Performance Analysis of the Matrix Pair Method for Blind Channel Identification

Wanzhi Qiu, *Member, IEEE*, and Yingbo Hua, *Senior Member, IEEE*

***Abstract*—We study the estimation variance performance of the matrix pair (MP) method for estimating the impulse responses of multiple FIR channels driven by an unknown input sequence. A first-order perturbation analysis of the large-data-size performance of the MP method is presented and an explicit expression for the estimation variance is derived. Both the theoretical and simulation results are used to investigate the statistical performance of the MP method and a number of new insights are revealed.**

***Index Terms*—Asymptotic analysis, blind identification, matrix pair method.**

I. Introduction

BLIND channel identification is useful in communications as it does not require a training sequence to equalize a channel and hence it could save the channel bandwidth. Most blind channel identification schemes begin by sampling the channel output at the baud rate to produce a stationary channel output sequence for processing. Consequently, higher order statistics (HOS) is required either explicitly or implicitly to identify a possibly nonminimum phase channel. Due to the large number of data samples and large amount of computation required to estimate HOS, their applications may be limited in fast changing environments, such as in mobile communications, where the channel has to be estimated within a short period of time.

The work by Tong–Xu–Kailath [1] appears to be a major breakthrough in the attempt to achieve fast blind channel identification. It is demonstrated in [1] that when sampled at a rate higher than the baud (symbol) rate, the output of a data communication channel can be described as that of a multichannel system which is driven by a sequence of unknown input symbols. This multichannel model allows the second-order statistics (SOS) to be sufficient to uniquely (up to a constant) estimate the system impulse response without knowing its input under a mild condition [1]. This result is believed to have inspired all the subsequent development in identifying a channel without using higher order statistics (see [2]–[12], for example).

Manuscript received April 20, 1995; revised October 27, 1996. This work was supported by the Australian Research Council and the Australian Cooperative Research Centre for Sensor Signal and Information Processing. The material in this paper was presented in part at the Asilomar Conference on Signals, Systems, and Computers, Monterey, CA, Nov. 1994.

W. Qiu was with the Department of ˜Electrical and Electronic Engineering, University of Melbourne, Parkville, Victoria 3052, Australia. He is now with Motorola Australian Research Centre, Botany 2019, Sydney, Australia.

Y. Hua is with the Department of Electrical and Electronic Engineering, University of Melbourne, Parkville, Victoria 3052, Australia.

Publisher Item Identifier S 0018-9448(97)03778-4.

The method developed by Tong–Xu–Kailath [1] exploits a pair ("matrix pair") of covariance matrices of the channel outputs and hence will be referred to as the matrix pair (MP) method. The MP method is the first dedicated algorithm for solving the blind channel identification problem where only SOS is used. It first demonstrated achieving satisfactory blind channel identification by using only a few hundred data samples. However, the estimation variance performance of this method has not been analyzed in depth before. In this paper, we provide an asymptotic (large-data-size) analysis of the MP method. In particular, we derive the asymptotic estimation variance of the MP method based on the first-order perturbation theory. We then investigate, using both theoretical and simulation results, the dependence of its estimation variance performance on the signal-to-noise ratio (SNR), the data size, the channel condition, the number of channels, and processing window length. The rest of this paper is organized as follows. Section II summarizes the MP method for easy reference. Section III derives the estimation variance of the MP method. Section IV presents numerical examples which verify the theoretical results and show some new insights into the MP method. Section V gives the conclusions.

II. The MP Method

As shown in [1], when a higher sampling rate (compared to the baud rate) is used at the output, a data communication system can be described as the following multichannel FIR system:

$$y_m(k) = \sum_{l=0}^{L} h_m(l)s(k-l) + w_m(k), \qquad k = 0, \cdots, N-1, \quad m = 1, \cdots, M \tag{1}$$

where $y_m(k)$ is the output of the mth channel, $h_m(l)$ the impulse response of the mth channel, $s(k)$ the common input to the M channels, L the maximum order of these M channels, N the data size, and $w_m(k)$ the (zero-mean) white noise.

The outputs of system (1) can be expressed in the following vector form by choosing a processing window of W baud intervals as in [1]:

$$\boldsymbol{y}(n) = \boldsymbol{H}\boldsymbol{s}(n) + \boldsymbol{w}(n) \quad MW \times 1 \quad n = 0, \cdots, N-W \tag{2}$$

where

$$\begin{aligned} \boldsymbol{y}(n) &= \left[\boldsymbol{y}_1^T(n),\ \boldsymbol{y}_2^T(n), \cdots, \boldsymbol{y}_W^T(n)\right]^T \\ \boldsymbol{y}_k(n) &= [y_1(n+k-1), y_2(n+k-1), \cdots, \\ &\qquad y_M(n+k-1)]^T, \quad k = 1, \cdots, W \\ \boldsymbol{s}(n) &= [s(n-L), s(n-L+1), \cdots, s(n+W-1)]^T \\ &\qquad (W+L) \times 1 \end{aligned} \tag{3}$$

with ^{T} denoting the transpose, and $\boldsymbol{H}$ is a generalized Sylvester matrix defined by

$$\boldsymbol{H} = \begin{pmatrix} \boldsymbol{h}_L & \boldsymbol{h}_{L-1} & \cdots & \boldsymbol{h}_0 & 0 & \cdots & 0 \\ 0 & \boldsymbol{h}_L & \boldsymbol{h}_{L-1} & \cdots & \boldsymbol{h}_0 & \cdots & 0 \\ \vdots & & \ddots & \ddots & & \ddots & \vdots \\ 0 & \cdots & 0 & \boldsymbol{h}_L & \boldsymbol{h}_{L-1} & \cdots & \boldsymbol{h}_0 \end{pmatrix} \quad MW \times (W+L) \tag{4}$$

with

$$\boldsymbol{h}_k = [h_1(k), h_2(k), \cdots, h_M(k)]^T, \quad k = 0, \cdots, L.$$

The noise vector $\boldsymbol{w}(n)$ is defined in the same way as $\boldsymbol{y}(n)$.

The channel impulse response vector is defined by

$$\boldsymbol{h} = \left[\boldsymbol{h}_0^T, \boldsymbol{h}_1^T, \cdots, \boldsymbol{h}_L^T\right]^T \tag{5}$$

and the objective of blind channel identification is to obtain $\boldsymbol{h}$ only from the received data $\{\boldsymbol{y}(n)\}$. For unique identification (up to a constant scalar), it is required by the MP method that the following conditions be satisfied [1], [6], [7]:

A1 Polynomials $H_m(z)$ do not share any common root, where

$$H_m(z) = \sum_{l=0}^{L} h_m(l) z^{-l}, \qquad m = 1, \cdots, M.$$

Note that this immediately implies $M \geq 2$.

A2 The matrix $\boldsymbol{H}$ has more rows than columns, i.e., $MW > (W+L) \triangleq d$.

A3 The noise samples are (zero-mean) white with variance σ^2.

A4 The input symbol $\{s(k)\}$ is a (zero-mean) white sequence.

It is assumed in [1] that $\{s(k)\}$ has unit variance (without loss of generality); this leads to

$$E\{\boldsymbol{s}(n)\boldsymbol{s}^H(n+k)\} = \boldsymbol{J}_k \quad (d \times d) \tag{6}$$

where $E\{\ \}$ denotes the expectation, ^{H} the complex conjugate transpose, and

$$(\boldsymbol{J}_k)_{uv} = \begin{cases} 1, & \text{if } u = v + k \text{ and } |k| \leq d-1 \\ 0, & \text{otherwise.} \end{cases}$$

Similarly

$$E\{\boldsymbol{w}(n)\boldsymbol{w}^H(n+k)\} = \sigma^2 \boldsymbol{J}'_{kM} \quad (MW \times MW) \tag{7}$$

where $\boldsymbol{J}'_k$ is defined in the same way as $\boldsymbol{J}_k$ except for the dimension.

With the above notations, one can write

$$\boldsymbol{R}(0) \triangleq E\{\boldsymbol{y}(n)\boldsymbol{y}^H(n)\} = \boldsymbol{H}\boldsymbol{H}^H + \sigma^2 \boldsymbol{I}_{MW} \tag{8}$$

$$\boldsymbol{R}(1) \triangleq E\{\boldsymbol{y}(n)\boldsymbol{y}^H(n+1)\} = \boldsymbol{H}\boldsymbol{J}_1\boldsymbol{H}^H + \sigma^2 \boldsymbol{J}'_M \tag{9}$$

where $\boldsymbol{I}_{MW}$ denotes the $MW \times MW$ identity matrix. It is the covariance property shown in (8) and (9) that Tong–Xu–Kailath exploited to estimate $\boldsymbol{H}$ (up to an unknown constant scalar $e^{j\phi}$) from $\boldsymbol{R}(0)$ and $\boldsymbol{R}(1)$. The impulse response vector $\boldsymbol{h}$ can be easily extracted once $\boldsymbol{H}$ is available. For easy reference, the MP method is summarized as follows.

Step 1. Compute the eigendecomposition of $\boldsymbol{R}(0)$, i.e.,

$$\boldsymbol{R}(0) = \boldsymbol{U}_s \boldsymbol{\Sigma}_s \boldsymbol{U}_s^H + \boldsymbol{U}_w \boldsymbol{\Sigma}_w \boldsymbol{U}_w^H$$

where

$$\boldsymbol{\Sigma}_s = \operatorname{diag}[\lambda_1, \cdots, \lambda_d]$$
$$\boldsymbol{\Sigma}_w = \operatorname{diag}[\lambda_{d+1}, \cdots, \lambda_{MW}]$$
$$\boldsymbol{U}_s = [\boldsymbol{u}_1, \boldsymbol{u}_2, \cdots, \boldsymbol{u}_d]$$
$$\boldsymbol{U}_w = [\boldsymbol{u}_{d+1}, \boldsymbol{u}_{d+2}, \cdots, \boldsymbol{u}_{MW}].$$

$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d > \lambda_{d+1} = \lambda_{d+2} = \cdots = \lambda_{MW} = \sigma^2$ are the eigenvalues of $\boldsymbol{R}(0)$, $\boldsymbol{U}_s$ and $\boldsymbol{U}_w$ contain orthonormal eigenvectors corresponding to $\boldsymbol{\Sigma}_s$ and $\boldsymbol{\Sigma}_w$, respectively. Range($\boldsymbol{U}_s$) and range($\boldsymbol{U}_w$) are referred to as the signal and noise subspaces, respectively.

Step 2. Form

$$\boldsymbol{Q} = \boldsymbol{\Lambda}^{-\frac{1}{2}} \boldsymbol{U}_s^H \tilde{\boldsymbol{R}}(1) \boldsymbol{U}_s \boldsymbol{\Lambda}^{-\frac{1}{2}} \tag{10}$$

where

$$\boldsymbol{\Lambda} = \boldsymbol{\Sigma}_s - \sigma^2 \boldsymbol{I}_d \tag{11}$$

$$\tilde{\boldsymbol{R}}(1) = \boldsymbol{R}(1) - \sigma^2 \boldsymbol{J}'_M. \tag{12}$$

Then compute $\boldsymbol{y}_d$, the left singular vector corresponding to the smallest singular value of $\boldsymbol{Q}$.

Step 3. Form an estimate of $\boldsymbol{H}$ (up to a constant scalar $e^{j\phi}$) by

$$\boldsymbol{H} = \boldsymbol{U}_s \boldsymbol{\Lambda}^{\frac{1}{2}} \boldsymbol{V}$$

where

$$\boldsymbol{V} = \left[\boldsymbol{y}_d, \boldsymbol{Q}\boldsymbol{y}_d, \cdots, \boldsymbol{Q}^{(d-1)}\boldsymbol{y}_d\right]$$

and extract the estimate of $\boldsymbol{h}_{L-n+1}$ by (as suggested in [1])

$$\boldsymbol{h}_{L-n+1} = \boldsymbol{W}\boldsymbol{\xi}_n \quad (n = 1, 2, \cdots, L+1) \tag{13}$$

where $\boldsymbol{\xi}_n$ denotes the nth column of $\boldsymbol{H}$, and $\boldsymbol{W}$ an $M \times MW$ matrix defined as $\boldsymbol{W} = [\boldsymbol{I}_M; \boldsymbol{0}]$ (see (4)).

III. Analysis of the MP Method

In practice, only finite data samples are available $(N < \infty)$, and the covariance matrices $\boldsymbol{R}(0)$ and $\boldsymbol{R}(1)$ must be estimated

$$\hat{\boldsymbol{R}}(0) = \frac{1}{T} \sum_{n=0}^{T-1} \boldsymbol{y}(n)\boldsymbol{y}^H(n) \tag{14}$$

$$\hat{\boldsymbol{R}}(1) = \frac{1}{T} \sum_{n=0}^{T-1} \boldsymbol{y}(n)\boldsymbol{y}^H(n+1) \tag{15}$$

where $T = N - W$, and $\hat{}$ denotes the estimate.

The finite data effect and the additive noise will result in an error in the estimate of $\boldsymbol{h}$. In this section, we derive the asymptotic estimation variance of the MP method using the first-order approximations. Our derivation development is orientated by the fact that the estimate of $\boldsymbol{h}$ is extracted from the estimate of the matrix $\boldsymbol{H}$ which is formed from certain estimated eigenvalues and eigenvectors. Accordingly, we first show some standard perturbation results on the eigenvalues and eigenvectors, which lay a foundation for our derivation of the estimation variance. We then derive the perturbation expressions for the columns of $\boldsymbol{H}$. Finally, we formulate the perturbation on $\boldsymbol{h}$ and present the estimation variance. In the following, we will denote a (first-order) perturbation by preceding the corresponding quantity by Δ and use the symbol $\approx$ to denote an equality in which the higher order terms are neglected. We assume that the noise samples $\{w_m(k)\}$ are complex-Gaussian [15].

Assuming that the dominant eigenvalues of $\boldsymbol{R}(0)$ (i.e., $\lambda_i : i = 1, \cdots, d$) are distinct, from [14, pp. 293–295] we have

$$\Delta\lambda_i \triangleq \hat{\boldsymbol{\lambda}}_i - \boldsymbol{\lambda}_i \approx \boldsymbol{u}_i^H \Delta\boldsymbol{R}(0)\boldsymbol{u}_i, \qquad i = 1, \cdots, d \tag{16}$$

and

$$\begin{aligned}\Delta\boldsymbol{u}_i &\triangleq \hat{\boldsymbol{u}}_i - \boldsymbol{u}_i \\ &\approx (\lambda_i \boldsymbol{I}_{MW} - \boldsymbol{R}(0))^{\dagger}\Delta\boldsymbol{R}(0)\boldsymbol{u}_i, \qquad i = 1, \cdots, d\end{aligned} \tag{17}$$

where † denotes the pseudoinverse.

We now derive the perturbation on the nth column of $\boldsymbol{H}$

$$\boldsymbol{\xi}_n \triangleq \boldsymbol{U}_s \boldsymbol{\Lambda}^{\frac{1}{2}} \boldsymbol{v}_n, \quad n = 1, \cdots, d \tag{18}$$

where

$$\boldsymbol{v}_n = \boldsymbol{Q}^{n-1}\boldsymbol{y}_d. \tag{19}$$

We obtain from (18) that

$$\Delta\boldsymbol{\xi}_n \approx \Delta\boldsymbol{U}_s\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{v}_n + \boldsymbol{U}_s\Delta\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{v}_n + \boldsymbol{U}_s\boldsymbol{\Lambda}^{\frac{1}{2}}\Delta\boldsymbol{v}_n \tag{20}$$

where, from (19), we have

$$\Delta\boldsymbol{v}_n \begin{cases} = \Delta\boldsymbol{y}_d, & n = 1 \\ \approx \displaystyle\sum_{i=1}^{n-1} \boldsymbol{Q}^{i-1}\Delta\boldsymbol{Q}\boldsymbol{Q}^{n-i-1}\boldsymbol{y}_d + \boldsymbol{Q}^{n-1}\Delta\boldsymbol{y}_d, & n > 1 \end{cases} \tag{21}$$

and we can show in Appendix I that

$$\Delta\boldsymbol{y}_d \approx -\boldsymbol{Q}\Delta\boldsymbol{Q}^H\boldsymbol{y}_d. \tag{22}$$

Since, due to (10)

$$\begin{aligned}\Delta\boldsymbol{Q} \approx\ & \Delta\boldsymbol{\Lambda}^{-\frac{1}{2}}\boldsymbol{U}_s^H\tilde{\boldsymbol{R}}(1)\boldsymbol{U}_s\boldsymbol{\Lambda}^{-\frac{1}{2}} + \boldsymbol{\Lambda}^{-\frac{1}{2}}\Delta\boldsymbol{U}_s^H\tilde{\boldsymbol{R}}(1)\boldsymbol{U}_s\boldsymbol{\Lambda}^{-\frac{1}{2}} \\ &+ \boldsymbol{\Lambda}^{-\frac{1}{2}}\boldsymbol{U}_s^H\Delta\tilde{\boldsymbol{R}}(1)\boldsymbol{U}_s\boldsymbol{\Lambda}^{-\frac{1}{2}} + \boldsymbol{\Lambda}^{-\frac{1}{2}}\boldsymbol{U}_s^H\tilde{\boldsymbol{R}}(1)\Delta\boldsymbol{U}_s\boldsymbol{\Lambda}^{-\frac{1}{2}} \\ &+ \boldsymbol{\Lambda}^{-\frac{1}{2}}\boldsymbol{U}_s^H\tilde{\boldsymbol{R}}(1)\boldsymbol{U}_s\Delta\boldsymbol{\Lambda}^{-\frac{1}{2}}\end{aligned} \tag{23}$$

and, due to the fact that $\boldsymbol{\Lambda}$ is diagonal

$$\Delta\boldsymbol{\Lambda}^{-\frac{1}{2}} \approx -\frac{1}{2}\boldsymbol{\Lambda}^{-\frac{3}{2}}\Delta\boldsymbol{\Lambda} \quad \text{and} \quad \Delta\boldsymbol{\Lambda}^{\frac{1}{2}} \approx \frac{1}{2}\boldsymbol{\Lambda}^{-\frac{1}{2}}\Delta\boldsymbol{\Lambda} \tag{24}$$

by substituting (21)–(24) into (20), one can show that for $n = 1$

$$\Delta\boldsymbol{\xi}_n \approx \sum_{i=1}^{2n}\boldsymbol{A}_i\Delta\boldsymbol{\Lambda}\boldsymbol{a}_i^{(1)} + \sum_{i=1}^{n+1}\boldsymbol{C}_i\Delta\boldsymbol{U}_s\boldsymbol{a}_i^{(3)} + \boldsymbol{E}_1\Delta\tilde{\boldsymbol{R}}(-1)\boldsymbol{a}_1^{(5)} \tag{25}$$

and for $n > 1$

$$\begin{aligned}\Delta\boldsymbol{\xi}_n \approx\ & \sum_{i=1}^{2n}\boldsymbol{A}_i\Delta\boldsymbol{\Lambda}\boldsymbol{a}_i^{(1)} + \sum_{i=1}^{n-1}\boldsymbol{B}_i\Delta\boldsymbol{U}_s^H\boldsymbol{a}_i^{(2)} \\ &+ \sum_{i=1}^{n+1}\boldsymbol{C}_i\Delta\boldsymbol{U}_s\boldsymbol{a}_i^{(3)} + \sum_{i=1}^{n-1}\boldsymbol{D}_i\Delta\tilde{\boldsymbol{R}}(1)\boldsymbol{a}_i^{(4)} \\ &+ \boldsymbol{E}_1\Delta\tilde{\boldsymbol{R}}(-1)\boldsymbol{a}_1^{(5)}\end{aligned} \tag{26}$$

where $\tilde{\boldsymbol{R}}(-1) = \tilde{\boldsymbol{R}}^H(1)$, and the matrices $\boldsymbol{A}_i, \boldsymbol{B}_i, \boldsymbol{C}_i, \boldsymbol{D}_i$, and $\boldsymbol{E}_1$ as well as the vectors $\boldsymbol{a}_i^{(k)} (k = 1, \cdots, 5)$ are given in Appendix II.

Next, using (16) and (17), we can further relate $\Delta\boldsymbol{\xi}_n$ to the perturbations of the covariance matrices by

$$\Delta\boldsymbol{\xi}_n \approx \sum_{j=1}^{I(n)}\boldsymbol{F}'_j\Delta\boldsymbol{R}(k_j)\boldsymbol{g}_j \tag{27}$$

where

$$I(n) = MW + nd + n, \qquad n = 1, 2, \cdots, d.$$

For $n = 1$

$$k_j = \begin{cases} 0, & j = 1, \cdots, MW + d \\ -1, & j = I(1). \end{cases} \tag{28}$$

For $n > 1$

$$k_j = \begin{cases} 0, & j = 1, \cdots, MW + nd \\ 1, & j = MW + nd + 1, \cdots, I(n) - 1 \\ -1, & j = I(n) \end{cases} \tag{29}$$

where the matrices $\boldsymbol{F}'_j$ and the vectors $\boldsymbol{g}_j$ are given in Appendix III, and their dependence on n is not shown explicitly for notational simplicity. See Appendix III for the derivation of (27).

Finally, considering (13), we obtain the perturbation expression for $\boldsymbol{h}_{L-n+1}$ as

$$\Delta\boldsymbol{h}_{L-n+1} \approx \sum_{j=1}^{I(n)}\boldsymbol{F}_j\Delta\boldsymbol{R}(k_j)\boldsymbol{g}_j, \qquad n = 1, 2, \cdots, L+1 \tag{30}$$

where

$$\boldsymbol{F}_j = \boldsymbol{W}\boldsymbol{F}'_j.$$

Before proceeding to Theorem 1, we introduce a lemma which is useful in the derivation of the estimation variance.

Lemma 1: Suppose that the input sequence $\{s(k)\}$ is independent and identically distributed (i.i.d.) with zero-mean, unit-variance and finite fourth-order moments and the noise $\{w_m(k)\}$ is circular-Gaussian, the random variable $\{\sqrt{T}\Delta R(k)_{ab}; -d < k < d,\ 1 \le a, b \le MW\}$, are jointly asymptotically normal with zero mean and covariance

$$\begin{aligned}&E\{T\Delta R(k)_{ab}\Delta R(l)_{cd}\} \\ &\quad = \sum_{\tau\in\mathcal{Z}} R(k+\tau)_{ad}R(l-\tau)_{cb} \\ &\qquad + \sum_{\tau\in\mathcal{Z}} \bar{R}(k+\tau)_{ac}\bar{R}^*(l+\tau)_{bd} + \kappa\tilde{R}(k)_{ab}\tilde{R}(l)_{cd}\end{aligned} \tag{31}$$

where

$$\boldsymbol{R}(k) \triangleq E\{\boldsymbol{y}(n)\boldsymbol{y}^H(n+k)\} = \boldsymbol{H}\boldsymbol{J}_k\boldsymbol{H}^H + \sigma^2 \boldsymbol{J}'_{kM} \quad (32)$$

$$\tilde{\boldsymbol{R}}(k) \triangleq \boldsymbol{R}(k) - \sigma^2 \boldsymbol{J}'_{kM} = \boldsymbol{H}\boldsymbol{J}_k\boldsymbol{H}^H \quad (33)$$

$$\bar{\boldsymbol{R}}(k) \triangleq E\{\boldsymbol{y}(n)\boldsymbol{y}^T(n+k)\} = E\{(s(k))^2\}\boldsymbol{H}\boldsymbol{J}_k\boldsymbol{H}^T \quad (34)$$

$$\kappa = E\{|s(k)|^4\} - |E\{(s(k))^2\}|^2 - 2$$

$\mathcal{Z}$ denotes the set of integers, * the circular conjugate, and A_{mn} the mnth element of matrix $\boldsymbol{A}$.

The proof of Lemma 1 is by direct application of [16, Theorem 14, p 228] (also see [8]). Note that the right-hand side of (31) involves only a finite number of terms due to the fact that $\boldsymbol{R}(k) = \bar{\boldsymbol{R}}(k) = \boldsymbol{0}$ when $|k| \geq d$.

Theorem 1: The variance of the large-data-size estimate of $\boldsymbol{h}$ is given by

$$E\left\{\Delta\boldsymbol{h}^H\Delta\boldsymbol{h}\right\} = \frac{a}{T}\sigma^4 + \frac{b}{T}\sigma^2 + \frac{c}{T} \quad (35)$$

where the scalars a, b, and c are independent of T and σ^2 and are given by

$$a = \sum_{n=1}^{L+1}\sum_{i,j=1}^{I(n)}\sum_{\tau\in\mathcal{Z}} \boldsymbol{g}_i^H \boldsymbol{J}'_{(\tau-k_i)M}\boldsymbol{g}_j T_r(\boldsymbol{G}\boldsymbol{J}'_{(k_j-\tau)M})$$

$$b = \sum_{n=1}^{L+1}\sum_{i,j=1}^{I(n)}\sum_{\tau\in\mathcal{Z}} \boldsymbol{g}_i^H \tilde{\boldsymbol{R}}(\tau-k_i)\boldsymbol{g}_j T_r\left(\boldsymbol{G}\boldsymbol{J}'_{(k_j-\tau)M}\right) + \sum_{n=1}^{L+1}\sum_{i,j=1}^{I(n)}\sum_{\tau\in\mathcal{Z}} \boldsymbol{g}_i^H \boldsymbol{J}'_{(\tau-k_i)M}\boldsymbol{g}_j T_r\left(\boldsymbol{G}\tilde{\boldsymbol{R}}(k_j-\tau)\right)$$

$$\begin{aligned} c = &\sum_{n=1}^{L+1}\sum_{i,j=1}^{I(n)}\sum_{\tau\in\mathcal{Z}} \boldsymbol{g}_i^H \tilde{\boldsymbol{R}}(\tau-k_i)\boldsymbol{g}_j T_r\left(\boldsymbol{G}\tilde{\boldsymbol{R}}(k_j-\tau)\right) \\ &+ |E\{(s(k))^2\}|^2 \sum_{n=1}^{L+1}\sum_{i,j=1}^{I(n)}\sum_{\tau\in\mathcal{Z}} \boldsymbol{g}_i^H \boldsymbol{H}\boldsymbol{J}_{\tau-k_i} \\ &\times \boldsymbol{H}^T\boldsymbol{G}^T\boldsymbol{H}^*\boldsymbol{J}_{\tau+k_j}\boldsymbol{H}^H\boldsymbol{g}_j \\ &+ \kappa \sum_{n=1}^{L+1}\sum_{i,j=1}^{I(n)} \boldsymbol{g}_i^H\tilde{\boldsymbol{R}}(-k_i)\boldsymbol{G}\tilde{\boldsymbol{R}}(k_j)\boldsymbol{g}_j \end{aligned}$$

with T_r denoting "the trace of" and

$$\boldsymbol{G} = \boldsymbol{F}_i^H\boldsymbol{F}_j.$$

Proof: Due to (5) and (30)

$$\begin{aligned} E\{\Delta\boldsymbol{h}^H\Delta\boldsymbol{h}\} &= \sum_{n=1}^{L+1} E\left\{\Delta\boldsymbol{h}_{L-n+1}^H\Delta\boldsymbol{h}_{L-n+1}\right\} \\ &= \sum_{n=1}^{L+1}\sum_{i,j=1}^{I(n)} E\left\{\boldsymbol{g}_i^H\Delta\boldsymbol{R}(-k_i)\boldsymbol{G}\Delta\boldsymbol{R}(k_j)\boldsymbol{g}_j\right\}. \end{aligned}$$

Denoting by $f(a)$ the ath element of vector $\boldsymbol{f}$ and using Lemma 1, we have

$$\begin{aligned} &E\{\Delta\boldsymbol{h}^H\Delta\boldsymbol{h}\} \\ &= \sum_{n=1}^{L+1}\sum_{i,j=1}^{I(n)}\sum_{a,b,c,d=1}^{MW} E\left\{g_i^*(a)\Delta R(-k_i)_{ab}G_{bc}\Delta R(k_j)_{cd}g_j(d)\right\} \\ &= \frac{1}{T}\sum_{n=1}^{L+1}\sum_{i,j=1}^{I(n)}\sum_{a,b,c,d=1}^{MW} g_i^*(a)G_{bc}g_j(d) \\ &\quad \times \sum_{\tau\in\mathcal{Z}} R(-k_i+\tau)_{ad}R(k_j-\tau)_{cb} \\ &\quad + \frac{1}{T}\sum_{n=1}^{L+1}\sum_{i,j=1}^{I(n)}\sum_{a,b,c,d=1}^{MW} g_i^*(a)G_{bc}g_j(d) \\ &\quad \times \sum_{\tau\in\mathcal{Z}} \bar{R}(-k_i+\tau)_{ac}\bar{R}^*(k_j+\tau)_{bd} \\ &\quad + \frac{\kappa}{T}\sum_{n=1}^{L+1}\sum_{i,j=1}^{I(n)}\sum_{a,b,c,d=1}^{MW} g_i^*(a)G_{bc}g_j(d)\tilde{R}(-k_i)_{ab}\tilde{R}(k_j)_{cd} \\ &= \frac{1}{T}\sum_{n=1}^{L+1}\sum_{i,j=1}^{I(n)}\sum_{\tau\in\mathcal{Z}} \boldsymbol{g}_i^H\boldsymbol{R}(-k_i+\tau)\boldsymbol{g}_jT_r(\boldsymbol{G}\boldsymbol{R}(k_j-\tau)) \\ &\quad + \frac{1}{T}\sum_{n=1}^{L+1}\sum_{i,j=1}^{I(n)}\sum_{\tau\in\mathcal{Z}} \boldsymbol{g}_i^H\bar{\boldsymbol{R}}(-k_i+\tau)\boldsymbol{G}^T\bar{\boldsymbol{R}}^*(k_j+\tau)\boldsymbol{g}_j \\ &\quad + \frac{\kappa}{T}\sum_{n=1}^{L+1}\sum_{i,j=1}^{I(n)} \boldsymbol{g}_i^H\tilde{\boldsymbol{R}}(-k_i)\boldsymbol{G}\tilde{\boldsymbol{R}}(k_j)\boldsymbol{g}_j. \end{aligned} \quad (36)$$

The conclusion follows by plugging (32)–(34) into (36).

Note that for large data size, it is much more economical to evaluate this theoretical variance than to run a Monte Carlo simulation. Hence, this expression could also be valuable in practical design.

IV. Performance Evaluation

This section investigates the statistical performance of the MP method using both theoretical and simulation results.

The asymptotic estimation variance of the MP method is shown as in (35) to be a quadratic function in terms of the noise variance σ^2. It is not surprising to note that there is a "constant" term in (35), which implies that, when N is finite, we cannot obtain an exact estimate even when the noise is absent ($\sigma^2 = 0$). This is due to the fact that, even when the noise is absent, an infinite number of data samples are still required to obtain the exact covariance matrices $\boldsymbol{R}(0)$ and $\boldsymbol{R}(1)$ and, consequently, the eigenvalues and eigenvectors to form the channel estimate. This "constant" term, i.e., c/T, tells us what estimation variance can be achieved at high SNR.

We now present a numerical Monte Carlo study which verifies our theoretical work and shows more insights into the performance of the MP method. As mentioned in Section II, the estimate of $\boldsymbol{h}$ is unique up to a constant scalar $e^{j\phi}$. In the simulation, we remove this angular ambiguity by replacing $\hat{\boldsymbol{x}}$, the estimate of the eigenvector $\boldsymbol{x}$ ($\boldsymbol{x} = \boldsymbol{u}_1, \cdots, \boldsymbol{u}_d$ successively in Step 1), and $\boldsymbol{x} = \boldsymbol{y}_d$ in Step 2) by $\hat{\boldsymbol{x}}e^{j\phi_o}$ where ϕ_o is given

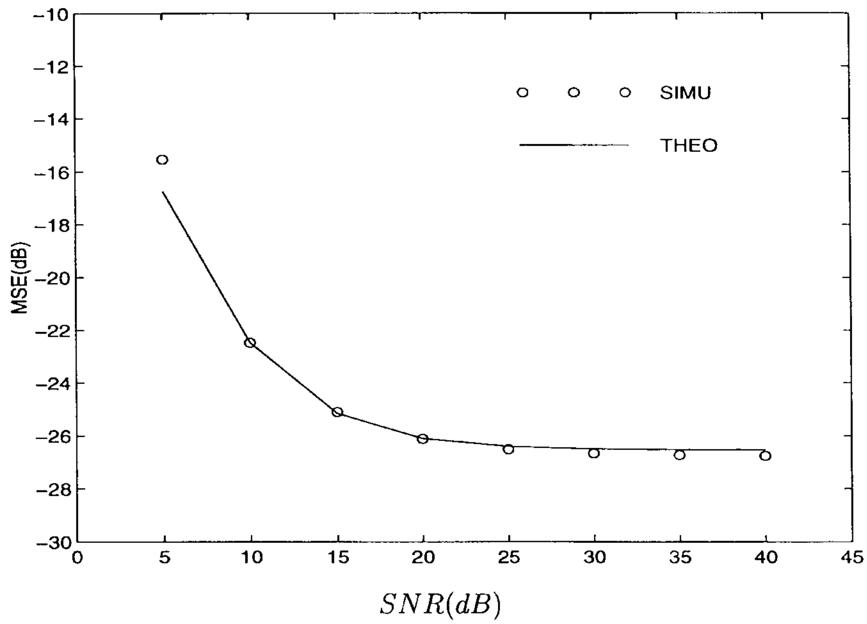

Fig. 1. Performance of the MP method versus SNR where $N = 600$, $L = 1$, $W = 5$, and $\lambda = 0.2$.

by

$$\phi_o = \arg\min_{\phi} ||\hat{\boldsymbol{x}} e^{j\phi} - \boldsymbol{x}||^2.$$

This minimization yields

$$e^{j\phi_o} = \frac{\hat{\boldsymbol{x}}^H \boldsymbol{x}}{||\hat{\boldsymbol{x}}^H \boldsymbol{x}||.}$$

The performance of the MP method in simulation is measured by the mean-square error in decibels

$$\text{MSE (dB)} \triangleq 20 \log_{10} \left(\frac{1}{||\boldsymbol{h}||} \sqrt{\frac{1}{N_r} \sum_{i=1}^{N_r} ||\hat{\boldsymbol{h}}_i - \boldsymbol{h}||^2} \right)$$

where $\hat{\boldsymbol{h}}_i$ denotes the estimate of $\boldsymbol{h}$ in the ith run. $N_r = 100$ independent runs are conducted for each simulation scenario, and both the input sequence and the noise sequence are independently chosen at each run. Accordingly, the theoretical performance is calculated by using Theorem 1 and

$$\text{MSE (dB)} \triangleq 20 \log_{10} \left(\frac{1}{||\boldsymbol{h}||} \sqrt{E\{\Delta \boldsymbol{h}^H \Delta \boldsymbol{h}\}} \right).$$

We also need to define the SNR as follows (see [11]):

$$\text{SNR (dB)} = 10 \log_{10} \frac{||\boldsymbol{h}||^2}{M\sigma^2}.$$

The following two-channel ($M = 2$) system used in [8] is considered for Figs. 1–3

$$H_1(z) = 1.000 + 0.6476 z^{-1} \quad (37)$$

$$H_2(z) = 1.000 + 0.6476 \lambda z^{-1} \quad (38)$$

where λ is used to control the distance between the roots of $H_1(z)$ and $H_2(z)$. For small λ, the system becomes ill-conditioned and a relatively large estimation error is expected (see condition **A1** in Section II). In all cases to be shown, the channel coefficients will be scaled such that $||\boldsymbol{h}|| = 1$ and the input sequence is 4-QAM scaled by $E\{s(k)s^*(k)\} = 1$. Unless otherwise stated, $\lambda = 0.2$, SNR $= 20$ dB, $N = 600$, and $W = 5$.

Fig. 1 shows the performance of the MP method against the SNR. It is seen in this case that, after 20 dB, the curve becomes nearly flat. This is due to the finite data effect discussed above which cannot be cured by increasing the SNR.

Fig. 2 shows the performance of the MP method against the data size N. It is shown that, as expected, a better estimate can be achieved by using more data samples. It is also shown that the (asymptotic) theoretical expression of the MSE is valid even for a data size as small as 100.

Fig. 3 shows the performance of the MP method against the channel condition (λ). Consistent with our previous discussion, the MSE increases as the zeros of the two channels become closer. This implies that the MP method fails for very poor channel conditions.

Fig. 4 shows the performance of the MP method against the number of channels (M). For $M = 2$, we consider the two-channel system (37) and (38). For $M = 3$, we consider the three-channel system (37), (38), and

$$H_3(z) = 1.000 + 0.6476 \delta z^{-1}$$

where δ is set to be $0.0, 0.2, 0.9$, and 1.0 in four cases, respectively. It is shown that adding different channels has different effects on the estimation error. This is due to the fact that, in the multichannel model, a larger M means a higher sampling rate, implying that more information is utilized from the channel output to carry out the identification and a better estimate is expected. On the other hand, a larger M results in a larger number of unknowns and more closely distributed zeros of the channels, which tends to increase the estimation error.

Fig. 5 shows the performance of the MP method against the processing window length W. Two two-channel systems

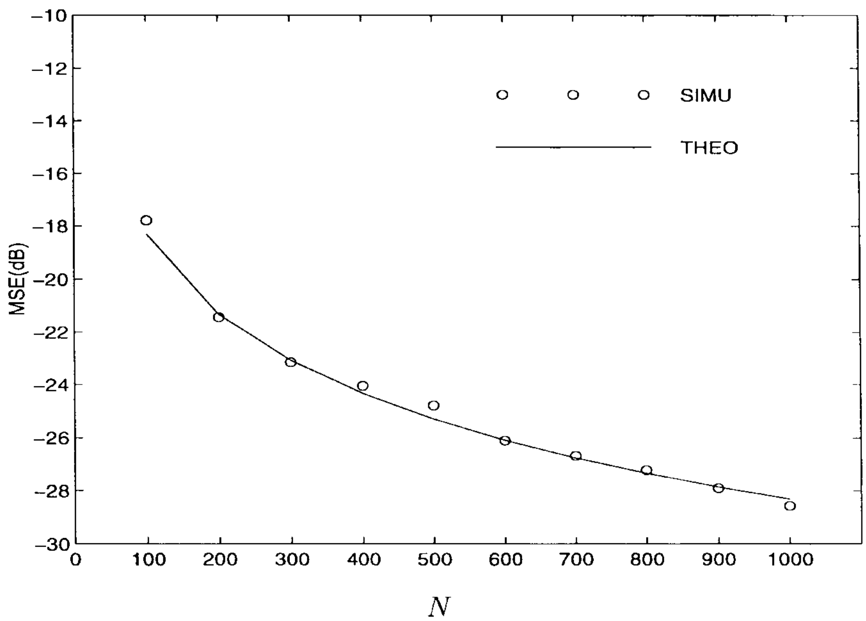

Fig. 2. Performance of the MP method versus N where SNR $= 20$ dB, $L = 1$, $W = 5$, and $\lambda = 0.2$.

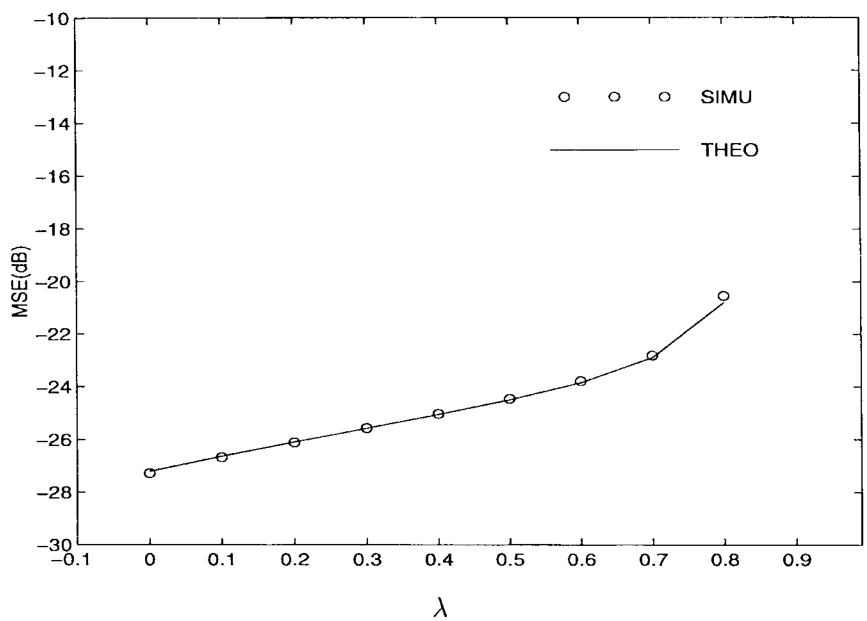

Fig. 3. Performance of the MP method versus λ where SNR $= 20$ dB, $N = 600$, $L = 1$, and $W = 5$.

are considered where system 1 consists of (37) and (38), and system 2 consists of the following two channels:

$$H_1(z) = 1 - 2\cos(0.1\pi)z^{-1} + z^{-2}$$
$$H_2(z) = 1 - 2\cos(0.4\pi)z^{-1} + z^{-2}.$$

It is observed that the effect of changing W depends on the particular system to be identified.

V. Concluding Remarks

We have analyzed the statistical performance of the MP method for blind identification of multiple FIR channels. In particular, we have obtained a closed-form expression for the large-data-size estimation variance, and further investigated how the SNR, the data size, the channel condition, the number of channels, and the processing window length affect the estimation accuracy. The explicit formula of the estimation variance and the insights revealed can be helpful for the identification system designers.

Appendix I
Proof of (22)

As shown in [1], V is unitary and

$$\boldsymbol{Q} = \boldsymbol{V}\boldsymbol{J}_1\boldsymbol{V}^H$$
$$\boldsymbol{Q}\boldsymbol{Q}^H = \boldsymbol{V}\mathrm{diag}(1, \cdots, 1, 0)\boldsymbol{V}^H.$$

Since $\boldsymbol{y}_d$ is the eigenvector corresponding to the single eigen-

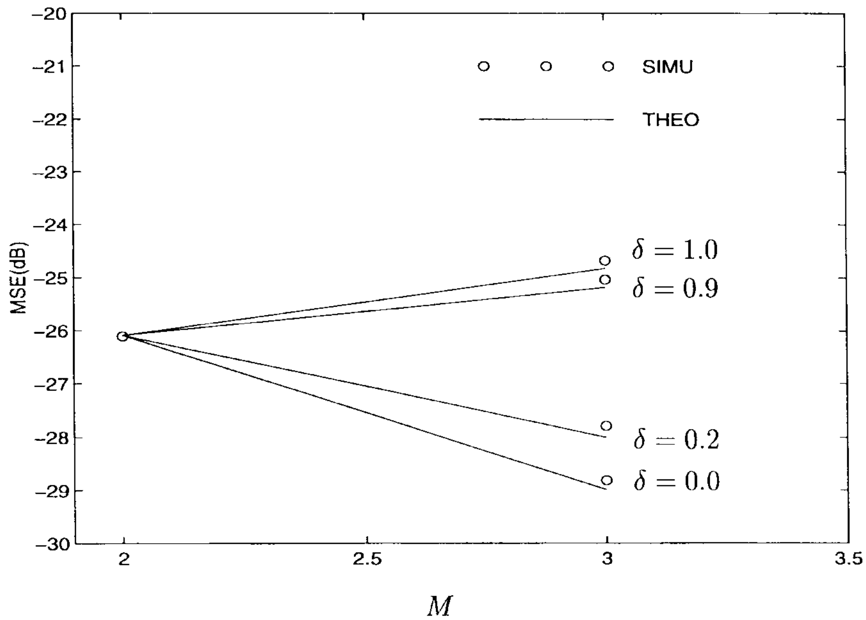

Fig. 4. Performance of the MP method versus M where SNR $= 20$ dB, $N = 600$, $L = 1$, and $W = 5$.

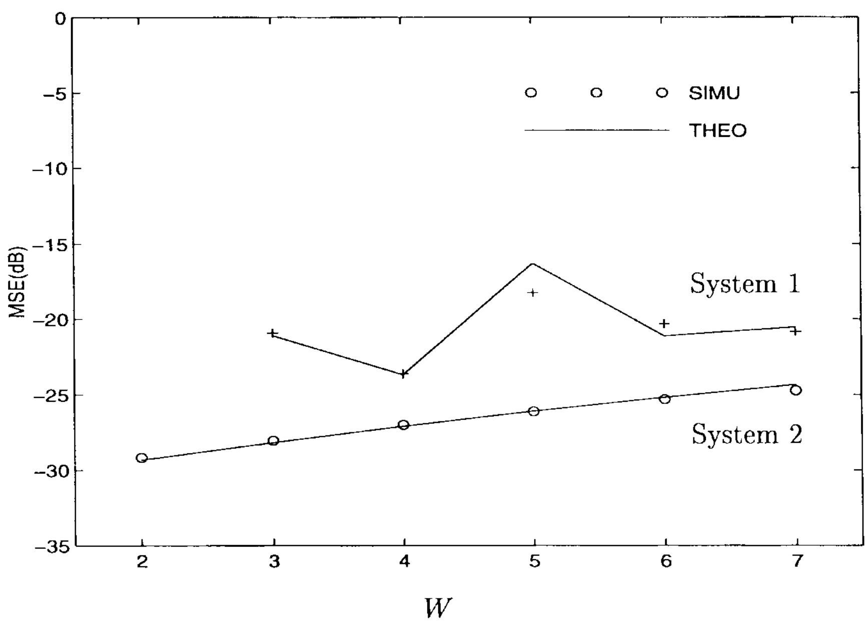

Fig. 5. Performance of the MP method versus W where SNR $= 20$ dB and $N = 600$.

value (i.e., 0) of $\boldsymbol{QQ}^H$, from [14, pp. 293–295], we have

$$\begin{aligned}\Delta \boldsymbol{y}_d &\approx \boldsymbol{V}\text{diag}(-1,\cdots,-1,0)\boldsymbol{V}^H\Delta(\boldsymbol{QQ}^H)\boldsymbol{y}_d\\ &\approx -\boldsymbol{QQ}^H(\Delta\boldsymbol{QQ}^H+\boldsymbol{Q}\Delta\boldsymbol{Q}^H)\boldsymbol{y}_d\\ &= -\boldsymbol{QQ}^H\boldsymbol{Q}\Delta\boldsymbol{Q}^H\boldsymbol{y}_d.\end{aligned}$$

In the last equation, we used the fact that $\boldsymbol{Q}^H\boldsymbol{y}_d = 0$. Noting that $\boldsymbol{QQ}^H\boldsymbol{Q} = \boldsymbol{Q}$ completes the proof.

Appendix II
Definitions of Deterministic Matrices and Vectors in (25) and (26)

Because $\boldsymbol{H} = \boldsymbol{U}_s\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{V}$ and $\boldsymbol{V}$ is unitary [1], it is very easy to verify that

$$\boldsymbol{H}^\dagger = \boldsymbol{V}^H\boldsymbol{\Lambda}^{-\frac{1}{2}}\boldsymbol{U}_s^H.$$

Using $\tilde{\boldsymbol{R}}(1) = \boldsymbol{HJ}_1\boldsymbol{H}^H$, and denoting by $\boldsymbol{\eta}_i$ the ith column of $(\boldsymbol{H}^\dagger)^H$, after some straightforward but tedious rearrangement, one can show that

For $n > 1$:

$$\boldsymbol{A}_i = \begin{cases} -\frac{1}{2}\boldsymbol{HJ}_{i-1}\boldsymbol{H}^\dagger\boldsymbol{U}_s, & i = 1,\cdots,n-1\\ \frac{1}{2}\boldsymbol{HJ}_n\boldsymbol{J}_{-1}\boldsymbol{H}^\dagger\boldsymbol{U}_s, & i = n\\ -\frac{1}{2}\boldsymbol{HJ}_{i-n}\boldsymbol{H}^\dagger\boldsymbol{U}_s, & i = n+1,\cdots,2n-1\\ \frac{1}{2}\boldsymbol{U}_s, & i = 2n\end{cases}$$

$$\boldsymbol{a}_i^{(1)} = \begin{cases}\boldsymbol{U}_s^H\boldsymbol{\eta}_{n-i+1}, & i = 1,\cdots,n\\ \boldsymbol{U}_s^H\boldsymbol{\eta}_{2n-i}, & i = n+1,\cdots,2n-1\\ \boldsymbol{U}_s^H\boldsymbol{\eta}_n, & i = 2n\end{cases}$$

$$\boldsymbol{B}_i = \boldsymbol{HJ}_{i-1}\boldsymbol{H}^\dagger\boldsymbol{U}_s, \quad i = 1,\cdots,n-1$$

$$\boldsymbol{a}_i^{(2)} = \boldsymbol{\xi}_{n-i+1}, \quad i = 1,\cdots,n-1$$

$$\boldsymbol{C}_i = \begin{cases} \boldsymbol{H}\boldsymbol{J}_i\boldsymbol{H}^H, & i=1,\cdots,n-1 \\ -\boldsymbol{H}\boldsymbol{J}_n\boldsymbol{J}_{-1}\boldsymbol{H}^H, & i=n \\ \boldsymbol{I}_{MW} & i=n+1, \end{cases}$$

$$\boldsymbol{a}_i^{(3)} = \begin{cases} \boldsymbol{U}_s^H\boldsymbol{\eta}_{n-i}, & i=1,\cdots,n-1 \\ \boldsymbol{U}_s^H\boldsymbol{\eta}_1, & i=n \\ \boldsymbol{U}_s^H\boldsymbol{\xi}_n, & i=n+1 \end{cases}$$

$$\boldsymbol{D}_i = \boldsymbol{H}\boldsymbol{J}_{i-1}\boldsymbol{H}^\dagger, \quad i=1,\cdots,n-1$$
$$\boldsymbol{a}_i^{(4)} = \boldsymbol{\eta}_{n-i}, \quad i=1,\cdots,n-1$$
$$\boldsymbol{E}_1 = -\boldsymbol{H}\boldsymbol{J}_n\boldsymbol{H}^\dagger$$
$$\boldsymbol{a}_1^{(5)} = \boldsymbol{\eta}_1.$$

For $n=1$:

$$\boldsymbol{A}_i = \begin{cases} \frac{1}{2}\boldsymbol{H}\boldsymbol{J}_1\boldsymbol{J}_{-1}\boldsymbol{H}^\dagger\boldsymbol{U}_s, & i=1 \\ \frac{1}{2}\boldsymbol{U}_s, & i=2 \end{cases}$$
$$\boldsymbol{a}_i^{(1)} = \boldsymbol{U}_s^H\boldsymbol{\eta}_i, \quad i=1,2$$
$$\boldsymbol{C}_i = \begin{cases} -\boldsymbol{H}\boldsymbol{J}_1\boldsymbol{J}_{-1}\boldsymbol{H}^H, & i=1 \\ \boldsymbol{I}_{MW}, & i=2 \end{cases}$$
$$\boldsymbol{a}_i^{(3)} = \begin{cases} \boldsymbol{U}_s^H\boldsymbol{\eta}_1, & i=1 \\ \boldsymbol{U}_s^H\boldsymbol{\xi}_1, & i=2 \end{cases}$$
$$\boldsymbol{E}_1 = -\boldsymbol{H}\boldsymbol{J}_1\boldsymbol{H}^\dagger$$
$$\boldsymbol{a}_1^{(5)} = \boldsymbol{\eta}_1.$$

APPENDIX III
PROOF OF (27)

Let us evaluate the terms in (25) and (26) one by one. Using (16), we can show that

$$\begin{aligned}\sum_{i=1}^{2n}\boldsymbol{A}_i\Delta\boldsymbol{\Lambda}\boldsymbol{a}_i^{(1)} &= \sum_{i=1}^{2n}\boldsymbol{A}_i\big(\Delta\boldsymbol{\Sigma}_s - \Delta\sigma^2\boldsymbol{I}_d\big)\boldsymbol{a}_i^{(1)} \\ &= \sum_{i=1}^{2n}\sum_{j=1}^{d}\boldsymbol{a}_j(i)\Delta\lambda_j\boldsymbol{a}_{ij}^{(1)} - \Delta\sigma^2\sum_{i=1}^{2n}\boldsymbol{A}_i\boldsymbol{a}_i^{(1)} \\ &\approx \sum_{j=1}^{d}\boldsymbol{A}'_j\Delta\boldsymbol{R}(0)\boldsymbol{u}_j - \Delta\sigma^2\boldsymbol{r}_1 \end{aligned} \quad (39)$$

where

$$\boldsymbol{A}'_j = \sum_{i=1}^{2n}\boldsymbol{a}_{ij}^{(1)}\boldsymbol{a}_j(i)\boldsymbol{u}_j^H$$
$$\boldsymbol{r}_1 = \sum_{i=1}^{2n}\boldsymbol{A}_i\boldsymbol{a}_i^{(1)}$$

$\boldsymbol{a}_j(i)$ is the jth column of $\boldsymbol{A}_i$ and $\boldsymbol{a}_{ij}^{(1)}$ the jth entry of $\boldsymbol{a}_i^{(1)}$. By assuming that the noise variance is estimated by the average of the estimated less-dominant eigenvalues of $\boldsymbol{R}(0)$ (i.e., $\hat{\sigma}^2 = \frac{\hat{\lambda}_{d+1}+\cdots+\hat{\lambda}_{MW}}{MW-d}$), we have

$$\Delta\sigma^2 \approx \frac{1}{MW-d}\sum_{i=1}^{MW}\boldsymbol{e}_i^H\Delta\boldsymbol{R}(0)\boldsymbol{e}_i - \frac{1}{MW-d}\sum_{i=1}^{d}\boldsymbol{u}_i^H\Delta\boldsymbol{R}(0)\boldsymbol{u}_i$$

where $\boldsymbol{e}_i$ is the $MW\times 1$ vector with 1 as the ith entry and 0's elsewhere.

Using (17) one can rewrite the second term in (26) as

$$\begin{aligned}\sum_{i=1}^{n-1}\boldsymbol{B}_i\Delta\boldsymbol{U}_s^H\boldsymbol{a}_i^{(2)} &= \sum_{i=1}^{n-1}\sum_{j=1}^{d}\boldsymbol{b}_j(i)\Delta\boldsymbol{u}_j^H\boldsymbol{a}_i^{(2)} \\ &\approx \sum_{i=1}^{n-1}\sum_{j=1}^{d}\boldsymbol{b}_j(i)\boldsymbol{u}_j^H\Delta\boldsymbol{R}(0)\big(\lambda_j\boldsymbol{I}_{MW} - \boldsymbol{R}(0)\big)^\dagger\boldsymbol{a}_i^{(2)} \\ &= \sum_{j=1}^{(n-1)d}\boldsymbol{B}'_j\Delta\boldsymbol{R}(0)\check{\boldsymbol{a}}_j^{(2)} \end{aligned} \quad (40)$$

where

$$\boldsymbol{B}'_j = \boldsymbol{b}_{(j-1)_d+1}\left(\frac{j-1-(j-1)_d}{d}+1\right)\boldsymbol{u}_{(j-1)_d+1}^H$$
$$\check{\boldsymbol{a}}_j^{(2)} = \big(\lambda_{(j-1)_d+1}\boldsymbol{I}_{MW} - \boldsymbol{R}(0)\big)^\dagger\boldsymbol{a}^{(2)}_{\frac{j-1-(j-1)_d}{d}+1}$$

$\boldsymbol{b}_j(i)$ is the jth column of $\boldsymbol{B}_i$, and $(j-1)_d \triangleq \bmod\,[j-1,d]$.

Using (17) one can show

$$\begin{aligned}\sum_{i=1}^{n+1}\boldsymbol{C}_i\Delta\boldsymbol{U}_s\boldsymbol{a}_i^{(3)} &= \sum_{i=1}^{n+1}\sum_{j=1}^{d}\boldsymbol{C}_i\Delta\boldsymbol{u}_j\boldsymbol{a}_{ij}^{(3)} \\ &\approx \sum_{i=1}^{n}\sum_{j=1}^{d}\boldsymbol{a}_{ij}^{(3)}\boldsymbol{C}_i(\lambda_j\boldsymbol{I}_{MW} - \boldsymbol{R}(0))^\dagger\Delta\boldsymbol{R}(0)\boldsymbol{u}_j \\ &= \sum_{j=1}^{d}\boldsymbol{C}'_j\Delta\boldsymbol{R}(0)\boldsymbol{u}_j \end{aligned} \quad (41)$$

where

$$\boldsymbol{C}'_j = \sum_{i=1}^{n}\boldsymbol{a}_{ij}^{(3)}\boldsymbol{C}_i(\lambda_j\boldsymbol{I}_{MW} - \boldsymbol{R}(0))^\dagger$$

$\boldsymbol{a}_{ij}^{(3)}$ is the jth entry of vector $\boldsymbol{a}_i^{(3)}$.

Using (12), we rewrite the fourth term in (26) by

$$\begin{aligned}\sum_{i=1}^{n-1}\boldsymbol{D}_i\Delta\tilde{\boldsymbol{R}}(1)\boldsymbol{a}_i^{(4)} &= \sum_{i=1}^{n-1}\boldsymbol{D}_i\big(\Delta\boldsymbol{R}(1) - \Delta\sigma^2\boldsymbol{J}'_M\big)\boldsymbol{a}_i^{(4)} \\ &= \sum_{i=1}^{n-1}\boldsymbol{D}_i\Delta\boldsymbol{R}(1)\boldsymbol{a}_i^{(4)} - \Delta\sigma^2\boldsymbol{r}_2 \end{aligned} \quad (42)$$

where

$$\boldsymbol{r}_2 = \sum_{i=1}^{n-1}\boldsymbol{D}_i\boldsymbol{J}'_M\boldsymbol{a}_i^{(4)}.$$

Similarly, we show

$$\begin{aligned}\boldsymbol{E}_1\Delta\tilde{\boldsymbol{R}}(-1)\boldsymbol{a}_1^{(5)} &= \boldsymbol{E}_1\big(\Delta\boldsymbol{R}(-1) - \Delta\sigma^2\boldsymbol{J}'_{-M}\big)\boldsymbol{a}_1^{(5)} \\ &= \boldsymbol{E}_1\Delta\boldsymbol{R}(-1)\boldsymbol{a}_1^{(5)} - \Delta\sigma^2\boldsymbol{r}_3 \end{aligned} \quad (43)$$

where

$$\boldsymbol{r}_3 = \boldsymbol{E}_1\boldsymbol{J}'_{-M}\boldsymbol{a}_1^{(5)}.$$

For $n > 1$: By summing up (39), (42), and (43), we get the following summation:

$$\sum_{j=1}^{d} \boldsymbol{A}'_j \Delta \boldsymbol{R}(0)\boldsymbol{u}_j + \sum_{i=1}^{n-1} \boldsymbol{D}_i \Delta \boldsymbol{R}(1)\boldsymbol{a}_i^{(4)} + \boldsymbol{E}_1 \Delta \boldsymbol{R}(-1)\boldsymbol{a}_1^{(5)} - \Delta\sigma^2(\boldsymbol{r}_1 + \boldsymbol{r}_2 + \boldsymbol{r}_3)$$
$$= \sum_{j=1}^{d}\left(\boldsymbol{A}'_j + \frac{\boldsymbol{r}_1+\boldsymbol{r}_2+\boldsymbol{r}_3}{MW-d}\boldsymbol{u}_j^H\right)\Delta\boldsymbol{R}(0)\boldsymbol{u}_j - \sum_{j=1}^{MW}\frac{\boldsymbol{r}_1+\boldsymbol{r}_2+\boldsymbol{r}_3}{MW-d}\boldsymbol{e}_j^H\Delta\boldsymbol{R}(0)\boldsymbol{e}_j + \sum_{j=1}^{n-1} D_j\Delta\boldsymbol{R}(1)\boldsymbol{a}_j^{(4)} + \boldsymbol{E}_1\Delta\boldsymbol{R}(-1)\boldsymbol{a}_1^{(5)}. \quad (44)$$

Furthermore, the summation of (40), (41), and (44) gives

$$\Delta\boldsymbol{\xi}_n = \sum_{j=1}^{I(n)} \boldsymbol{F}'_j \Delta\boldsymbol{R}(k_j)\boldsymbol{g}_j$$

where

$$I(n) = MW + nd + n$$

$$\boldsymbol{F}'_j = \begin{cases} \boldsymbol{A}'_j + \frac{\boldsymbol{r}_1+\boldsymbol{r}_2+\boldsymbol{r}_3}{MW-d}\boldsymbol{u}_j^H + \boldsymbol{C}'_j, & j = 1,\cdots,d \\ -\frac{\boldsymbol{r}_1+\boldsymbol{r}_2+\boldsymbol{r}_3}{MW-d}\boldsymbol{e}_{j-d}^H, & j = d+1,\cdots, d+MW \\ \boldsymbol{B}'_{j-(d+MW)}, & j = d+MW+1,\cdots, MW+nd \\ D_{j-(MW+nd)}, & j = MW+nd+1,\cdots, I(n)-1 \\ \boldsymbol{E}_1, & j = I(n) \end{cases}$$

$$\boldsymbol{g}_j = \begin{cases} \boldsymbol{u}_j, & j = 1,\cdots,d \\ \boldsymbol{e}_{j-d}, & j = d+1,\cdots,d+MW \\ \check{\boldsymbol{a}}_{j-(d+MW)}^{(2)}, & j = d+MW+1,\cdots,MW+nd \\ \boldsymbol{a}_{j-(MW+nd)}^{(4)}, & j = MW+nd+1,\cdots,I(n)-1 \\ \boldsymbol{a}_1^{(5)}, & j = I(n) \end{cases}$$

$$\Delta\boldsymbol{R}(k_j) = \begin{cases} \Delta\boldsymbol{R}(0), & j = 1,\cdots,MW+nd \\ \Delta\boldsymbol{R}(1) & j = MW+nd+1,\cdots,I(n)-1 \\ \Delta\boldsymbol{R}(-1), & j = I(n). \end{cases} \quad (45)$$

Note that (45) immediately leads to (29).

For $n = 1$: The summation of (39), (41), and (43) gives

$$\Delta\boldsymbol{\xi}_n = \sum_{j=1}^{I(1)} \boldsymbol{F}'_j \Delta\boldsymbol{R}(k_j)\boldsymbol{g}_j$$

where

$$I(1) = MW + d + 1$$

$$\boldsymbol{F}'_j = \begin{cases} \boldsymbol{A}'_j + \frac{\boldsymbol{r}_1+\boldsymbol{r}_3}{MW-d}\boldsymbol{u}_j^H + \boldsymbol{C}'_j, & j = 1,\cdots,d \\ -\frac{\boldsymbol{r}_1+\boldsymbol{r}_3}{MW-d}\boldsymbol{e}_{j-d}^H, & j = d+1,\cdots,I(1)-1 \\ \boldsymbol{E}_1, & j = I(1) \end{cases}$$

$$\boldsymbol{g}_j = \begin{cases} \boldsymbol{u}_j, & j = 1,\cdots,d \\ \boldsymbol{e}_{j-d}, & j = d+1,\cdots,I(1)-1 \\ \boldsymbol{a}_1^{(5)}, & j = I(1) \end{cases}$$

$$\Delta\boldsymbol{R}(k_j) = \begin{cases} \Delta\boldsymbol{R}(0), & j = 1,\cdots,I(1)-1 \\ \Delta\boldsymbol{R}(-1), & j = I(1). \end{cases} \quad (46)$$

Noting that (28) is the immediate results of (46) completes the proof.

ACKNOWLEDGMENT

The authors wish to thank the reviewers for their stimulating and helpful comments.

REFERENCES

[1] L. Tong, G. Xu, and T. Kailath, "Blind identification and equalization based on second-order statistics: A time domain approach," *IEEE Trans. Inform. Theory*, vol. 40, pp. 340–349, Mar. 1994. A short version in *Proc. 25th Asilomar Conf. on Signals, Systems, and Computers*, Nov. 1991.

[2] Z. Ding, "Blind identification and equalization using spectral measures: Part I," in *Cyclostationarity in Communications and Signal Processing*, W. A. Gardner, Ed. New York: IEEE Press, 1993.

[3] G. B. Giannakis, "Linear cyclic correlation approach for blind identification of FIR channels," in *Proc. 28th Asilomar Conf. on Signals, Systems, and Computers* (Pacific Grove, CA, Nov. 1994).

[4] M. Shao and C. L. Nikias, "An ML/MMSE estimation approach to blind equalization," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing* (Adelaide, Australia, Apr. 1994), vol. 4, pp. 569–572.

[5] D. T. M. Slock, "Blind fractionally-space equalization, perfect-reconstruction filter bank and multichannel linear prediction," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing* (Adelaide, Australia, Apr. 1994), vol. 4, pp. 585–588.

[6] L. Tong, G. Xu, B. Hassibi, and T. Kailath, "Blind identification and equalization based on second-order statistics: A frequency domain approach," *IEEE Trans. Inform. Theory*, vol. 41, pp. 329–333, Jan. 1995.

[7] G. Xu, H. Liu, L. Tong, and T. Kailath, "A least-squares approach to blind channel identification," *IEEE Trans. Signal Processing*, vol. 43, pp. 2982–2993, Dec. 1995.

[8] K. Abed-Meraim, P. Loubaton, and E. Moulines, "A subspace algorithm for certain blind identification problems," *IEEE Trans. on Information Theory*, vol. 43, pp. 499–511, Mar. 1997.

[9] H. Liu and G. Xu, "A deterministic approach to blind symbol estimation," *IEEE Signal Processing Letters*, vol. 1, no. 12, Dec. 1994.

[10] E. Moulines, P. Duhamel, J.-F. Cardoso, and S. Mayrargue, "Subspace methods for the blind identification of multichannel FIR filters," *IEEE Trans. Signal Processing*, vol. 43, pp. 516–525, Feb. 1995.

[11] Y. Hua, "Fast maximum likelihood for blind identification of multiple FIR channels," *IEEE Trans. Signal Processing*, vol. 44, pp. 661–672, Mar. 1996.

[12] Y. Hua, H. Yang, and W. Qiu, "Source correlation compensation for blind channel identification based on second-order statistics," *IEEE Signal Processing Lett.*, vol. 1, pp. 119–120, Aug. 1994.

[13] P. Stoica and T. Soderstrom, "Statistical analysis of MUSIC and subspace rotation estimates of sinusoidal frequencies," *IEEE Trans. Signal Processing*, vol. 39, pp. 1836–1846, Aug. 1991.

[14] G. W. Stewart, *Introduction to Matrix Computations*. New York: Academic, 1973.

[15] S. U. Pillai and B. H. Kwon, "Performance analysis of MUSIC-type high resolution estimators for direction finding in correlated coherent scenes," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 1176–1189, Aug. 1989.

[16] E. Hannan, *Multiple Time Series*. New York: Wiley, 1970.