# An Iterative Multichannel Wiener Filter Based on a Kronecker Product Decomposition

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

Constantin Paleologu, Claudia-Cristina Oprea, Silviu Ciochină
*Department of Telecommunications*
*University Politehnica of Bucharest*
Bucharest, Romania
{pale, cristina, silviu}@comm.pub.ro

***Abstract*—Multiple-input single-output (MISO) system identification problems appear in the context of many important applications. Due to their nature, they are usually addressed based on multichannel algorithms. However, the identification of long length impulse responses (e.g., like in echo cancellation) raises significant challenges, especially in terms of complexity and accuracy of the solution. In this paper, we develop an iterative multichannel Wiener filter for such MISO system identification scenarios. This algorithm is based on a Kronecker product decomposition of the impulse response, in conjunction with low-rank approximations. Simulation results indicate a good accuracy of the proposed solution, even when a small amount of data is available for the estimation of the statistics.**

***Index Terms*—Echo cancellation, Kronecker product decomposition, low-rank approximation, multichannel Wiener filter, multiple-input single-output (MISO) system identification.**

## I. Introduction

The celebrated Wiener filter is widely used for solving system identification problems. Moreover, it represents a benchmark for the related solutions based on adaptive filters [1]. Nevertheless, the performance of the Wiener filter relies heavily on the estimates of some signal statistics, i.e., the covariance matrix of the input and the cross-correlation vector (between the input and the reference signal). Furthermore, the number of data samples required to this purpose is basically related to the filter length. Consequently, a challenging situation appears when a small amount of data is available for the estimation of the statistics. This aspect becomes critical in case of long length impulse responses and multichannel systems.

Recently, an efficient approach to handle this challenge was proposed in [2], in the framework of echo cancellation. The basic idea is to exploit a Kronecker product decomposition of the impulse response, together with low-rank approximations. It also relies on the identification of bilinear forms, as previously indicated in [3]. These techniques are mainly related to tensor decomposition and modelling, which allow to reformulate different high-dimension system identification problems and finally combine low-dimension solutions that match the original purpose, e.g., see [4]– [7] and the references therein. In this context, one of the main features of the solution proposed in [2] is that it is applicable not only for perfectly decomposable systems, but for more general (and realistic) forms of impulse responses, like the echo paths.

The iterative Wiener filter developed in [2] is able to obtain accurate results using less amount of data (as compared to the filter length). However, in [2], the framework was limited to a single-channel echo cancellation scenario, which is basically a single-input single-output system identification problem. A more challenging case would be the identification of multiple-input single-output (MISO) systems, like in the framework of stereophonic acoustic echo cancellation (SAEC) [8]. This represents the motivation behind the development of the multichannel Wiener filter proposed in this paper.

In the following, Section II presents the system model and different equivalent forms, together with the related "best" (low-rank) approximation. The proposed iterative multichannel Wiener filter is developed in Section III. Its performance is analyzed in Section IV, based on simulations performed in the context of multichannel echo cancellation, including SAEC. Finally, conclusions are summarized in Section V.

## II. System Model and Best Approximation

Let us consider the framework of a linear MISO (LMISO) system identification problem, with $M$ inputs (or channels) in the system. The channel impulse responses are denoted by $\mathbf{h}_{\mathrm{c},m}$, with $m = 1, 2, \ldots, M$, and each one of them is of length $L$. Also, $\mathbf{x}_m(k) = \begin{bmatrix} x_m(k) & x_m(k-1) & \cdots & x_m(k-L+1) \end{bmatrix}^T$ denotes a vector containing the most recent $L$ time samples of the $m$th ($m = 1, 2, \ldots, M$) input signal, $x_m(k)$, where $k$ is the discrete-time index and the superscript $^T$ is the transpose operator. In this context, the reference signal is given by

$$d(k) = \sum_{m=1}^{M} \mathbf{h}_{\mathrm{c},m}^T \mathbf{x}_m(k) + w(k), \tag{1}$$

where $w(k)$ is the additive noise. In this work, it is assumed that all random (input, output, and noise) signals are real valued and zero mean. Also, it is customary to consider that $x_m(k)$ ($m = 1, 2, \ldots, M$) and $w(k)$ are uncorrelated.

The reference signal from (1) can be rewritten as

$$d(k) = \overline{\mathbf{h}}^T \overline{\mathbf{x}}(k) + w(k), \tag{2}$$

This work was supported by a grant of the Romanian Ministry of Research and Innovation, CNCS – UEFISCDI, project number: PN-III-P1-1.1-TE-2019-0420, within PNCDI III.

where $\overline{\mathbf{h}} = \left[ \begin{array}{ccc} \mathbf{h}_{c,1}^T & \cdots & \mathbf{h}_{c,M}^T \end{array} \right]^T$ and $\overline{\mathbf{x}}(k) = \left[ \begin{array}{ccc} \mathbf{x}_1^T(k) & \cdots & \mathbf{x}_M^T(k) \end{array} \right]^T$. Next, assuming that $L = L_1 L_2$ (e.g., with $ML_1 \geq L_2$), the $m$th ($m = 1, 2, \ldots, M$) input vector can be decomposed as $\mathbf{x}_m(k) = \left[ \begin{array}{cccc} \mathbf{x}_m'^T(k) & \mathbf{x}_m'^T(k - L_1) & \cdots & \mathbf{x}_m'^T[k - (L_2 - 1)L_1] \end{array} \right]^T$, where $\mathbf{x}_m'(k) = \left[ \begin{array}{ccc} x_m(k) & \cdots & x_m(k - L_1 + 1) \end{array} \right]^T$. Furthermore, we can combine the $M$ input signals as $\underline{\mathbf{x}}(k) = \left[ \begin{array}{cccc} \underline{\mathbf{x}}'^T(k) & \underline{\mathbf{x}}'^T(k - L_1) & \cdots & \underline{\mathbf{x}}'^T[k - (L_2 - 1)L_1] \end{array} \right]^T$, where $\underline{\mathbf{x}}'(k) = \left[ \begin{array}{ccc} \mathbf{x}_1'^T(k) & \cdots & \mathbf{x}_M'^T(k) \end{array} \right]^T$ is a vector of length $ML_1$. Therefore, an equivalent way to write the LMISO system in (1) or (2) is

$$d(k) = \underline{\mathbf{h}}^T \underline{\mathbf{x}}(k) + w(k) = y(k) + w(k), \quad (3)$$

where $\underline{\mathbf{h}}$ is the system spatiotemporal impulse response of length $ML$, whose coefficients are the same of those of $\overline{\mathbf{h}}$, obtained by simple permutations according to the inputs.

As we can notice, $y(k) = \underline{\mathbf{h}}^T \underline{\mathbf{x}}(k)$ is linear in $\underline{\mathbf{h}}$. On the other hand, a bilinear MISO (BMISO) system can be derived from the LMISO system in (1) and according to (3). Let us form the matrix of size $ML_1 \times L_2$, $\underline{\mathbf{X}}(k) = \mathrm{ivec}\left[\underline{\mathbf{x}}(k)\right]$, where $\mathrm{ivec}[\cdot]$ denotes the conversion of a vector into a matrix (i.e., the inverse of the vectorization operation) [7]. Consequently, the reference signal is given by

$$d(k) = \sum_{l=1}^{L_2} \mathbf{h}_{\mathrm{st1},l}^T \underline{\mathbf{X}}(k) \mathbf{h}_{\mathrm{t2},l} + w(k) = y(k) + w(k), \quad (4)$$

where $\mathbf{h}_{\mathrm{st1},l},\ l = 1, 2, \ldots, L_2$ is the set of the system spatiotemporal impulse responses of length $ML_1$ and $\mathbf{h}_{\mathrm{t2},l},\ l = 1, 2, \ldots, L_2$ is the set of the system temporal impulse responses of length $L_2$. In (4), the bilinear form of this BMISO system can be expressed as

$$y(k) = \sum_{l=1}^{L_2} \mathbf{h}_{\mathrm{st1},l}^T \underline{\mathbf{X}}(k) \mathbf{h}_{\mathrm{t2},l} = \mathrm{vec}^T\left(\mathbf{H}_{\mathrm{st}}\right) \underline{\mathbf{x}}(k), \quad (5)$$

where $\mathrm{vec}[\cdot]$ is the vectorization operation (which consists of converting a matrix into a vector) and $\mathbf{H}_{\mathrm{st}} = \sum_{l=1}^{L_2} \mathbf{h}_{\mathrm{st1},l} \mathbf{h}_{\mathrm{t2},l}^T$ is a matrix of size $ML_1 \times L_2$ of rank equal to $L_2$ in general. In (3), we can decompose the spatiotemporal impulse response of the LMISO system as $\underline{\mathbf{h}} = \left[ \begin{array}{ccc} \underline{\mathbf{h}}_1^T & \cdots & \underline{\mathbf{h}}_{L_2}^T \end{array} \right]^T$, where $\underline{\mathbf{h}}_l,\ l = 1, 2, \ldots, L_2$ are impulse responses of length $ML_1$ each. Thus, the linear term can be written as

$$y(k) = \underline{\mathbf{h}}^T \underline{\mathbf{x}}(k) = \mathrm{vec}^T\left(\mathbf{H}_{\mathrm{st}}\right) \underline{\mathbf{x}}(k), \quad (6)$$

with $\mathbf{H}_{\mathrm{st}} = \mathrm{ivec}\left(\underline{\mathbf{h}}\right)$. As a result, by simple identification between (5) and (6), we deduce that the LMISO and BMISO systems are equivalent.

In this framework, our objective is the identification of the LMISO system in (1) [or, equivalently, in (2) or (3)]. This identification is done in light with what we know on bilinear forms and how they are best approximated [2], [3]. Let us consider the impulse response of the BMISO system in (4), i.e., the matrix $\mathbf{H}_{\mathrm{st}} = \sum_{l=1}^{L_2} \mathbf{h}_{\mathrm{st1},l} \mathbf{h}_{\mathrm{t2},l}^T$ of size $ML_1 \times L_2$, with $ML_1 \geq L_2$ [see (5)]. This system is equivalent to the LMISO system in (3) as we already know. Using the singular value decomposition (SVD), $\mathbf{H}_{\mathrm{st}}$ can be factorized as

$$\mathbf{H}_{\mathrm{st}} = \mathbf{U} \boldsymbol{\Sigma} \mathbf{V}^T = \sum_{l=1}^{L_2} \sigma_l \mathbf{u}_l \mathbf{v}_l^T, \quad (7)$$

where $\mathbf{U}$ and $\mathbf{V}$ are two orthogonal matrices of sizes $ML_1 \times ML_1$ and $L_2 \times L_2$, respectively, and $\boldsymbol{\Sigma}$ is an $ML_1 \times L_2$ rectangular diagonal matrix with nonnegative real numbers on the main diagonal. The columns of $\mathbf{U}$ (resp. $\mathbf{V}$) are called the left-singular (resp. right-singular) vectors of $\mathbf{H}_{\mathrm{st}}$, while the diagonal entries $\sigma_l,\ l = 1, 2, \ldots, L_2$ of $\boldsymbol{\Sigma}$ are known as the singular values of $\mathbf{H}_{\mathrm{st}}$ with $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_{L_2} \geq 0$. We deduce that $\mathbf{h}_{\mathrm{st1},l} = \alpha(\sigma_l) \mathbf{u}_l$ and $\mathbf{h}_{\mathrm{t2},l} = \beta(\sigma_l) \mathbf{v}_l$, for $l = 1, 2, \ldots, L_2$, where $\alpha(\sigma_l)\beta(\sigma_l) = \sigma_l$, $\mathbf{u}_l$ are the first $L_2$ columns of $\mathbf{U}$, and $\mathbf{v}_l$ are the columns of $\mathbf{V}$. It is easy to verify[1] that $\left\|\mathbf{H}_{\mathrm{st}}\right\|_{\mathrm{F}} = \sqrt{\sum_{l=1}^{L_2} \sigma_l^2}$ and $\left\|\mathbf{H}_{\mathrm{st}}\right\|_2 = \sigma_1$.

In many practical applications, because of the redundancies in $\mathbf{H}_{\mathrm{st}}$ due to the reflections and/or sparseness in the system, this matrix is never really full rank. Under these circumstances, let $P \ll L_2$ and let us define the matrix:

$$\mathbf{H}_{\mathrm{st}}(P) = \sum_{p=1}^{P} \sigma_p \mathbf{u}_p \mathbf{v}_p^T. \quad (8)$$

The main and important question is if $\mathbf{H}_{\mathrm{st}}$ can be well approximated by $\mathbf{H}_{\mathrm{st}}(P)$. In the positive scenario, we could express the LMISO system as

$$\begin{aligned} d(k) &= \mathrm{vec}^T\left(\mathbf{H}_{\mathrm{st}}\right) \underline{\mathbf{x}}(k) + w(k) \\ &= \mathrm{vec}^T\left[\mathbf{H}_{\mathrm{st}}(P)\right] \underline{\mathbf{x}}(k) + b(k) + w(k), \end{aligned} \quad (9)$$

where $b(k) = \mathrm{vec}^T\left(\boldsymbol{\Upsilon}\right) \underline{\mathbf{x}}(k)$ is some negligible correlated noise, with $\boldsymbol{\Upsilon} = \sum_{i=P+1}^{L_2} \sigma_i \mathbf{u}_i \mathbf{v}_i^T$. Then, our objective would be to identify $\mathbf{H}_{\mathrm{st}}(P)$ instead of $\mathbf{H}_{\mathrm{st}}$ (as explained in the next section, there might be some advantages with this new approach). The answer to the above question is given by a theorem shown in [9], [10], which can be stated as follows. Let $\mathrm{rank}\left(\mathbf{H}_{\mathrm{st}}\right) = R \leq L_2$ and let $\mathcal{S}$ be the set of $ML_1 \times L_2$ matrices with a rank equal to $P < R$. Then, the solution of

$$\min_{\mathbf{H} \in \mathcal{S}} \left\|\mathbf{H}_{\mathrm{st}} - \mathbf{H}\right\|_2 \quad \text{or} \quad \min_{\mathbf{H} \in \mathcal{S}} \left\|\mathbf{H}_{\mathrm{st}} - \mathbf{H}\right\|_{\mathrm{F}} \quad (10)$$

is given by (8). Furthermore, we have $\min_{\mathbf{H} \in \mathcal{S}} \left\|\mathbf{H}_{\mathrm{st}} - \mathbf{H}\right\|_2 = \left\|\mathbf{H}_{\mathrm{st}} - \mathbf{H}_{\mathrm{st}}(P)\right\|_2 = \sigma_{P+1}$ and $\min_{\mathbf{H} \in \mathcal{S}} \left\|\mathbf{H}_{\mathrm{st}} - \mathbf{H}\right\|_{\mathrm{F}} = \left\|\mathbf{H}_{\mathrm{st}} - \mathbf{H}_{\mathrm{st}}(P)\right\|_{\mathrm{F}} = \sqrt{\sum_{i=P+1}^{L_2} \sigma_i^2}$. Therefore, as long as the normalized misalignment, $\mathcal{M}(P) = \left\|\mathbf{H}_{\mathrm{st}} - \mathbf{H}_{\mathrm{st}}(P)\right\|_{\mathrm{F}} / \left\|\mathbf{H}_{\mathrm{st}}\right\|_{\mathrm{F}}$, is very small, it is enough in practice to estimate the impulse responses $\mathbf{h}_{\mathrm{st1},p}$ and $\mathbf{h}_{\mathrm{t2},p}$ for $p = 1, 2, \ldots, P$.

[1] The Frobenius norm and the 2-norm of a real-valued rectangular matrix $\mathbf{A}$ are, respectively, $\left\|\mathbf{A}\right\|_{\mathrm{F}} = \sqrt{\mathrm{tr}\left(\mathbf{A}^T \mathbf{A}\right)}$ and $\left\|\mathbf{A}\right\|_2 = \max_{\left\|\mathbf{x}\right\|_2 = 1} \left\|\mathbf{A}\mathbf{x}\right\|_2$.

## III. Multichannel Wiener Filter

The classical concern in the identification of the LMISO system [see (3)] is to estimate $\underline{\mathbf{h}}$ with a real-valued filter, $\widehat{\underline{\mathbf{h}}}$, of length $ML$. It makes sense to define the error signal:

$$e(k) = d(k) - \widehat{y}(k), \tag{11}$$

where $\widehat{y}(k) = \widehat{\underline{\mathbf{h}}}^T \underline{\mathbf{x}}(k)$. Then, the optimal filter is found from the mean-squared error (MSE) criterion:

$$\mathcal{J}\left(\widehat{\underline{\mathbf{h}}}\right) = E\left[e^2(k)\right] = \sigma_d^2 - 2\widehat{\underline{\mathbf{h}}}^T \mathbf{p} + \widehat{\underline{\mathbf{h}}}^T \mathbf{R}\widehat{\underline{\mathbf{h}}}, \tag{12}$$

where $E[\cdot]$ denotes mathematical expectation, $\mathbf{p} = E\left[\underline{\mathbf{x}}(k)d(k)\right]$ is the cross-correlation vector between $\underline{\mathbf{x}}(k)$ and $d(k)$, and $\mathbf{R} = E\left[\underline{\mathbf{x}}(k)\underline{\mathbf{x}}^T(k)\right]$ is the covariance matrix of $\underline{\mathbf{x}}(k)$. From the minimization of $\mathcal{J}\left(\widehat{\underline{\mathbf{h}}}\right)$, we easily get the celebrated (multichannel) Wiener filter:

$$\widehat{\underline{\mathbf{h}}}_{\mathrm{W}} = \mathbf{R}^{-1}\mathbf{p}. \tag{13}$$

From the previous expression, we observe that we deal with a large matrix of size $ML \times ML$ and, to hope to have a reliable Wiener filter, we need much more than $ML$ data samples.

Instead of identifying the LMISO system in (3) and estimate $\underline{\mathbf{h}}$ as in the conventional approach, we can, equivalently, identify the LMISO system in (9) and estimate $\mathbf{H}_{\mathrm{st}}$. In the rest of the paper, we drop the subscripts $_{\mathrm{st}}$ and $_{\mathrm{t}}$ to simplify the notation, so that $\mathbf{H}_{\mathrm{st}} = \sum_{l=1}^{L_2} \mathbf{h}_{\mathrm{st}1,l}\mathbf{h}_{\mathrm{t}2,l}^T$ becomes $\mathbf{H} = \sum_{l=1}^{L_2} \mathbf{h}_{1,l}\mathbf{h}_{2,l}^T$. Let us assume that $\mathrm{rank}(\mathbf{H}) = P \ll L_2$, so that $\underline{\mathbf{h}}$ can be decomposed as

$$\underline{\mathbf{h}} = \sum_{p=1}^{P} \mathbf{h}_{2,p} \otimes \mathbf{h}_{1,p}, \tag{14}$$

where $\mathbf{h}_{1,p}$ and $\mathbf{h}_{2,p}$ are impulse responses of lengths $ML_1$ and $L_2$, respectively, and $\otimes$ is the Kronecker product [11]. As a consequence, we can choose to decompose the filter $\widehat{\underline{\mathbf{h}}}$ as

$$\widehat{\underline{\mathbf{h}}} = \sum_{p=1}^{P} \widehat{\mathbf{h}}_{2,p} \otimes \widehat{\mathbf{h}}_{1,p}, \tag{15}$$

where $\widehat{\mathbf{h}}_{1,p}$ and $\widehat{\mathbf{h}}_{2,p}$ are filters of lengths $ML_1$ and $L_2$, respectively. From the following relationships [12]:

$$\widehat{\mathbf{h}}_{2,p} \otimes \widehat{\mathbf{h}}_{1,p} = \left(\widehat{\mathbf{h}}_{2,p} \otimes \mathbf{I}_{ML_1}\right)\widehat{\mathbf{h}}_{1,p} = \left(\mathbf{I}_{L_2} \otimes \widehat{\mathbf{h}}_{1,p}\right)\widehat{\mathbf{h}}_{2,p},$$

where $\mathbf{I}_{ML_1}$ and $\mathbf{I}_{L_2}$ are the identity matrices of sizes $ML_1 \times ML_1$ and $L_2 \times L_2$, respectively, we can express (15) as

$$\widehat{\underline{\mathbf{h}}} = \sum_{p=1}^{P} \widehat{\mathbf{H}}_{2,p}\widehat{\mathbf{h}}_{1,p} = \sum_{p=1}^{P} \widehat{\mathbf{H}}_{1,p}\widehat{\mathbf{h}}_{2,p}, \tag{16}$$

where $\widehat{\mathbf{H}}_{2,p} = \widehat{\mathbf{h}}_{2,p} \otimes \mathbf{I}_{ML_1}$ and $\widehat{\mathbf{H}}_{1,p} = \mathbf{I}_{L_2} \otimes \widehat{\mathbf{h}}_{1,p}$ are matrices of sizes $ML \times ML_1$ and $ML \times L_2$, respectively. Thus, the error signal in (11) can be expressed into two different manners:

$$\begin{aligned} e(k) &= d(k) - \sum_{p=1}^{P} \widehat{\mathbf{h}}_{1,p}^T \widehat{\mathbf{H}}_{2,p}^T \underline{\mathbf{x}}(k) \qquad (17) \\ &= d(k) - \sum_{p=1}^{P} \widehat{\mathbf{h}}_{1,p}^T \mathbf{x}_{2,p}(k) = d(k) - \widehat{\underline{\mathbf{h}}}_1^T \underline{\mathbf{x}}_2(k), \\ e(k) &= d(k) - \sum_{p=1}^{P} \widehat{\mathbf{h}}_{2,p}^T \widehat{\mathbf{H}}_{1,p}^T \underline{\mathbf{x}}(k) \qquad (18) \\ &= d(k) - \sum_{p=1}^{P} \widehat{\mathbf{h}}_{2,p}^T \mathbf{x}_{1,p}(k) = d(k) - \widehat{\underline{\mathbf{h}}}_2^T \underline{\mathbf{x}}_1(k), \end{aligned}$$

where

$$\begin{aligned} \mathbf{x}_{2,p}(k) &= \widehat{\mathbf{H}}_{2,p}^T \underline{\mathbf{x}}(k),\ \widehat{\underline{\mathbf{h}}}_1 = \begin{bmatrix} \widehat{\mathbf{h}}_{1,1}^T & \widehat{\mathbf{h}}_{1,2}^T & \cdots & \widehat{\mathbf{h}}_{1,P}^T \end{bmatrix}^T, \\ \underline{\mathbf{x}}_2(k) &= \begin{bmatrix} \mathbf{x}_{2,1}^T(k) & \mathbf{x}_{2,2}^T(k) & \cdots & \mathbf{x}_{2,P}^T(k) \end{bmatrix}^T, \\ \mathbf{x}_{1,p}(k) &= \widehat{\mathbf{H}}_{1,p}^T \underline{\mathbf{x}}(k),\ \widehat{\underline{\mathbf{h}}}_2 = \begin{bmatrix} \widehat{\mathbf{h}}_{2,1}^T & \widehat{\mathbf{h}}_{2,2}^T & \cdots & \widehat{\mathbf{h}}_{2,P}^T \end{bmatrix}^T, \\ \underline{\mathbf{x}}_1(k) &= \begin{bmatrix} \mathbf{x}_{1,1}^T(k) & \mathbf{x}_{1,2}^T(k) & \cdots & \mathbf{x}_{1,P}^T(k) \end{bmatrix}^T. \end{aligned}$$

Thus, we can write the MSE criterion as

$$\begin{aligned} \mathcal{J}\left(\widehat{\underline{\mathbf{h}}}_1, \widehat{\underline{\mathbf{h}}}_2\right) &= \sigma_d^2 - 2\widehat{\underline{\mathbf{h}}}_1^T \underline{\mathbf{p}}_2 + \widehat{\underline{\mathbf{h}}}_1^T \underline{\mathbf{R}}_2 \widehat{\underline{\mathbf{h}}}_1 \qquad (19) \\ &= \sigma_d^2 - 2\widehat{\underline{\mathbf{h}}}_2^T \underline{\mathbf{p}}_1 + \widehat{\underline{\mathbf{h}}}_2^T \underline{\mathbf{R}}_1 \widehat{\underline{\mathbf{h}}}_2, \qquad (20) \end{aligned}$$

where $\underline{\mathbf{p}}_2 = \begin{bmatrix} \mathbf{p}^T\widehat{\mathbf{H}}_{2,1} & \mathbf{p}^T\widehat{\mathbf{H}}_{2,2} & \cdots & \mathbf{p}^T\widehat{\mathbf{H}}_{2,P} \end{bmatrix}^T$,

$$\underline{\mathbf{R}}_2 = \begin{bmatrix} \widehat{\mathbf{H}}_{2,1}^T\mathbf{R}\widehat{\mathbf{H}}_{2,1} & \widehat{\mathbf{H}}_{2,1}^T\mathbf{R}\widehat{\mathbf{H}}_{2,2} & \cdots & \widehat{\mathbf{H}}_{2,1}^T\mathbf{R}\widehat{\mathbf{H}}_{2,P} \\ \widehat{\mathbf{H}}_{2,2}^T\mathbf{R}\widehat{\mathbf{H}}_{2,1} & \widehat{\mathbf{H}}_{2,2}^T\mathbf{R}\widehat{\mathbf{H}}_{2,2} & \cdots & \widehat{\mathbf{H}}_{2,2}^T\mathbf{R}\widehat{\mathbf{H}}_{2,P} \\ \vdots & \vdots & \ddots & \vdots \\ \widehat{\mathbf{H}}_{2,P}^T\mathbf{R}\widehat{\mathbf{H}}_{2,1} & \widehat{\mathbf{H}}_{2,P}^T\mathbf{R}\widehat{\mathbf{H}}_{2,2} & \cdots & \widehat{\mathbf{H}}_{2,P}^T\mathbf{R}\widehat{\mathbf{H}}_{2,P} \end{bmatrix},$$

$$\underline{\mathbf{p}}_1 = \begin{bmatrix} \mathbf{p}^T\widehat{\mathbf{H}}_{1,1} & \mathbf{p}^T\widehat{\mathbf{H}}_{1,2} & \cdots & \mathbf{p}^T\widehat{\mathbf{H}}_{1,P} \end{bmatrix}^T,$$

$$\underline{\mathbf{R}}_1 = \begin{bmatrix} \widehat{\mathbf{H}}_{1,1}^T\mathbf{R}\widehat{\mathbf{H}}_{1,1} & \widehat{\mathbf{H}}_{1,1}^T\mathbf{R}\widehat{\mathbf{H}}_{1,2} & \cdots & \widehat{\mathbf{H}}_{1,1}^T\mathbf{R}\widehat{\mathbf{H}}_{1,P} \\ \widehat{\mathbf{H}}_{1,2}^T\mathbf{R}\widehat{\mathbf{H}}_{1,1} & \widehat{\mathbf{H}}_{1,2}^T\mathbf{R}\widehat{\mathbf{H}}_{1,2} & \cdots & \widehat{\mathbf{H}}_{1,2}^T\mathbf{R}\widehat{\mathbf{H}}_{1,P} \\ \vdots & \vdots & \ddots & \vdots \\ \widehat{\mathbf{H}}_{1,P}^T\mathbf{R}\widehat{\mathbf{H}}_{1,1} & \widehat{\mathbf{H}}_{1,P}^T\mathbf{R}\widehat{\mathbf{H}}_{1,2} & \cdots & \widehat{\mathbf{H}}_{1,P}^T\mathbf{R}\widehat{\mathbf{H}}_{1,P} \end{bmatrix}.$$

We observe that the sizes of the matrices $\underline{\mathbf{R}}_1$ and $\underline{\mathbf{R}}_2$ are $PL_2 \times PL_2$ and $PML_1 \times PML_1$, respectively, which can be much smaller than the size of $\mathbf{R}$, which is $ML \times ML$. Also, we need at least $PML_1$ data samples to estimate the statistics in the MSE in (19) or (20), while at least $ML$ data samples are needed to estimate the statistics in the conventional MSE in (12). When $\widehat{\underline{\mathbf{h}}}_2$ is fixed, we write (19) as

$$\mathcal{J}_{\widehat{\underline{\mathbf{h}}}_2}\left(\widehat{\underline{\mathbf{h}}}_1\right) = \sigma_d^2 - 2\widehat{\underline{\mathbf{h}}}_1^T \underline{\mathbf{p}}_2 + \widehat{\underline{\mathbf{h}}}_1^T \underline{\mathbf{R}}_2 \widehat{\underline{\mathbf{h}}}_1 \tag{21}$$

and when $\widehat{\underline{\mathbf{h}}}_1$ is fixed, we express (20) as

$$\mathcal{J}_{\widehat{\underline{\mathbf{h}}}_1}\left(\widehat{\underline{\mathbf{h}}}_2\right) = \sigma_d^2 - 2\widehat{\underline{\mathbf{h}}}_2^T \underline{\mathbf{p}}_1 + \widehat{\underline{\mathbf{h}}}_2^T \underline{\mathbf{R}}_1 \widehat{\underline{\mathbf{h}}}_2. \tag{22}$$

To find the optimal filters, we can derive an iterative algorithm similar to those proposed in [2] and [3], but applicable for MISO system identification problems. At iteration 0, we set $\widehat{\mathbf{h}}_{2,p}^{(0)} = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}^T$, $p = 1, 2, \ldots, P$, from which we form $\widehat{\mathbf{H}}_{2,p}^{(0)} = \widehat{\mathbf{h}}_{2,p}^{(0)} \otimes \mathbf{I}_{ML_1}$ and

$$\underline{\mathbf{p}}_2^{(0)} = \begin{bmatrix} \mathbf{p}^T\widehat{\mathbf{H}}_{2,1}^{(0)} & \cdots & \mathbf{p}^T\widehat{\mathbf{H}}_{2,P}^{(0)} \end{bmatrix}^T,$$

$$\underline{\mathbf{R}}_2^{(0)} = \begin{bmatrix} \left(\widehat{\mathbf{H}}_{2,1}^{(0)}\right)^T \mathbf{R}\widehat{\mathbf{H}}_{2,1}^{(0)} & \cdots & \left(\widehat{\mathbf{H}}_{2,1}^{(0)}\right)^T \mathbf{R}\widehat{\mathbf{H}}_{2,P}^{(0)} \\ \left(\widehat{\mathbf{H}}_{2,2}^{(0)}\right)^T \mathbf{R}\widehat{\mathbf{H}}_{2,1}^{(0)} & \cdots & \left(\widehat{\mathbf{H}}_{2,2}^{(0)}\right)^T \mathbf{R}\widehat{\mathbf{H}}_{2,P}^{(0)} \\ \vdots & \ddots & \vdots \\ \left(\widehat{\mathbf{H}}_{2,P}^{(0)}\right)^T \mathbf{R}\widehat{\mathbf{H}}_{2,1}^{(0)} & \cdots & \left(\widehat{\mathbf{H}}_{2,P}^{(0)}\right)^T \mathbf{R}\widehat{\mathbf{H}}_{2,P}^{(0)} \end{bmatrix}.$$

Substituting these quantities into the MSE in (21), we obtain at iteration 1:

$$\mathcal{J}_{\widehat{\underline{\mathbf{h}}}_2}\left(\widehat{\underline{\mathbf{h}}}_1^{(1)}\right) = \sigma_d^2 - 2\left(\widehat{\underline{\mathbf{h}}}_1^{(1)}\right)^T \underline{\mathbf{p}}_2^{(0)} + \left(\widehat{\underline{\mathbf{h}}}_1^{(1)}\right)^T \underline{\mathbf{R}}_2^{(0)}\widehat{\underline{\mathbf{h}}}_1^{(1)},$$

whose minimization with respect to $\widehat{\underline{\mathbf{h}}}_1^{(1)}$ gives

$$\widehat{\underline{\mathbf{h}}}_1^{(1)} = \left(\underline{\mathbf{R}}_2^{(0)}\right)^{-1} \underline{\mathbf{p}}_2^{(0)}. \tag{23}$$

Then, from the above $\widehat{\underline{\mathbf{h}}}_1^{(1)}$, we form $\widehat{\mathbf{H}}_{1,p}^{(1)} = \mathbf{I}_{L_2} \otimes \widehat{\mathbf{h}}_{1,p}^{(1)}$ and

$$\underline{\mathbf{p}}_1^{(1)} = \begin{bmatrix} \mathbf{p}^T\widehat{\mathbf{H}}_{1,1}^{(1)} & \cdots & \mathbf{p}^T\widehat{\mathbf{H}}_{1,P}^{(1)} \end{bmatrix}^T,$$

$$\underline{\mathbf{R}}_1^{(1)} = \begin{bmatrix} \left(\widehat{\mathbf{H}}_{1,1}^{(1)}\right)^T \mathbf{R}\widehat{\mathbf{H}}_{1,1}^{(1)} & \cdots & \left(\widehat{\mathbf{H}}_{1,1}^{(1)}\right)^T \mathbf{R}\widehat{\mathbf{H}}_{1,P}^{(1)} \\ \left(\widehat{\mathbf{H}}_{1,2}^{(1)}\right)^T \mathbf{R}\widehat{\mathbf{H}}_{1,1}^{(1)} & \cdots & \left(\widehat{\mathbf{H}}_{1,2}^{(1)}\right)^T \mathbf{R}\widehat{\mathbf{H}}_{1,P}^{(1)} \\ \vdots & \ddots & \vdots \\ \left(\widehat{\mathbf{H}}_{1,P}^{(1)}\right)^T \mathbf{R}\widehat{\mathbf{H}}_{1,1}^{(1)} & \cdots & \left(\widehat{\mathbf{H}}_{1,P}^{(1)}\right)^T \mathbf{R}\widehat{\mathbf{H}}_{1,P}^{(1)} \end{bmatrix}.$$

As a result, the MSE in (22) is

$$\mathcal{J}_{\widehat{\underline{\mathbf{h}}}_1}\left(\widehat{\underline{\mathbf{h}}}_2^{(1)}\right) = \sigma_d^2 - 2\left(\widehat{\underline{\mathbf{h}}}_2^{(1)}\right)^T \underline{\mathbf{p}}_1^{(1)} + \left(\widehat{\underline{\mathbf{h}}}_2^{(1)}\right)^T \underline{\mathbf{R}}_1^{(1)}\widehat{\underline{\mathbf{h}}}_2^{(1)}.$$

The minimization of the previous expression with respect to $\widehat{\underline{\mathbf{h}}}_2^{(1)}$ gives

$$\widehat{\underline{\mathbf{h}}}_2^{(1)} = \left(\underline{\mathbf{R}}_1^{(1)}\right)^{-1} \underline{\mathbf{p}}_1^{(1)}. \tag{24}$$

Continuing to iterate up to iteration $n$, we easily get

$$\widehat{\underline{\mathbf{h}}}_1^{(n)} = \left(\underline{\mathbf{R}}_2^{(n-1)}\right)^{-1} \underline{\mathbf{p}}_2^{(n-1)}, \tag{25}$$

$$\widehat{\underline{\mathbf{h}}}_2^{(n)} = \left(\underline{\mathbf{R}}_1^{(n)}\right)^{-1} \underline{\mathbf{p}}_1^{(n)}, \tag{26}$$

where $\underline{\mathbf{R}}_2^{(n-1)}$, $\underline{\mathbf{p}}_2^{(n-1)}$, $\underline{\mathbf{R}}_1^{(n)}$, and $\underline{\mathbf{p}}_1^{(n)}$ are formed in a similar way to $\underline{\mathbf{R}}_2^{(0)}$, $\underline{\mathbf{p}}_2^{(0)}$, $\underline{\mathbf{R}}_1^{(1)}$, and $\underline{\mathbf{p}}_1^{(1)}$, respectively. Finally, we deduce that the Wiener filter at iteration $n$ is

$$\widehat{\underline{\mathbf{h}}}_{\mathrm{W}}^{(n)} = \sum_{p=1}^{P} \widehat{\mathbf{h}}_{2,p}^{(n)} \otimes \widehat{\mathbf{h}}_{1,p}^{(n)}. \tag{27}$$

In practice, to avoid any potential problems related to the matrix inversion operations, a small positive constant can be added to the diagonal elements of the matrices in (25)–(26).

This iterative algorithm resembles a block coordinate descent approach. The convergence of this type of methods was analyzed in [2] and [13]. Regarding the computational complexity, when $P \ll L_2$, the proposed solution "transforms" a system identification problem of size $ML_1L_2$ into two "smaller" problems of size $PML_1$ and $PL_2$, respectively.

## IV. Simulation Results

Simulations are performed in the framework of a MISO system identification problem, where the reference signal is obtained based on (1) [or (3)]. The additive noise $w(k)$ is white and Gaussian; the signal-to-noise ratio (SNR) is defined as SNR $= \sigma_y^2/\sigma_w^2$, where $\sigma_y^2$ and $\sigma_w^2$ are the variances of $y(k)$ and $w(k)$, respectively. In the experiments, we set SNR $= 30$ dB. The performance measure is the normalized misalignment (in dB). In case of the conventional Wiener filter, it is defined as $20\log_{10}\left(\left\|\underline{\mathbf{h}} - \widehat{\underline{\mathbf{h}}}_{\mathrm{W}}\right\|_2 / \|\underline{\mathbf{h}}\|_2\right)$, where $\widehat{\underline{\mathbf{h}}}_{\mathrm{W}}$ results from (13). For the iterative Wiener filter, this measure is evaluated as $20\log_{10}\left(\left\|\underline{\mathbf{h}} - \widehat{\underline{\mathbf{h}}}_{\mathrm{W}}^{(n)}\right\|_2 / \|\underline{\mathbf{h}}\|_2\right)$, where $\widehat{\underline{\mathbf{h}}}_{\mathrm{W}}^{(n)}$ results from (27).

Let us consider that $N$ data samples are available to estimate the covariance matrix $\mathbf{R}$ and the cross-correlation vector $\mathbf{p}$, which result in $\widehat{\mathbf{R}} = (1/N)\sum_{k=1}^{N} \underline{\mathbf{x}}(k)\underline{\mathbf{x}}^T(k)$ and $\widehat{\mathbf{p}} = (1/N)\sum_{k=1}^{N} \underline{\mathbf{x}}(k)d(k)$. These terms are a priori computed and then are used (instead of $\mathbf{R}$ and $\mathbf{p}$) within both the conventional and iterative Wiener filters. In other words, we consider that the statistics are available, which is in the spirit of the Wiener filter. However, in case of the conventional Wiener filter from (13), we need to solve a linear system with a matrix of size $ML \times ML$; consequently, more than $ML$ data samples are required to estimate the statistics $\widehat{\mathbf{R}}$ and $\widehat{\mathbf{p}}$, in order to obtain a reliable solution. On the other hand, the proposed iterative Wiener filter uses two "shorter" filters, of lengths $PML_1$ and $PL_2$ (usually, $P \ll L_2$).

In the first experiment, we consider $M = 4$ and use the first, the second, the fifth, and the sixth impulse responses from from G168 Recommendation [14]. These network echo paths are clusters of 64, 96, 96, and 120 coefficients, respectively, each one of them padded with zeros up to the length $L = 500$. The decomposition parameters are chosen as $L_1 = 25$ and $L_2 = 20$; in this case, $\mathrm{rank}(\mathbf{H}) = 5$. The input signals are independent AR(1) processes, where each one of them is generated by filtering a white Gaussian noise through a first-order system $1/\left(1 - 0.9z^{-1}\right)$. Based on the previous considerations, we evaluate two scenarios: (i) $N = 10000$, which should lead to an accurate solution of the conventional Wiener filter, and (ii) $N = 2000$, which basically represents the limit $ML$ (for $M = 4$ and $L = 500$). The results are provided in Figs. 1(a) and (b), respectively. As we can notice, the iterative Wiener filter using $P \ll L_2$ is able to outperform the conventional benchmark, especially when less data samples are available to estimate the statistics [2].

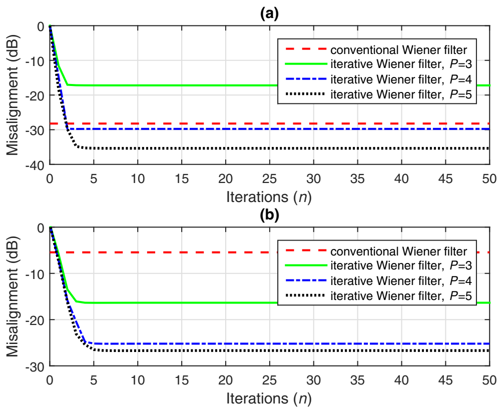


Fig. 1. Normalized misalignment of the conventional and iterative Wiener filters, using different numbers of data samples to estimate the statistics: (a) $N = 10000$ and (b) $N = 2000$. The impulse responses are network echo paths from G168 Recommendation [14], with $M = 4$ and $L = 500$. The input signals are independent AR(1) processes. The iterative Wiener filter uses $L_1 = 25$, $L_2 = 20$, and different values of $P$.

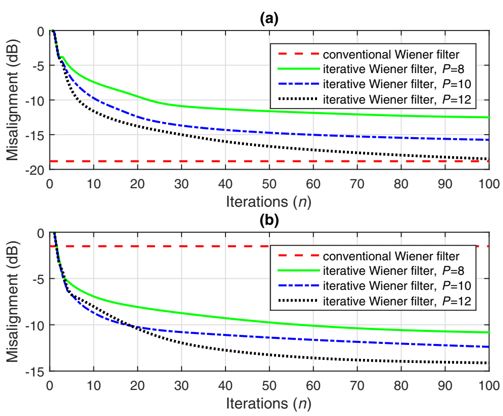


Fig. 2. Normalized misalignment of the conventional and iterative Wiener filters, using different numbers of data samples to estimate the statistics: (a) $N = 10000$ and (b) $N = 2000$. An SAEC scenario is considered ($M = 2$), where the length of each acoustic impulse response is $L = 1024$. The source signal is a white Gaussian noise and the input (or loudspeaker) signals are preprocessed with positive and negative half-wave rectifiers, using a distortion parameter equal to 0.5 [8]. The iterative Wiener filter uses $L_1 = L_2 = 32$ and different values of $P$.

In the second experiment, we consider a more challenging case when the input signals are correlated, which is similar to an SAEC scenario. In this case, the loudspeaker (input) signals are linearly related, which results in the so-called nonuniqueness problem [8]. This issue can be addressed by manipulating the signals transmitted to the receiving room, e.g., using a preprocessor on the loudspeaker signals to make them less coherent, but without affecting much the stereo perception and the signal quality. A simple but efficient nonlinear method uses positive and negative half-wave rectifiers on each channel respectively [8], controlling the amount of nonlinearity based on a "distortion" parameter. In our experiment, the source signal is a white Gaussian noise and the input (or loudspeaker) signals are preprocessed with positive and negative half-wave rectifiers, using a distortion parameter equal to $0.5$ [8]. The length of each acoustic echo path from the near-end room is $L = 1024$ and $M = 2$. In this case, the iterative Wiener filter uses $L_1 = L_2 = 32$ and different values of $P < L_2$. As expected, the matrix $\mathbf{H}$ is closer to full rank for the acoustic impulse responses [2]. However, even for a sufficiently large amount of data [e.g., $N = 10000$ in Fig. 2(a)], the iterative algorithm is able to match the performance of the conventional Wiener filter for a reasonably low value of $P$ (as compared to $L_2$). The gain becomes apparent when less data samples are available to estimate the statistics [e.g., $N = 2000$ in Fig. 2(b), which is slightly less than the limit $ML = 2048$]. In this case, the conventional Wiener filter cannot lead to a reliable result, while the iterative solution still provides a reasonable attenuation of the misalignment.

## V. Conclusions

In this paper, an iterative multichannel Wiener filter was developed in the context of MISO system identification. This algorithm is based on a Kronecker product decomposition of the impulse response, followed by low-rank approximations. Due to these features, it is suitable for the identification of low-rank systems, like the echo paths. Simulations performed in the context of multichannel echo cancellation (including SAEC) indicate that the proposed iterative algorithm is able to outperform the conventional Wiener filter, especially when a small amount of data is available to estimate the statistics.

## References

[1] S. Haykin, *Adaptive Filter Theory*. Fourth Edition, Upper Saddle River, NJ: Prentice-Hall, 2002.
[2] 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.
[3] 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.
[4] 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.
[5] M. Rupp and S. Schwarz, "Gradient-based approaches to learn tensor products," in *Proc. EUSIPCO*, 2015, pp. 2486–2490.
[6] 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.
[7] 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.
[8] J. Benesty, D. R. Morgan, and M. M. Sondhi, "A better understanding and an improved solution to the specific problems of stereophonic acoustic echo cancellation," *IEEE Trans. Speech, Audio Processing*, vol. 6, pp. 156–165, Mar. 1998.
[9] G. H. Golub and C. F. Van Loan, *Matrix Computations, 3rd Edition*. Baltimore, MD: The John Hopkins University Press, 1996.
[10] W. Gander, M. J. Gander, and F. Kwok, *Scientific Computing–An Introduction Using Maple and MATLAB*. Berlin: Springer-Verlag, 2014.
[11] C. F. Van Loan, "The ubiquitous Kronecker product," *J. Computational Applied Mathematics*, vol. 123, pp. 85–100, Nov. 2000.
[12] D. A. Harville, *Matrix Algebra From a Statistician's Perspective*. New York: Springer-Verlag, 1997.
[13] D. P. Bertsekas, *Nonlinear Programming*. Second Edition, Belmont, Massachusetts: Athena Scientific, 1999.
[14] *Digital Network Echo Cancellers*, ITU-T Recommendations G.168, 2002.