\Sigma}$ is an $L_1 \times L_2$ rectangular diagonal matrix with nonnegative real numbers on the main diagonal. The columns of $\mathbf{U}_1$ (resp. $\mathbf{U}_2$) are the left-singular (resp. right-singular) vectors of $\mathbf{H}$, while the diagonal entries $\sigma_l,\ l=1,2,\ldots,L_2$ of $\mathbf{\Sigma}$ represent the singular values of $\mathbf{H}$, with $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_{L_2} \geq 0$. In this context, the optimal impulse responses that minimize $\mathcal{M}(\mathbf{h}_1,\mathbf{h}_2)$ are $\overline{\mathbf{h}}_1 = \sqrt{\sigma_1}\mathbf{u}_{1,1}$ and $\overline{\mathbf{h}}_2 = \sqrt{\sigma_1}\mathbf{u}_{2,1}$, where $\mathbf{u}_{1,1}$ (resp. $\mathbf{u}_{2,1}$) is the first column of $\mathbf{U}_1$ (resp. $\mathbf{U}_2$). Hence, the optimal approximation of $\mathbf{h}$ is $\overline{\mathbf{h}} = \overline{\mathbf{h}}_2 \otimes \overline{\mathbf{h}}_1$.

In the more general case when the short impulse responses from (1) may not be that linearly dependent, we can use the approximation $\mathbf{h} \approx \sum_{p=1}^{P} \mathbf{h}_{2,p} \otimes \mathbf{h}_{1,p} = \text{vec}\left(\mathbf{H}_1\mathbf{H}_2^T\right)$, where $P \leq L_2$, $\mathbf{h}_{1,p}$ and $\mathbf{h}_{2,p}$ are impulse responses of lengths L_1 and L_2, respectively, $\text{vec}(\cdot)$ denotes the vectorization operation [29], and $\mathbf{H}_1 = \left[\begin{array}{cccc} \mathbf{h}_{1,1} & \mathbf{h}_{1,2} & \ldots & \mathbf{h}_{1,P} \end{array}\right]$ and $\mathbf{H}_2 = \left[\begin{array}{cccc} \mathbf{h}_{2,1} & \mathbf{h}_{2,2} & \ldots & \mathbf{h}_{2,P} \end{array}\right]$ are matrices of sizes $L_1 \times P$ and $L_2 \times P$, respectively. In this context, the problem is to minimize $\mathcal{M}(\mathbf{H}_1,\mathbf{H}_2) = \left\|\mathbf{H} - \mathbf{H}_1\mathbf{H}_2^T\right\|_{\text{F}} / \left\|\mathbf{H}\right\|_{\text{F}}$, which leads to the optimal solutions:

$$
\begin{aligned}
\overline{\mathbf{H}}_1 &= \left[\begin{array}{cccc} \overline{\mathbf{h}}_{1,1} & \overline{\mathbf{h}}_{1,2} & \ldots & \overline{\mathbf{h}}_{1,P} \end{array}\right] \qquad (3)\\
&= \left[\begin{array}{cccc} \sqrt{\sigma_1}\mathbf{u}_{1,1} & \sqrt{\sigma_2}\mathbf{u}_{1,2} & \ldots & \sqrt{\sigma_P}\mathbf{u}_{1,P} \end{array}\right],\\
\overline{\mathbf{H}}_2 &= \left[\begin{array}{cccc} \overline{\mathbf{h}}_{2,1} & \overline{\mathbf{h}}_{2,2} & \ldots & \overline{\mathbf{h}}_{2,P} \end{array}\right] \qquad (4)\\
&= \left[\begin{array}{cccc} \sqrt{\sigma_1}\mathbf{u}_{2,1} & \sqrt{\sigma_2}\mathbf{u}_{2,2} & \ldots & \sqrt{\sigma_P}\mathbf{u}_{2,P} \end{array}\right],
\end{aligned}
$$

where $\mathbf{u}_{1,p},\ p=1,2,\ldots,P$ (resp. $\mathbf{u}_{2,p},\ p=1,2,\ldots,P$) are the first P columns of $\mathbf{U}_1$ (resp. $\mathbf{U}_2$). Consequently, the optimal approximation of $\mathbf{h}$ is

$$\overline{\mathbf{h}}(P) = \sum_{p=1}^{P} \overline{\mathbf{h}}_{2,p} \otimes \overline{\mathbf{h}}_{1,p} = \sum_{p=1}^{P} \sigma_p \mathbf{u}_{2,p} \otimes \mathbf{u}_{1,p}, \qquad (5)$$

while for $P = L_2$, we have $\mathbf{h} = \sum_{l=1}^{L_2} \sigma_l \mathbf{u}_{2,l} \otimes \mathbf{u}_{1,l}$, which represents the exact decomposition.

Concluding, if for $P \ll L_2$, the approximation in (5) is a good one, we can take advantage of the nearest Kronecker product decomposition for system identification problems. In other words, if the rank of the matrix $\mathbf{H}$ is equal to $P \ll L_2$, then we can estimate $\mathbf{h}$ at least as well as in the conventional approach. This framework fits very well in the context of echo cancellation scenarios, where this matrix is never really full rank (because of the redundancies due to the reflections and/or sparseness in the system [30]–[33]).

3. RLS ALGORITHM BASED ON THE NEAREST KRONECKER PRODUCT DECOMPOSITION

The optimal filters required in (5) can be found using the iterative Wiener filter proposed in [8]. However, the Wiener filter may not always be very efficient or convenient to use in practice. In the following, we derive an RLS algorithm based on the previous approach.

Let us consider the signal model specific to a system identification problem (like in echo cancellation):

$$d(t) = \mathbf{h}^T\mathbf{x}(t) + w(t) = y(t) + w(t), \qquad (6)$$

where $d(t)$ is the zero-mean desired signal at the discrete-time index t, $\mathbf{h}$ is the impulse response of the unknown system (of length L), $\mathbf{x}(t) = \left[\begin{array}{cccc} x(t) & x(t-1) & \cdots & x(t-L+1) \end{array}\right]^T$ is a vector containing the most recent L time samples of the zero-mean input signal $x(t)$, and $w(t)$ is the zero-mean additive noise. It is assumed that all the data is real valued and $\mathbf{x}(t)$ and $w(t)$ are uncorrelated. The goal is to estimate $\mathbf{h}$ with an adaptive filter, $\widehat{\mathbf{h}}(t)$, of length L. Thus, we can define the error signal between the desired signal, $d(t)$, and the estimated signal, $\widehat{y}(t)$, as

$$e(t) = d(t) - \widehat{y}(t) = d(t) - \widehat{\mathbf{h}}^T(t-1)\mathbf{x}(t). \qquad (7)$$

As explained in Section 2, let us consider that $L = L_1L_2$ (with $L_1 \geq L_2$) and $\text{rank}(\mathbf{H}) = P \ll L_2$, where $\mathbf{H}$ is defined in (2), so that the impulse response can be decomposed as in (5). As a consequence, we can also choose to decompose the adaptive filter as

$$\widehat{\mathbf{h}}(t) = \sum_{p=1}^{P} \widehat{\mathbf{h}}_{2,p}(t) \otimes \widehat{\mathbf{h}}_{1,p}(t), \qquad (8)$$

where $\widehat{\mathbf{h}}_{1,p}(t)$ and $\widehat{\mathbf{h}}_{2,p}(t)$ are filters of lengths L_1 and L_2, respectively. Using the relationships $\left[\widehat{\mathbf{h}}_{2,p}(t) \otimes \mathbf{I}_{L_1}\right]\widehat{\mathbf{h}}_{1,p}(t) = \left[\mathbf{I}_{L_2} \otimes \widehat{\mathbf{h}}_{1,p}(t)\right]\widehat{\mathbf{h}}_{2,p}(t) = \widehat{\mathbf{h}}_{2,p}(t) \otimes \widehat{\mathbf{h}}_{1,p}(t)$, where $\mathbf{I}_{L_1}$ and $\mathbf{I}_{L_2}$ are the identity matrices of sizes $L_1 \times L_1$ and $L_2 \times L_2$, respectively, into (8), the error signal can be expressed into two different manners:

$$
\begin{aligned}
e(t) &= d(t) - \sum_{p=1}^{P} \widehat{\mathbf{h}}_{1,p}^T(t-1)\left[\widehat{\mathbf{h}}_{2,p}(t-1) \otimes \mathbf{I}_{L_1}\right]^T\mathbf{x}(t) \qquad (9)\\
&= d(t) - \sum_{p=1}^{P} \widehat{\mathbf{h}}_{1,p}^T(t-1)\mathbf{x}_{2,p}(t) = d(t) - \underline{\widehat{\mathbf{h}}}_1^T(t-1)\underline{\mathbf{x}}_2(t),\\
e(t) &= d(t) - \sum_{p=1}^{P} \widehat{\mathbf{h}}_{2,p}^T(t-1)\left[\mathbf{I}_{L_2} \otimes \widehat{\mathbf{h}}_{1,p}(t-1)\right]^T\mathbf{x}(t) \qquad (10)\\
&= d(t) - \sum_{p=1}^{P} \widehat{\mathbf{h}}_{2,p}^T(t-1)\mathbf{x}_{1,p}(t) = d(t) - \underline{\widehat{\mathbf{h}}}_2^T(t-1)\underline{\mathbf{x}}_1(t),
\end{aligned}
$$

where $\mathbf{x}_{2,p}(t) = \left[\widehat{\mathbf{h}}_{2,p}(t-1) \otimes \mathbf{I}_{L_1}\right]^T\mathbf{x}(t)$,

$$
\begin{aligned}
\underline{\widehat{\mathbf{h}}}_1(t) &= \left[\begin{array}{cccc} \widehat{\mathbf{h}}_{1,1}^T(t) & \widehat{\mathbf{h}}_{1,2}^T(t) & \cdots & \widehat{\mathbf{h}}_{1,P}^T(t) \end{array}\right]^T, \qquad (11)\\
\underline{\mathbf{x}}_2(t) &= \left[\begin{array}{cccc} \mathbf{x}_{2,1}^T(t) & \mathbf{x}_{2,2}^T(t) & \cdots & \mathbf{x}_{2,P}^T(t) \end{array}\right]^T,\\
\mathbf{x}_{1,p}(t) &= \left[\mathbf{I}_{L_2} \otimes \widehat{\mathbf{h}}_{1,p}(t-1)\right]^T\mathbf{x}(t),\\
\underline{\widehat{\mathbf{h}}}_2(t) &= \left[\begin{array}{cccc} \widehat{\mathbf{h}}_{2,1}^T(t) & \widehat{\mathbf{h}}_{2,2}^T(t) & \cdots & \widehat{\mathbf{h}}_{2,P}^T(t) \end{array}\right]^T, \qquad (12)\\
\underline{\mathbf{x}}_1(t) &= \left[\begin{array}{cccc} \mathbf{x}_{1,1}^T(t) & \mathbf{x}_{1,2}^T(t) & \cdots & \mathbf{x}_{1,P}^T(t) \end{array}\right]^T.
\end{aligned}
$$

Following the least-squares (LS) error criterion [1], we can define the cost functions:

$$\mathcal{J}_{\underline{\widehat{\mathbf{h}}}_2}\left[\underline{\widehat{\mathbf{h}}}_1(t)\right] = \sum_{i=1}^{t} \lambda_1^{t-i}\left[d(i) - \underline{\widehat{\mathbf{h}}}_1^T(t)\underline{\mathbf{x}}_2(i)\right]^2, \qquad (13)$$

$$\mathcal{J}_{\underline{\widehat{\mathbf{h}}}_1}\left[\underline{\widehat{\mathbf{h}}}_2(t)\right] = \sum_{i=1}^{t} \lambda_2^{t-i}\left[d(i) - \underline{\widehat{\mathbf{h}}}_2^T(t)\underline{\mathbf{x}}_1(i)\right]^2, \qquad (14)$$

where λ_1 $(0 \ll \lambda_1 < 1)$ and λ_2 $(0 \ll \lambda_2 < 1)$ are the forgetting factors. The minimization of $\mathcal{J}_{\underline{\widehat{\mathbf{h}}}_2}\left[\underline{\widehat{\mathbf{h}}}_1(t)\right]$ and $\mathcal{J}_{\underline{\widehat{\mathbf{h}}}_1}\left[\underline{\widehat{\mathbf{h}}}_2(t)\right]$ with respect to $\underline{\widehat{\mathbf{h}}}_1(t)$ and $\underline{\widehat{\mathbf{h}}}_2(t)$, respectively, lead to the normal equations:

$$\underline{\mathbf{R}}_2(t)\underline{\widehat{\mathbf{h}}}_1(t) = \underline{\mathbf{p}}_2(t), \qquad (15)$$

$$\underline{\mathbf{R}}_1(t)\underline{\widehat{\mathbf{h}}}_2(t) = \underline{\mathbf{p}}_1(t), \qquad (16)$$

where the terms from (15)–(16) can be recursively evaluated as

$$
\begin{aligned}
\underline{\mathbf{R}}_2(t) &= \lambda_1\underline{\mathbf{R}}_2(t-1) + \underline{\mathbf{x}}_2(t)\underline{\mathbf{x}}_2^T(t),\\
\underline{\mathbf{p}}_2(t) &= \lambda_1\underline{\mathbf{p}}_2(t-1) + \underline{\mathbf{x}}_2(t)d(t),\\
\underline{\mathbf{R}}_1(t) &= \lambda_2\underline{\mathbf{R}}_1(t-1) + \underline{\mathbf{x}}_1(t)\underline{\mathbf{x}}_1^T(t),\\
\underline{\mathbf{p}}_1(t) &= \lambda_2\underline{\mathbf{p}}_1(t-1) + \underline{\mathbf{x}}_1(t)d(t).
\end{aligned}
$$

From the previous equations, it is straightforward to derive the RLS algorithm based on the nearest Kronecker product decomposition, namely RLS-NKP, which is defined by the updates:

$$\underline{\widehat{\mathbf{h}}}_1(t) = \underline{\widehat{\mathbf{h}}}_1(t-1) + \mathbf{k}_2(t)e(t), \tag{17}$$

$$\underline{\widehat{\mathbf{h}}}_2(t) = \underline{\widehat{\mathbf{h}}}_2(t-1) + \mathbf{k}_1(t)e(t), \tag{18}$$

where the Kalman gain vectors can be computed as follows:

$$\mathbf{k}_2(t) = \frac{\underline{\mathbf{R}}_2^{-1}(t-1)\underline{\mathbf{x}}_2(t)}{\lambda_1 + \underline{\mathbf{x}}_2^T(t)\underline{\mathbf{R}}_2^{-1}(t-1)\underline{\mathbf{x}}_2(t)}, \tag{19}$$

$$\mathbf{k}_1(t) = \frac{\underline{\mathbf{R}}_1^{-1}(t-1)\underline{\mathbf{x}}_1(t)}{\lambda_2 + \underline{\mathbf{x}}_1^T(t)\underline{\mathbf{R}}_1^{-1}(t-1)\underline{\mathbf{x}}_1(t)}, \tag{20}$$

and $e(t)$ is defined in (9) or (10). Finally, using (17)–(18) and the related decompositions (11)–(12), the estimated impulse response can be evaluated as in (8). Also, based on the matrix inversion lemma [1], the updates of $\underline{\mathbf{R}}_2^{-1}(t)$ and $\underline{\mathbf{R}}_1^{-1}(t)$ result in

$$\underline{\mathbf{R}}_2^{-1}(t) = \lambda_1^{-1}\left[\underline{\mathbf{R}}_2^{-1}(t-1) - \mathbf{k}_2(t)\underline{\mathbf{x}}_2^T(t)\mathbf{R}_2^{-1}(t-1)\right], \tag{21}$$

$$\underline{\mathbf{R}}_1^{-1}(t) = \lambda_2^{-1}\left[\underline{\mathbf{R}}_1^{-1}(t-1) - \mathbf{k}_1(t)\underline{\mathbf{x}}_1^T(t)\mathbf{R}_1^{-1}(t-1)\right]. \tag{22}$$

As we can notice, the proposed RLS-NKP algorithm involves matrices of size $PL_1 \times PL_1$ and $PL_2 \times PL_2$, while the regular RLS algorithm would involve matrices of size $L_1L_2 \times L_1L_2$. Basically, a system identification problem of size $L = L_1L_2$ is transformed into two "smaller" problems of size PL_1 and PL_2, respectively. Consequently, when $P \ll L_2$, it is much more convenient to use the proposed RLS-NKP algorithm instead of the regular RLS algorithm. In this context, in order to maximize the gain in terms of complexity, it is reasonable to select the value of L_1 close to the value of L_2, such that the sum $L_1 + L_2$ should be as small as possible as compared to the product L_1L_2 [8]. Besides, since the proposed RLS-NKP algorithm uses two shorter filters of lengths PL_1 and PL_2, improved convergence and tracking capabilities are expected (as compared to the regular RLS algorithm, which uses a longer adaptive filter).

The overall complexity of the regular RLS algorithm [1] is $\mathcal{O}(L^2) = \mathcal{O}[(L_1L_2)^2]$, while the proposed RLS-NKP algorithm requires $\mathcal{O}[(PL_1)^2 + (PL_2)^2]$ operations. Clearly, for $P \ll L_2$, the computational amount of the RLS-NKP algorithm is much lower as compared to its regular counterpart. Of course, for large values of P (i.e., closer to L_2), the computational complexity of the RLS-NKP algorithm exceeds the conventional RLS algorithm. However, the proposed algorithm is not designed for such purposes, but for the identification of low-rank systems (e.g., echo paths), where the rank of the matrix $\mathbf{H}$ is usually much lower as compared to L_2.

4. SIMULATION RESULTS

Simulations are performed in the context of echo cancellation. The input signal $x(t)$ is either an AR(1) process [generated by filtering a white Gaussian noise through a first-order system $1/\left(1 - 0.9z^{-1}\right)$] or a speech sequence; the sampling rate is 8 kHz. The additive noise $w(t)$ is white and Gaussian. The signal-to-noise ratio (SNR), which is defined as $E[y^2(t)]/E[w^2(t)]$ (where $E[\cdot]$ denotes the expectation), is set to 20 dB. The performance measure is the normalized misalignment (in dB), defined as $20\log_{10}\left(\left\|\mathbf{h} - \widehat{\mathbf{h}}(t)\right\|_2 / \|\mathbf{h}\|_2\right)$.

In the experiments, we consider five impulse responses, as depicted in Fig. 1. The plot in Fig. 1(a) shows the first impulse response from G168 Recommendation [34] (i.e., a cluster of 64 coefficients padded with zeros). The second impulse response depicted in Fig. 1(b) is obtained by concatenating the first and fifth impulse responses from G168 Recommendation [34] (i.e., two clusters of 64 and 96 coefficients, respectively, which are padded with zeros). The burst impulse response from Fig. 1(c) contains a cluster of 64 random taps (with Gaussian distribution), while the rest are zeros. Similarly, the burst impulse response from Fig. 1(d) contains two clusters of 64 and 96 random taps. Each of these four impulse responses has the length $L = 500$, so that the decomposition can be performed using $L_1 = 25$ and $L_2 = 20$. Finally, the impulse response from Fig. 1(e) represents a typical acoustic echo path of length $L = 1024$, so that we can use $L_1 = L_2 = 32$.

These impulse responses have different sparseness degrees, which can be evaluated using a sparseness measure based on the ℓ_1 and ℓ_2 norms [32] (denoted by ξ_{12} in the caption of Fig. 1). The closer this measure is to 1, the sparser is the impulse response; on the contrary, the closer the measure is to 0, the denser or less sparse is the impulse response. However, as outlined in the previous sections, the sparsity of the system helps (by reducing the rank of the matrix $\mathbf{H}$), but it is not the only factor. The main issue is related to the decomposition and, consequently, to the low-rank approximation of the impulse response. In this regard, the singular values of $\mathbf{H}$ (nor-

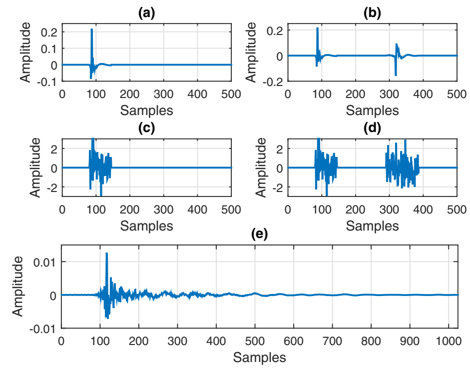

Fig. 1. Impulse responses used in the experiments: (a) $L = 500$, $\xi_{12} = 0.8957$; (b) $L = 500$, $\xi_{12} = 0.8080$; (c) $L = 500$, $\xi_{12} = 0.7549$; (d) $L = 500$, $\xi_{12} = 0.5795$; and (e) $L = 1024$, $\xi_{12} = 0.6880$.

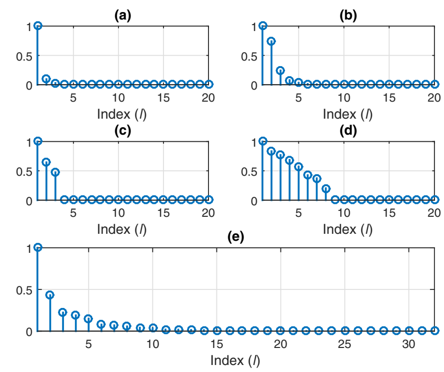

Fig. 2. Singular values (normalized with respect to the maximum one) of $\mathbf{H}$ for the corresponding impulse responses from Fig. 1: (a)–(d) $L_1 = 25$, $L_2 = 20$; and (e) $L_1 = L_2 = 32$.

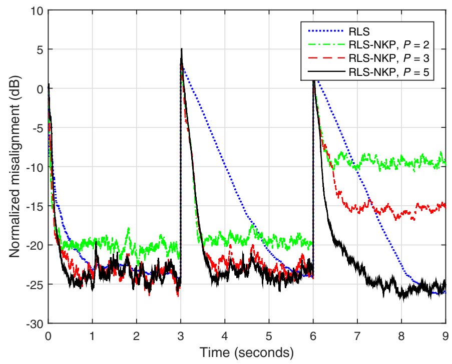

Fig. 3. Misalignment of the RLS algorithm ($L = 500$) and RLS-NKP algorithm (using $L_1 = 25$, $L_2 = 20$, and $P < L_2$), for the identification of the impulse responses from Figs. 1(a) and (b). The input signal is an AR(1) process and the impulse response changes at times 3 and 6 seconds.

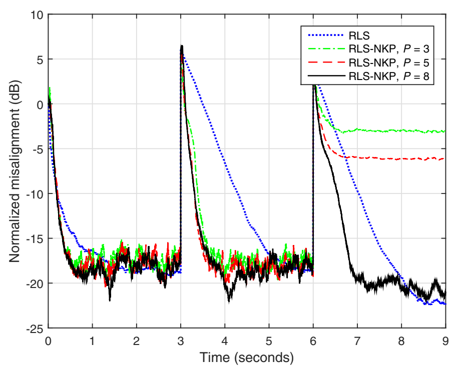

Fig. 4. Misalignment of the RLS algorithm ($L = 500$) and RLS-NKP algorithm (using $L_1 = 25$, $L_2 = 20$, and $P < L_2$), for the identification of the impulse responses from Figs. 1(c) and (d). The input signal is an AR(1) process and the impulse response changes at times 3 and 6 seconds.

malized with respect to the maximum one) are depicted in Fig. 2, for the corresponding impulse responses from Fig. 1. The faster these values decrease to zero, the better is the approximation behind (5).

Next, we focus on the identification of the impulse responses from Figs. 1(a)–(d). The proposed RLS-NKP algorithm (using $L_1 = 25$, $L_2 = 20$, and different values of P) is compared to the regular RLS algorithm (with $L = 500$). For a fair comparison, the forgetting factors of the RLS-NKP algorithm are set to $\lambda_1 = 1 - 1/(KPL_1)$ and $\lambda_2 = 1 - 1/(KPL_2)$, while the regular RLS algorithm uses the forgetting factor $\lambda = 1 - 1/(KL)$, where K is a positive integer. In these experiments, the input signal is an AR(1) process and $K = 10$. In order to test the tracking capabilities of the algorithms, two abrupt changes of the impulse responses are introduced in each simulation. First, in Fig. 3, the impulse response from Fig. 1(a) is shifted to the right by 12 samples after 3 seconds; then, at time 6 seconds, the impulse response from Fig. 1(b) is used. Second, in Fig. 4, an abrupt change of the impulse response from Fig. 1(c) is introduced after 3 seconds, by changing the sign of the coefficients; also, at time 6 seconds, the impulse response from Fig. 1(d) is involved.

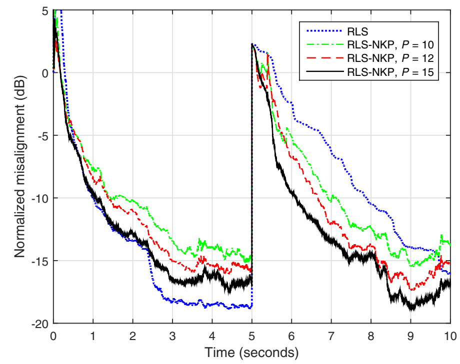

Fig. 5. Misalignment of the RLS algorithm ($L = 1024$) and RLS-NKP algorithm (using $L_1 = L_2 = 32$ and $P < L_2$), for the identification of the impulse responses from Fig. 1(e). The input signal is a speech sequence and the impulse response changes at time 5 seconds.

As we can notice in the first two parts of Fig. 3 [using the impulse response from Fig. 1(a), with $\text{rank}(\mathbf{H}) = 3$], the RLS-NKP algorithm using $P \geq 3$ achieves a similar initial convergence rate and steady-state misalignment as the regular RLS algorithm, but it performs better in terms of tracking. The same tracking reaction can be observed in case of the RLS-NKP algorithm using $P = 2$, which achieves a slightly higher misalignment level. Similar conclusions result from the last part of Fig. 3, where the impulse response from Fig. 1(b) is identified. However, since the rank of the matrix $\mathbf{H}$ is higher in this case, the value of P should be larger, e.g., $P = 5$. Also, the same better tracking reaction can be noticed in case of the RLS-NKP algorithm as compared to its regular counterpart.

The results from Fig. 4 follow a similar pattern, even if the impulse responses from Figs. 1(c) and (d) are less sparse. Basically, the impulse response from Fig. 1(c) can be estimated using $P = 3$ [since $\text{rank}(\mathbf{H}) = 3$], while the value of P should be higher in case of the impulse response from Fig. 1(d), where $\text{rank}(\mathbf{H}) = 8$.

Finally, the acoustic impulse response from Fig. 1(e) is identified, using a speech signal as input. The results are reported in Fig. 5, where the RLS-NKP algorithm (using $L_1 = L_2 = 32$ and different values of P) is compared to the RLS algorithm (with $L = 1024$). In this case, the matrix $\mathbf{H}$ is closer to full-rank, so that we obtain an approximative solution when $P < L_2$. However, a reasonable attenuation of the misalignment is achieved for a reasonable low value of P (as compared to L_2). Also, the tracking capabilities of the RLS-NKP algorithm are better as compared to its regular counterpart.

5. CONCLUSIONS

In this paper, we have proposed the RLS-NKP algorithm, which exploits the nearest Kronecker product decomposition of the impulse response, together with low-rank approximation. Consequently, this algorithm is suitable for the identification of low-rank models, like the echo paths. As compared to the regular RLS algorithm, the gain of the proposed solution is twofold. First, the computational complexity of the RLS-NKP algorithm could be much lower as compared to its regular counterpart. Second, the tracking capabilities of the proposed algorithm are better as compared to the conventional RLS algorithm. Simulations performed in the context of echo cancellation support these advantages.

6. REFERENCES

[1] S. Haykin, *Adaptive Filter Theory*. Fourth Edition, Upper Saddle River, NJ: Prentice-Hall, 2002.

[2] S. L. Gay and J. Benesty, Eds., *Acoustic Signal Processing for Telecommunication*. Boston, MA: Kluwer Academic Publisher, 2000.

[3] J. Benesty, T. Gaensler, D. R. Morgan, M. M. Sondhi, and S. L. Gay, *Advances in Network and Acoustic Echo Cancellation*. Berlin, Germany: Springer-Verlag, 2001.

[4] Y. V. Zakharov, G. P. White, and J. Liu, "Low-complexity RLS algorithms using dichotomous coordinate descent iterations," *IEEE Trans. Signal Processing*, vol. 56, pp. 3150–3161, July 2008.

[5] Y. V. Zakharov and V. H. Nascimento, "DCD-RLS adaptive filters with penalties for sparse identification," *IEEE Trans. Signal Processing*, vol. 61, pp. 3198–3213, June 2013.

[6] Y. V Zakharov, V. H. Nascimento, R. C De Lamare, F. G. De Almeida Neto, "Low-complexity DCD-based sparse recovery algorithms," *IEEE Access*, vol. 5, pp. 12737–12750, July 2017.

[7] Y. Zhang, Y. V. Zakharov, and J. Li, "Soft-decision-driven sparse channel estimation and turbo equalization for MIMO underwater acoustic communications," *IEEE Access*, vol. 6, pp. 4955–4973, Jan. 2018.

[8] 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.

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

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

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

[12] R. Abrahamsson, S. M. Kay, and P. Stoica, "Estimation of the parameters of a bilinear model with applications to submarine detection and system identification," *Digital Signal Processing*, vol. 17, pp. 756–773, July 2007.

[13] N. Vervliet, O. Debals, L. Sorber, and L. De Lathauwer, "Breaking the curse of dimensionality using decompositions of incomplete tensors: Tensor-based scientific computing in big data analysis," *IEEE Signal Processing Mag.*, vol. 31, pp. 71–79, Sept. 2014.

[14] A. Cichocki, D. Mandic, L. De Lathauwer, G. Zhou, Q. Zhao, C. Caiafa, and A. H. Phan, "Tensor decompositions for signal processing applications: From two-way to multiway component analysis," *IEEE Signal Processing Mag.*, vol. 32, pp. 145–163, Mar. 2015.

[15] L. N. Ribeiro, A. L. F. de Almeida, and J. C. M. Mota, "Identification of separable systems using trilinear filtering," in *Proc. IEEE CAMSAP*, 2015, pp. 189–192.

[16] A. P. da Silva, P. Comon, and A. L. F. de Almeida, "A finite algorithm to compute rank-1 tensor approximations," *IEEE Signal Processing Lett.*, vol. 23, pp. 959–963, July 2016.

[17] M. Bousse, O. Debals, and L. De Lathauwer, "A tensor-based method for large-scale blind source separation using segmentation," *IEEE Trans. Signal Processing*, vol. 65, pp. 346–358, Jan. 2017.

[18] 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.

[19] N. Sidiropoulos, L. De Lathauwer, X. Fu, K. Huang, E. Papalexakis, and C. Faloutsos, "Tensor decomposition for signal processing and machine learning," *IEEE Trans. Signal Processing*, vol. 65, pp. 3551–3582, July 2017.

[20] L. N. Ribeiro, S. Schwarz, M. Rupp, A. L. F. de Almeida, and J. C. M. Mota, "A low-complexity equalizer for massive MIMO systems based on array separability," in *Proc. EUSIPCO*, 2017, pp. 2522–2526.

[21] M. N. da Costa, G. Favier, and J. M. T. Romano, "Tensor modelling of MIMO communication systems with performance analysis and Kronecker receivers," *Signal Processing*, vol. 145, pp. 304–316, Apr. 2018.

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

[23] M. Rupp and S. Schwarz, "Gradient-based approaches to learn tensor products," in *Proc. EUSIPCO*, 2015, pp. 2486–2490.

[24] C. Paleologu, J. Benesty, and S. Ciochină, "Adaptive filtering for the identification of bilinear forms," *Digital Signal Processing*, vol. 75, pp. 153–167, Apr. 2018.

[25] L. M. Dogariu, C. Paleologu, S. Ciochină, J. Benesty, and P. Piantanida, "Identification of bilinear forms with the Kalman filter," in *Proc. IEEE ICASSP*, 2018, pp. 4134–4138.

[26] C. Elisei-Iliescu, C. Stanciu, C. Paleologu, J. Benesty, C. Anghel, and S. Ciochină, "Low-complexity RLS algorithms for the identification of bilinear forms," in *Proc. EUSIPCO*, 2018, pp. 460–464.

[27] C. Elisei-Iliescu, C. Stanciu, C. Paleologu, J. Benesty, C. Anghel, and S. Ciochină, "Efficient recursive least-squares algorithms for the identification of bilinear forms," *Digital Signal Processing*, vol. 83, pp. 280–296, Dec. 2018.

[28] G. H. Golub and C. F. Van Loan, *Matrix Computations, 3rd Edition*. Baltimore, MD: The John Hopkins University Press, 1996.

[29] D. A. Harville, *Matrix Algebra From a Statistician's Perspective*. New York: Springer-Verlag, 1997.

[30] D. L. Duttweiler, "Proportionate normalized least-mean-squares adaptation in echo cancelers," *IEEE Trans. Speech, Audio Processing*, vol. 8, pp. 508–518, Sept. 2000.

[31] J. Benesty and S. L. Gay, "An improved PNLMS algorithm," in *Proc. IEEE ICASSP*, 2002, pp. II-1881–II-1884.

[32] C. Paleologu, J. Benesty, and S. Ciochină, *Sparse Adaptive Filters for Echo Cancellation*. Morgan & Claypool Publishers, 2010.

[33] J. Liu and S. L. Grant, "Proportionate adaptive filtering for block-sparse system identification," *IEEE/ACM Trans. Audio, Speech, Language Processing*, vol. 24, pp. 623–630, Apr. 2016.

[34] *Digital Network Echo Cancellers*, ITU-T Recommendations G.168, 2002.