Blind Channel Equalization With Colored Sources Based on Second-Order Statistics: A Linear Prediction Approach

Roberto López-Valcarce, *Member, IEEE,* and Soura Dasgupta, *Fellow, IEEE*

***Abstract*—We consider the blind equalization and estimation of single-user, multichannel models from the second-order statistics of the channel output when the channel input statistics are colored but known. By exploiting certain results from linear prediction theory, we generalize the algorithm of Tong *et al.*, which was derived under the assumption of a white transmitted sequence. In particular, we show that all one needs to estimate the channel to within an unitary scaling constant, and thus to find its equalizers, is a) that a standard channel matrix have full column rank, and b) a vector of the input signal and its delays have positive definite lag zero autocorrelation. An algorithm is provided to determine the equalizer under these conditions. We argue that because this algorithm makes explicit use of the input statistics, the equalizers thus obtained should outperform those obtained by other methods that neither require, nor exploit, the knowledge of the input statistics. Simulation results are provided to verify this fact.**

***Index Terms*—Blind channel estimation, blind equalization, colored sources, linear prediction theory, second-order statistics.**

I. Introduction

In MANY digital communication systems, intersymbol interference (ISI), which is a result of the dispersive characteristics of the channel, becomes the main limitation to performance. Traditionally, these systems have relied on known training sequences that are used to estimate and equalize the channel. To conserve bandwidth lost through the transmission of training signals, an alternative to training is the use of blind estimation/equalization techniques. Such techniques have received considerable attention in the literature and are the subject of this paper.

Early blind techniques exploited higher (than second) order statistics of the channel output to estimate the channel and compute the equalizer [5], [6]. More recently, there has been considerable interest in methods based on second-order statistics (SOS) after the seminal work in [16] showed that finite impulse response (FIR) single-input multiple-output (SIMO) channels can be perfectly equalized by means of a bank of FIR equalizers, which can be computed from the channel output SOS. Following [16], many SOS-based blind methods have been proposed; see, e.g., [4], [13], [14], and [19]. Other types of methods have also been developed, such as deterministic [18] and maximum likelihood schemes [8]; see also the review [15].

We consider the blind equalization/estimation of SIMO models from the SOS of the channel output. Of specific interest is the case where the channel input statistics *are colored but known*. This is not to be confused with semi-blind channel estimation, which assumes additional knowledge of the symbol sequence: specifically, that part of the data vector is known [3]. By contrast, in our framework, the symbol sequence is completely unknown on a sample-by-sample basis, and only its second-order statistical information is available to the receiver. Colored sources may arise, for example, as a result of channel encoding [12], and the knowledge of the encoding scheme alone will provide the required source statistics to the receiver.

The precise channel model to be considered is the FIR SIMO model

$$\overline{\mathbf{x}}_n = \sum_{i=0}^{l} \mathbf{h}_i a_{n-i} + \overline{\mathbf{w}}_n \tag{1}$$

where $\{a_n\}$ is the zero mean, wide sense stationary sequence of transmitted symbols, $\{\overline{\mathbf{x}}_n\}$ is the $p\times 1$ vector of channel outputs, $\{\overline{\mathbf{w}}_n\}$ is a $p \times 1$ white noise vector, and the $p \times 1$ vectors $\{\mathbf{h}_i\}$ represent the channel impulse response; the number of subchannels is thus p. Such a multichannel model may arise by a variety of means, e.g., by deploying multiple sensors, by fractional sampling the channel output when the continuous-time channel has excess bandwidth [16], or by introducing cyclic redundancy at the channel input [17].

A typical reformulation of this problem [16] involves the $mp \times 1$ vector processes

$$\mathbf{x}_n = [\,\overline{\mathbf{x}}_n^T \quad \overline{\mathbf{x}}_{n-1}^T \quad \cdots \quad \overline{\mathbf{x}}_{n-m+1}^T\,]^T$$
$$\mathbf{w}_n = [\,\overline{\mathbf{w}}_n^T \quad \overline{\mathbf{w}}_{n-1}^T \quad \cdots \quad \overline{\mathbf{w}}_{n-m+1}^T\,]^T$$

which are related via

$$\mathbf{x}_n = \mathcal{H}\mathbf{s}_n + \mathbf{w}_n \tag{2}$$

Manuscript received June 5, 2000; revised May 3, 2001. This work was supported by the National Science Foundation under Grants ECS-9970105 and CCR-9973133. The associate editor coordinating the review of this paper and approving it for publication was Prof. Dimitrios Hatzinakos.

R. López-Valcarce was with the Department of Electrical and Computer Engineering, University of Iowa, Iowa City 52242 USA. He is now with the Departamento de Tecnologías de las Comunicaciones, Universidad de Vigo, Vigo, Spain (e-mail: valcarce@gts.tsc.uvigo.es).

S. Dasgupta is with the Department of Electrical and Computer Engineering, University of Iowa, Iowa City 52242 USA (e-mail: dasgupta@engineering.uiowa.edu).

Publisher Item Identifier S 1053-587X(01)07058-1.

where $\mathcal{H}$ is an $mp \times (m+l)$ generalized Sylvester matrix constructed from the channel impulse response, i.e.,

$$\mathcal{H} = \begin{bmatrix} \mathbf{h}_0 & \mathbf{h}_1 & \cdots & \mathbf{h}_l & 0 & \cdots & 0 & 0 \\ 0 & \mathbf{h}_0 & \mathbf{h}_1 & \cdots & \mathbf{h}_l & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & \mathbf{h}_0 & \mathbf{h}_1 & \cdots & \mathbf{h}_l \end{bmatrix}. \quad (3)$$

With $d = m + l$, the signal vector $\mathbf{s}_n$ is given by

$$\mathbf{s}_n = [\, a_n \quad a_{n-1} \quad \cdots \quad a_{n-d+1} \,]^T.$$

As with most SOS-based methods, [13], [16], we assume that $\mathcal{H}$ has full column rank: a condition equivalent to requiring that all the subchannels be coprime [10].

There are two classes of SOS-based methods of particular note. The first, which was pioneered by Moulines *et al.* [13], relies on subspaced-based methods and requires no knowledge of input statistics whatsoever. There are obviously clear advantages to such a scheme. At the same time, the knowledge of the input statistics is often available, and its use should intuitively improve performance. The second class of schemes originates from the work of Tong *et al.* [16] and does exploit such knowledge. In fact, this original algorithm assumed that the transmitted sequence is white. The whiteness assumption on $\{a_n\}$ is crucial for the TXK algorithm, which uses the channel output autocorrelation matrices of lags 0 and 1

$$\mathcal{R}_x(0) = E[\mathbf{x}_n \mathbf{x}_n^H], \qquad \mathcal{R}_x(1) = E[\mathbf{x}_n \mathbf{x}_{n-1}^H] \quad (4)$$

to estimate $\mathcal{H}$ to within an unitary scaling constant. This was noted in [9], where a modification was proposed in order to deal with weakly correlated sources with unknown correlation (which must then be estimated).

Among the follow up papers, we note the interesting work of Afkhamie and Luo [1], who treat the colored source case. Using the notation outlined above, [1] assumes that the input autocorrelation sequence

$$r(k) = E[a_n a_{n-k}^*]$$

obeys for some integer a and constant $0 \le c < 1$

$$r(k) = c \cdot r(k-1) \qquad \text{for all } a \le k \le a + 2d - 2. \quad (5)$$

Under these conditions, [1] shows that the channel matrix can be estimated to within a unitary scaling constant from *all of*

$$\mathcal{R}_x(0), \mathcal{R}_x(1), \ldots, \mathcal{R}_x(a+d-1).$$

Even if $a = 0$, this requires the compilation of greater amount of output statistics than just $\mathcal{R}_x(0)$ and $\mathcal{R}_x(1)$, as is the case with [16]. It is thus computationally more onerous.

By contrast, our algorithm constitutes a direct extension of the TXK method, and therefore, it makes use of only $\mathcal{R}_x(0)$ and $\mathcal{R}_x(1)$. Further, using linear prediction techniques, we show that the only assumption needed on the input statistics to estimate $\mathcal{H}$ to within a scaling constant is the intuitively appealing condition:

$$E\left\{ \begin{bmatrix} a_n \\ a_{n-1} \\ \vdots \\ a_{n-d} \end{bmatrix} [\, a_n^* \quad a_{n-1}^* \quad \cdots \quad a_{n-d}^* \,] \right\} > 0. \quad (6)$$

Note the input vector in (6) has one more element than $\mathbf{s}_n$. Thus, the computational complexity *is comparable with that of the original TXK algorithm*.

As will be evident in this paper, such an extension, particularly the proof that the resulting channel estimate is to within a scaling constant of the "true" channel, is highly nontrivial. To this end, we make extensive use of the rich literature on linear prediction theory. *En route to our main result, we derive a singular value decomposition (SVD) of the normalized lag-1 source autocorrelation matrix, which we regard as a contribution of independent significance to linear prediction theory*. Simulation results are presented to show how, at high SNR values, the equalizer designed on the basis of the new algorithm consistently outperforms those designed by the method of [13] *for equalization delay values that lead to the best performance*. At these delays and SNR, the comparison with [1], which requires more output statistics, varies from comparable to favorable.

In our notation, $\mathbf{J}$ denotes the square shift matrix with ones in the first subdiagonal and zeros elsewhere. $\mathbf{X}$ is the exchange matrix with ones in the antidiagonal and zeros elsewhere. Superscripts $(\cdot)^*$, $(\cdot)^T$, and $(\cdot)^H$ denote, respectively, the conjugate, the transpose, and the conjugate transpose.

In Section II, the basics of the problem are presented, together with several results from linear prediction theory that will be useful. Some intermediate results are in Section III. The new algorithm is given in Section IV, and simulation results are in Section V.

II. Preliminaries

A. Assumptions

Define

$$\mathcal{R}_s(i) = E[\mathbf{s}_n \mathbf{s}_{n-i}^H],$$

as the $d \times d$ lag i autocorrelation matrix of the vector of input symbols $\mathbf{s}_n$. Let $\sigma_a^2 = E[|a_n|^2]$ and also define the autocorrelation vector

$$\mathbf{r}^H = \mathbf{e}_1^H \mathcal{R}_s(1) = E[a_n \mathbf{s}_{n-1}^H]. \quad (7)$$

In the sequel, we will make the following standard assumptions.

Assumption 1: The $(d+1) \times (d+1)$ source autocorrelation matrix

$$\begin{bmatrix} \sigma_a^2 & \mathbf{r}^H \\ \mathbf{r} & \mathcal{R}_s(0) \end{bmatrix}$$

i.e., the matrix in (6), is positive definite.

Assumption 2: The channel matrix $\mathcal{H}$ is tall and has full column rank.

In the sequel, we will assume that the noise is zero, as the noise component can be subtracted from the output autocorrelation matrices using a standard device [16]. In this case, one has, for all $i \geq 0$

$$\mathcal{R}_x(i) = \mathcal{H}\mathcal{R}_s(i)\mathcal{H}^H. \tag{8}$$

Thus, our goal is to find an estimate of $\mathcal{H}$ from (8) for $i = 0, 1$ and from the knowledge of $\mathcal{R}_s(0)$ and $\mathcal{R}_s(1)$.

To this end, we first undertake a whitening step. Under Assumption 1, where $\mathcal{L}$ is a lower triangular matrix with positive diagonal elements, one has the Cholesky decomposition

$$\mathcal{R}_s(0) = \mathcal{L}\mathcal{L}^H. \tag{9}$$

Introduce the normalized matrices

$$\mathbf{H} = \mathcal{H}\mathcal{L} \tag{10}$$

and

$$\mathbf{R}_s(1) = \mathcal{L}^{-1}\mathcal{R}_s(1)\mathcal{L}^{-H}. \tag{11}$$

Then, from (8), one has

$$\mathcal{R}_x(0) = \mathbf{H}\mathbf{H}^H \tag{12}$$

and

$$\mathcal{R}_x(1) = \mathbf{H}\mathbf{R}_s(1)\mathbf{H}^H. \tag{13}$$

Since $\mathcal{L}$ is known, the problem amounts to identifying $\mathbf{H}$ from (12) and (13).

Here is where the key point of departure from [16] lies. Under the assumption of white $\{a_n\}$ underlying [16], $\mathcal{L}$ is identity, and $\mathbf{R}_s(1)$ is $\mathbf{J}$. These two facts are critically exploited in [16] to obtain the algorithm that estimates $\mathbf{H}$ and to show that the class of all matrices $\mathbf{H}$ that simultaneously obey (12) and (13) are scaled versions of each other.

A possible modification of the TXK algorithm was proposed in [1, Sec. III-A]. It is claimed that the vector of colored symbols $\mathbf{s}_n$ can be seen as being generated by passing a corresponding white sequence through a coloring filter $\mathcal{L}$: $\mathbf{s}_n = \mathcal{L} \cdot \mathbf{s}_n^{\text{white}}$. In this way, one has $\mathbf{x}_n = \mathcal{H} \cdot \mathcal{L} \cdot \mathbf{s}_n^{\text{white}} = \mathbf{H}\mathbf{s}_n^{\text{white}}$, which suggests directly using the original TXK algorithm to identify $\mathbf{H}$. However, in order for the TXK approach to work, it is required that $\mathcal{R}_x(0) = \mathbf{H}\mathbf{H}^H$ and $\mathcal{R}_x(1) = \mathbf{H}\mathbf{J}\mathbf{H}^H$. Although the first condition is satisfied by construction, the second does not necessarily hold. This is because $E[(\mathbf{s}_n^{\text{white}})(\mathbf{s}_{n-1}^{\text{white}})^H] = \mathcal{L}^{-1}\mathcal{R}_s(1)\mathcal{L}^{-H} = \mathbf{R}_s(1) \neq \mathbf{J}$ in general. Thus, this approach is not guaranteed to work.

To treat the colored case of this paper, we must exploit the structure of $\mathbf{R}_s(1)$ and its relation to $\mathcal{L}$. To this end, in Section II-B, we establish certain connections between these matrices and the optimum forward prediction error filter (FPEF) of order d for the sequence $\{a_n\}$. These connections are crucial in resolving the colored source problem.

B. Optimum Forward Prediction Error Interpretation

Consider the standard linear prediction problem of finding coefficients θ_i such that the following quantity is minimized:

$$E\left\{\left|a_n - \sum_{k=1}^{d} \theta_k^* a_{n-k}\right|^2\right\}.$$

Suppose $\theta_k = \alpha_k$ are the minimizing parameters. Then, $1 + \sum_{k=1}^{d} \alpha_k^* z^{-k}$ is the transfer function of the FPEF of order d for the process $\{a_n\}$. With $\mathbf{r}$ as in (7), it is well known [7] that the coefficients of the dth-order predictor α_k are given by

$$\boldsymbol{\alpha} = [\,\alpha_1 \quad \cdots \quad \alpha_d\,]^T = -\mathcal{R}_s^{-1}(0)\mathbf{r}. \tag{14}$$

Now, since the last (respectively, first) $d-1$ rows (resp. columns) of $\mathcal{R}_s(1)$ are the first (resp. last) $d-1$ rows (resp. columns) of $\mathcal{R}_s(0)$, we have that

$$\mathcal{R}_s(1) = \mathbf{J}\mathcal{R}_s(0) + \mathbf{e}_1\mathbf{r}^H = \mathcal{R}_s(0)\mathbf{J} + \mathbf{X}\mathbf{r}^*\mathbf{e}_d^H. \tag{15}$$

Consequently, because of (7), the normalized matrix $\mathbf{R}_s(1)$ becomes

$$\mathbf{R}_s(1) = \mathcal{L}^{-1}(\mathbf{J} - \mathbf{e}_1\boldsymbol{\alpha}^H)\mathcal{L}. \tag{16}$$

This is the connection between $\mathbf{R}_s(1)$ and the FPEF of order d. Observe that $\mathbf{J} - \mathbf{e}_1\boldsymbol{\alpha}^H$ is a companion matrix whose eigenvalues coincide with the zeros of the FPEF.

Now, let the last row of $\mathcal{L}^{-1}$ be

$$\boldsymbol{\beta}^T = \mathbf{e}_d^H\mathcal{L}^{-1} = [\,\beta_{d-1} \quad \cdots \quad \beta_1 \quad \beta_0\,]. \tag{17}$$

Note from (17) and the fact that $\mathcal{L}$ is lower triangular with positive real diagonal elements that

$$\mathbf{X}\mathcal{L}^*\mathbf{e}_d = \beta_0^{-1}\mathbf{e}_1 \quad \text{and} \quad \beta_0\mathbf{e}_d = \mathcal{L}^{-1}\mathbf{e}_d \tag{18}$$

which will prove useful later.

It is known from linear prediction theory [7] that β_i^*/β_0 are the coefficients of the FPEF of order $d-1$, with β_0 real positive. In view of the order-update property of prediction filters [7], the vectors $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ are related via

$$\begin{bmatrix} 1 \\ \boldsymbol{\alpha} \end{bmatrix} = \frac{1}{\beta_0}\left(\begin{bmatrix} \mathbf{X}\boldsymbol{\beta} \\ 0 \end{bmatrix} + \alpha_d \begin{bmatrix} 0 \\ \boldsymbol{\beta}^* \end{bmatrix}\right). \tag{19}$$

The following fact about the order d FPEF is well known and is very important to the subsequent development.

Theorem 1: Under Assumption 1, all the zeros of $1 + \sum_{k=1}^{d} \alpha_k^* z^{-k}$, which is the order d FPEF for the process $\{a_n\}$, lie strictly inside the unit circle, and thus

$$|\alpha_d| < 1. \tag{20}$$

Another basic property is that $1/\beta_0^2$ is the variance of the forward (or backward) prediction error of order $d-1$. On the other hand, the variance of the prediction errors of order d is given by [7]

$$\gamma^2 = \frac{1}{\beta_0^2}(1 - |\alpha_d|^2). \tag{21}$$

Finally, it can be shown [2] that $\mathcal{R}_s(0)$ and the companion matrix $\mathbf{J}-\mathbf{e}_1\boldsymbol{\alpha}^H$ satisfy the following Lyapunov equation:

$$\mathcal{R}_s(0)-(\mathbf{J}-\mathbf{e}_1\boldsymbol{\alpha}^H)\mathcal{R}_s(0)(\mathbf{J}-\mathbf{e}_1\boldsymbol{\alpha}^H)^H=\gamma^2\mathbf{e}_1\mathbf{e}_1^H. \quad (22)$$

III. Some Intermediate Results

As in [16], consider an SVD of $\mathcal{R}_x(0)$:

$$\mathcal{R}_x(0)=[\,\mathbf{U}_1 \quad \mathbf{U}_2\,]\begin{bmatrix}\boldsymbol{\Sigma}^2 & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\mathbf{U}_1^H\\ \mathbf{U}_2^H\end{bmatrix}. \quad (23)$$

Here, $\boldsymbol{\Sigma}>\mathbf{0}$ is $d\times d$ diagonal. By Assumptions 1 and 2, $\mathbf{H}$ has full column rank. Thus, from (12)

$$\mathbf{H}=\mathbf{U}_1\boldsymbol{\Sigma}\mathbf{V} \quad (24)$$

for some unitary $\mathbf{V}$, which has to be estimated. For this purpose, consider, as in [16], the matrix

$$\mathbf{R}=\boldsymbol{\Sigma}^{-1}\mathbf{U}_1^H\mathcal{R}_x(1)\mathbf{U}_1\boldsymbol{\Sigma}^{-1}. \quad (25)$$

By direct verification using (13), one finds that $\mathbf{R}=\mathbf{V}\mathbf{R}_s(1)\mathbf{V}^H$. Thus, in view of (16)

$$\begin{aligned}\mathbf{R}&=\mathbf{V}\mathbf{R}_s(1)\mathbf{V}^H=\mathbf{V}\mathcal{L}^{-1}(\mathbf{J}-\mathbf{e}_1\boldsymbol{\alpha}^H)\mathcal{L}\mathbf{V}^H\\ &=\tilde{\mathbf{V}}(\mathbf{J}-\mathbf{e}_1\boldsymbol{\alpha}^H)\overline{\mathbf{V}}^H\end{aligned} \quad (26)$$

where we have introduced the matrices

$$\tilde{\mathbf{V}}=\mathbf{V}\mathcal{L}^{-1},\qquad \overline{\mathbf{V}}=\mathbf{V}\mathcal{L}^H. \quad (27)$$

Note that $\overline{\mathbf{V}}^H\tilde{\mathbf{V}}=\mathbf{I}$. Thus, (26) implies

$$\mathbf{R}\tilde{\mathbf{V}}=\tilde{\mathbf{V}}(\mathbf{J}-\mathbf{e}_1\boldsymbol{\alpha}^H),\qquad \mathbf{R}^H\overline{\mathbf{V}}=\overline{\mathbf{V}}(\mathbf{J}-\mathbf{e}_1\boldsymbol{\alpha}^H)^H.$$

Thus, partitioning

$$\tilde{\mathbf{V}}=[\,\tilde{\mathbf{v}}_1 \quad \cdots \quad \tilde{\mathbf{v}}_d\,] \quad (28)$$

and

$$\overline{\mathbf{V}}=[\,\overline{\mathbf{v}}_1 \quad \cdots \quad \overline{\mathbf{v}}_d\,] \quad (29)$$

columnwise, one has

$$\mathbf{R}\tilde{\mathbf{v}}_i=\tilde{\mathbf{v}}_{i+1}-\alpha_i^*\tilde{\mathbf{v}}_1,\qquad i=1,\ldots,d-1 \quad (30)$$

$$\mathbf{R}\tilde{\mathbf{v}}_d=-\alpha_d^*\tilde{\mathbf{v}}_1 \quad (31)$$

$$\mathbf{R}^H\overline{\mathbf{v}}_i=\overline{\mathbf{v}}_{i-1},\qquad i=2,\ldots,d \quad (32)$$

$$\mathbf{R}^H\overline{\mathbf{v}}_1=-\sum_{i=1}^{d}\alpha_i\overline{\mathbf{v}}_i. \quad (33)$$

Since $\mathcal{L}$ is known, it suffices to estimate either $\tilde{\mathbf{V}}$ or $\overline{\mathbf{V}}$.

We will exploit now the linear prediction approach of the previous section to present two lemmas concerning $\tilde{\mathbf{V}}$ and $\overline{\mathbf{V}}$ that have two benefits.

- They lead to an efficient means of estimating $\mathbf{V}$.
- They show that this $\mathbf{V}$ is unique to within a scaling constant.

Note that neither fact is immediately obvious from (30)–(33) alone.

The first lemma exposes the structure of the singular values of $\mathbf{R}$. Note that these singular values are the same as those of $\mathbf{R}_s(1)$. Therefore, they are independent of the channel matrix because they are determined by the autocorrelation of the symbols $\{a_n\}$ alone.

Lemma 1: There exists a $d\times d$ unitary matrix $\mathbf{Q}$ such that $\mathbf{R}_s(1)=\mathbf{Q}\mathbf{D}$, where $\mathbf{D}$ is $d\times d$ diagonal given by

$$\mathbf{D}=\operatorname{diag}(\,1 \quad \cdots \quad 1 \quad |\alpha_d|\,). \quad (34)$$

Proof: See Appendix A.

Observe that $\mathbf{R}_s(1)=\mathbf{Q}\mathbf{D}$ constitutes an SVD of $\mathbf{R}_s(1)$. Therefore

$$\mathbf{R}=(\mathbf{V}\mathbf{Q})\mathbf{D}\mathbf{V}^H$$

constitutes an SVD of $\mathbf{R}$. In particular, the columns of $\mathbf{V}$ are right singular vectors of $\mathbf{R}$.

The significance of lemma 1 resides in the fact that because of Theorem 1, the smallest singular value of $\mathbf{R}_s(1)$ (and, therefore, of $\mathbf{R}$) is given by $|\alpha_d|$ and is hence unique. This uniqueness allows us to extract the matrix $\mathbf{V}$ from $\mathbf{R}$, up to a unitary scaling constant. The fact that, under the nonsingularity assumption on the $(d+1)\times(d+1)$ lag-0 source autocorrelation matrix, the smallest singular value of the normalized lag-1 autocorrelation matrix has multiplicity one is, in our opinion, a result of independent interest. The next lemma provides the key to the estimation of $\mathbf{V}$.

Lemma 2: The vector $\beta_0^{-1}\tilde{\mathbf{v}}_1$ is a unit-norm left singular vector of the matrix $\mathbf{R}$ associated with its smallest singular value (under Assumption 1) $|\alpha_d|$.

Proof: See Appendix B.

IV. Modified Algorithm

In view of the results of the previous section, it is possible to estimate the columns of the matrix $\tilde{\mathbf{V}}$ as follows: First, extract $\tilde{\mathbf{v}}_1$ as β_0 times the left singular vector of $\mathbf{R}$ associated with the smallest singular value; then, use the recurrence (30) in order to estimate the remaining columns. For convenience, the algorithm is detailed next.

1) Perform an SVD of $\mathcal{R}_x(0)$ as in (23), and form the matrix $\mathbf{R}=\boldsymbol{\Sigma}^{-1}\mathbf{U}_1^H\mathcal{R}_x(1)\mathbf{U}_1\boldsymbol{\Sigma}^{-1}$.
2) Let $\hat{\mathbf{v}}_1$ be a unit-norm left singular vector of $\mathbf{R}$ associated with the smallest singular value.
3) Compute the FPEF coefficients α_i via (14). Then, for $i=1,2,\ldots,d-1$, let $\hat{\mathbf{v}}_{i+1}=\mathbf{R}\hat{\mathbf{v}}_i+\alpha_i^*\hat{\mathbf{v}}_1$.
4) With $\mathcal{L}$ the Cholesky factor of $\mathcal{R}_s(0)$, obtain β_0 via (17). The normalized channel matrix estimate is then

$$\hat{\mathbf{H}}=\beta_0\mathbf{U}_1\boldsymbol{\Sigma}[\,\hat{\mathbf{v}}_1 \quad \cdots \quad \hat{\mathbf{v}}_d\,]\mathcal{L}$$

so that the unnormalized channel matrix estimate $\hat{\mathcal{H}}=\hat{\mathbf{H}}\mathcal{L}^{-1}$ is given by

$$\hat{\mathcal{H}}=\beta_0\mathbf{U}_1\boldsymbol{\Sigma}[\,\hat{\mathbf{v}}_1 \quad \cdots \quad \hat{\mathbf{v}}_d\,].$$

5) The columns of the matrix $\mathcal{G}_{\mathrm{ZF}}$ given by

$$\mathcal{G}_{\mathrm{ZF}}=\beta_0\mathbf{U}_1\boldsymbol{\Sigma}^{-1}[\,\hat{\mathbf{v}}_1 \quad \cdots \quad \hat{\mathbf{v}}_d\,]\mathcal{R}_s(0)$$

constitute zero-forcing equalizers. In the absence of noise

$$\mathcal{G}_{\mathrm{ZF}}^{H}\mathbf{x}_n = e^{-j\theta}\mathbf{s}_n \tag{35}$$

for some real θ.

We will comment on this algorithm after stating our main result.

Theorem 2: Consider the algorithm above with the various quantities in Step 1 defined in (24). Consider any $\mathcal{H}$ that simultaneously satisfies (8) for $i = 0, 1$. Then, under Assumptions 1 and 2, there exists real θ such that $\hat{\mathcal{H}}$ obtained in Step 4 obeys

$$\hat{\mathcal{H}} = e^{j\theta}\mathcal{H}.$$

Further, with $\mathbf{w}_n = 0$ in (2), the matrix $\mathcal{G}_{\mathrm{ZF}}$ obtained in step 5 obeys (35).

Proof: See Appendix C.

Several comments on the algorithm are in order. First, note that since the space of left singular vectors of $\mathbf{R}$ corresponding to this singular value has dimension 1, $\hat{\mathbf{v}}_1$ is easily determined. The chain of equations in Step 3 also determine $[\hat{\mathbf{v}}_2 \ \cdots \ \hat{\mathbf{v}}_d]$ efficiently (no matrix inversion). The statistics of $\{a_n\}$ provide β_0 and $\mathcal{L}$, just as the output statistics provide $\mathcal{R}_x(1)$ and $\mathbf{U}_1$, $\mathbf{\Sigma}$ (via an SVD) and, hence, $\mathbf{R}$. Note also that when $\{a_n\}$ is white, i.e., the case covered in [16], one has $\alpha_i = 0$, $\beta_0 = \sigma_a^2$, and $\mathcal{L} = \sigma_a\mathbf{I}$, and the algorithm recovers as a special case its counterpart in [16].

V. Simulation Results

A series of simulation experiments have been performed to test the new algorithm. For comparison purposes, the algorithms of Moulines *et al.* [13] and Afkhamie and Luo [1] were also implemented in the same environment.

The channel impulse response corresponds to $p = 2$ subchannels with length $l = 5$ and coefficients

$$[\mathbf{h}_0 \ \cdots \ \mathbf{h}_5] = \begin{bmatrix} 0.1 & -0.12 & 0.43 & 0.87 & -0.12 & 0.04 \\ 0.15 & 0.45 & -0.76 & 0.21 & -0.15 & 0.11 \end{bmatrix}.$$

The input symbols $\{a_n\}$ are drawn from a 4-QAM constellation according to the following rule. Let $\{b_n\}$ be the input stream of independent and identically distributed bits, i.e., $b_n \in \{0, 1\}$. Then

$$a_n = \begin{cases} -1 + j & \text{if } (b_n\ b_{n-1}) = (0\ 0) \\ +1 + j & \text{if } (b_n\ b_{n-1}) = (0\ 1) \\ -1 - j & \text{if } (b_n\ b_{n-1}) = (1\ 0) \\ +1 - j & \text{if } (b_n\ b_{n-1}) = (1\ 1). \end{cases}$$

This generates a colored symbol sequence $\{a_n\}$ with autocorrelation

$$E[a_n a_{n-k}^*] = \begin{cases} 2, & k = 0, \\ \pm j, & k = \pm 1, \\ 0, & \text{else.} \end{cases}$$

Two sets of experiments were conducted: First, we took an equalizer length of $m = 6$, which yields $d = m + l = 11$. Then, we considered the case $m = 12$ for which $d = m + l = 17$. The algorithm of Afkhamie and Luo makes use of the autocorrelation matrices $\mathcal{R}_x(0), \ldots, \mathcal{R}_x(a + d - 2)$, where a is a parameter that should be chosen in terms of the source autocorrelation [1]. In this case, as this autocorrelation vanishes for lags greater than 1, it suffices to take $a = 2$.

Additive white Gaussian noise $\mathbf{w}_n$ was added to the channel output so that the model becomes $\mathbf{x}_n = \mathcal{H}\mathbf{s}_n + \mathbf{w}_n$. The signal-to-noise ratio (SNR) is defined as

$$\mathrm{SNR} = 10 \log \frac{\mathrm{trace}\ E[(\mathcal{H}\mathbf{s}_n)(\mathcal{H}\mathbf{s}_n)^H]}{\mathrm{trace}\ E[\mathbf{w}_n\mathbf{w}_n^H]}.$$

The noise variance estimate $\hat{\sigma}_w^2$ was taken as the smallest eigenvalue of the matrix $\mathcal{R}_x(0)$ and then substracted to provide the algorithms with denoised autocorrelation estimates. For simplicity, knowledge of the channel length l was assumed; see [11] for a discussion on how to blindly estimate the channel order from the output statistics.

Once the channel matrix has been estimated by the algorithm of Moulines *et al.*, the zero-forcing equalizers are obtained as the rows of the pseudoinverse: $\hat{\mathcal{H}}^{\#} = \mathcal{G}_{\mathrm{ZF}}^{H}$. For all the algorithms, the minimum mean-squared error (MMSE) equalizers are computed as

$$\mathcal{G}_{\mathrm{MMSE}} = (\mathbf{I} - \hat{\sigma}_w^2\mathcal{R}_x^{-1}(0))\mathcal{G}_{\mathrm{ZF}} \tag{36}$$

where $\mathcal{R}_x(0)$ represents the *undenoised* autocorrelation matrix of the channel output. The expression (36) was originally derived in [14] for white symbols, but it can be readily checked that it is also valid for colored symbols. Finally, the phase ambiguity (inherent to all SOS-based methods) is removed before evaluating the equalizers' performance.

A. Equalizer Length $m = 6$

The different rows of $\mathcal{G}_{\mathrm{MMSE}}$ correspond to different equalization delays, and therefore, they yield different performances. Fig. 1 shows the symbol error rate (SER) as a function of the delay (between 0 and $d - 1$) for four different values of the SNR and using $K = 2000$ symbols for the estimation of the autocorrelation matrices. It is seen that for low SNR, the three methods perform poorly. As the SNR is increased, the performance of the equalizers associated with extremal delays improves much more slowly than that of the intermediate delay equalizers. For these intermediate delays, the new algorithm consistently yields lower SER than the method of Moulines *et al.* The equalizers of Afkhamie and Luo are seen to lie somewhere in between, although in some cases they may outperform the other two methods or even present the highest SER. Fig. 2 shows the variation of SER with SNR for the equalizers with delays 1, 3 and, 7 with $K = 2000$ symbols.

In Fig. 3, the normalized root-mean-square error (NRMSE) of the channel estimate is shown: first as a function of the SNR with $K = 2000$ and then as a function of K for SNR $= 20$ dB. The NRMSE is defined as

$$\mathrm{NRMSE} = \frac{1}{||h||}\sqrt{\frac{1}{T}\sum_{t=1}^{T}||\hat{h}(t) - h||^2}$$

where h is the vector of channel coefficients, and T is the number of Monte Carlo trials (100 for our experiment). The

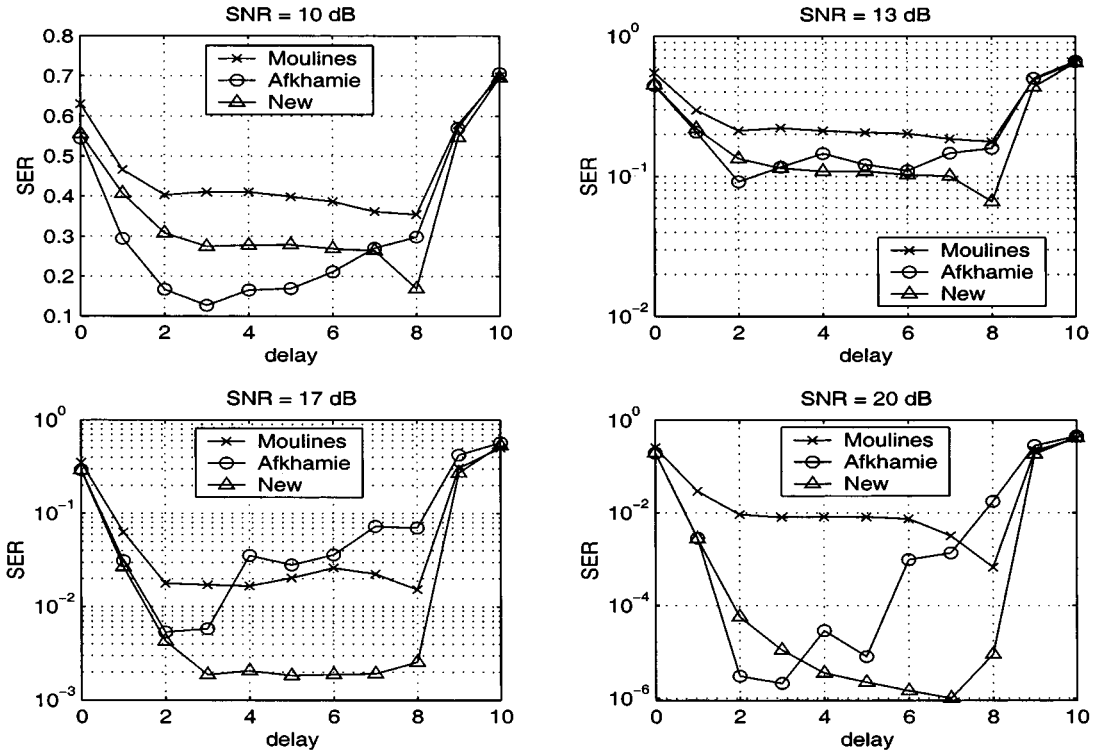

Fig. 1. SER obtained by the length $m = 6$ MMSE equalizers (averaged over 100 independent runs).

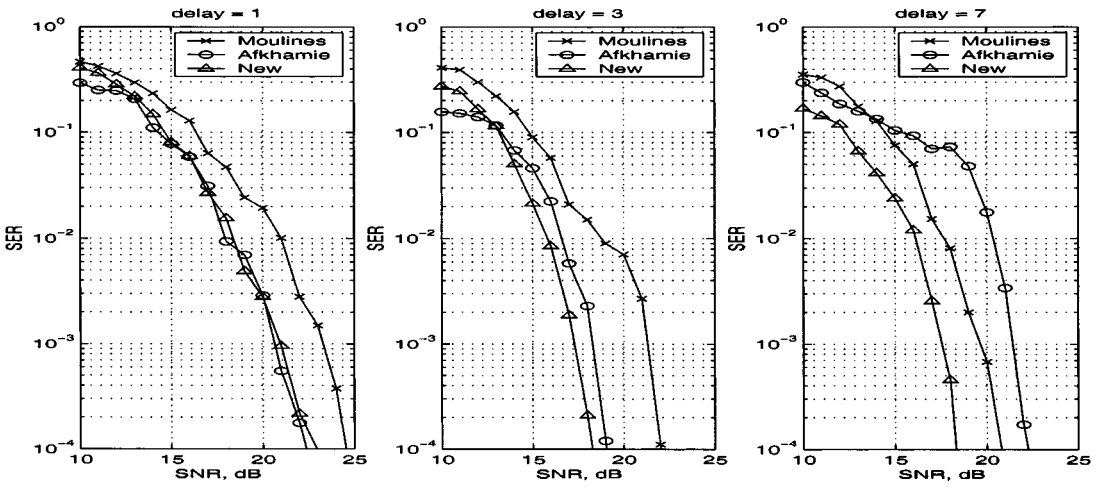

Fig. 2. SER obtained by the length $m = 6$ MMSE equalizers (averaged over 100 independent runs).

new method seems to present a smaller error for all values of the SNR and K than the other two schemes.

B. Equalizer Length $m = 12$

In this case, one obtains equalizers with associated delays 0 through 16. The performance of the MMSE equalizers as a function of delay using $K = 2000$ is shown in Fig. 4. Again, as the SNR increases, the SER decreases much faster for the intermediate delays than for the extremal ones. For extremal delays, the three methods perform similarly. For intermediate delays and low SNR, the equalizers obtained by Moulines' and Afkhamie's methods yield lower SER than that of the new scheme. However, as the SNR is increased, the new algorithm eventually outperforms the other two. Fig. 5 shows the variation of SER with SNR for the equalizers with delays 1, 5, 10, and 15 with $K = 2000$ symbols.

Surprisingly, in this case, the Afkhamie and Luo method does not present a clear advantage with respect Moulines' and the new scheme, despite the fact that it uses more statistical information about the transmitted symbol sequence. The main difficulty seems to be the need to resolve the phase ambiguities in the columns of the estimate of the unitary matrix $\mathbf{V}$. Although in the formulation of Afkhamie's algorithm [1] only sign ambiguities of the type ± 1 were considered, when dealing with complex signals, the resulting ambiguities are of the type $e^{j\theta}$. Even though these can be resolved in principle in the same manner as

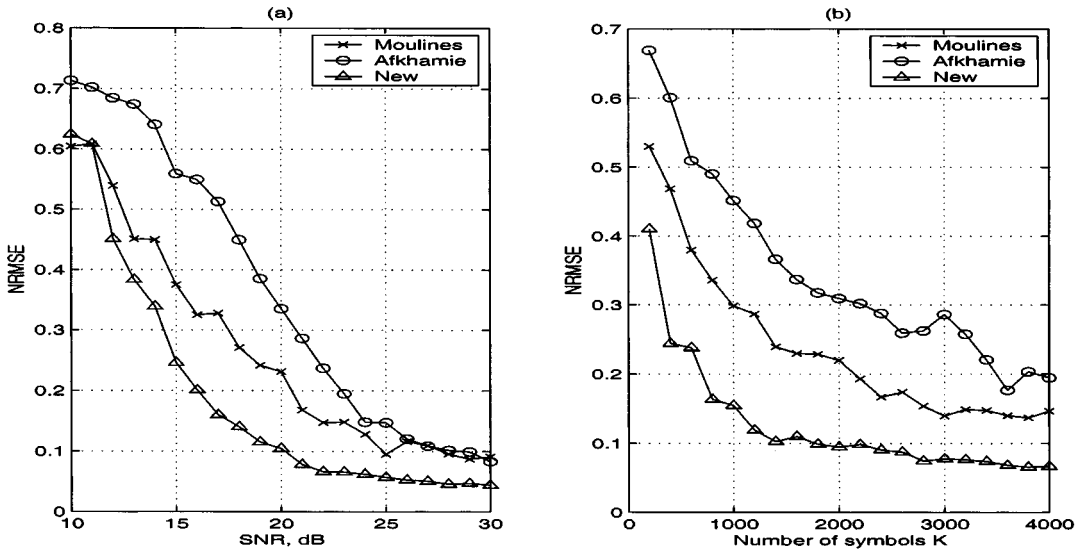

Fig. 3. NRMSE of the estimated channel response using $m = 6$ versus (a) SNR using $K = 2000$ data samples. (b) Number of samples K for SNR $= 20$ dB (averaged over 100 runs).

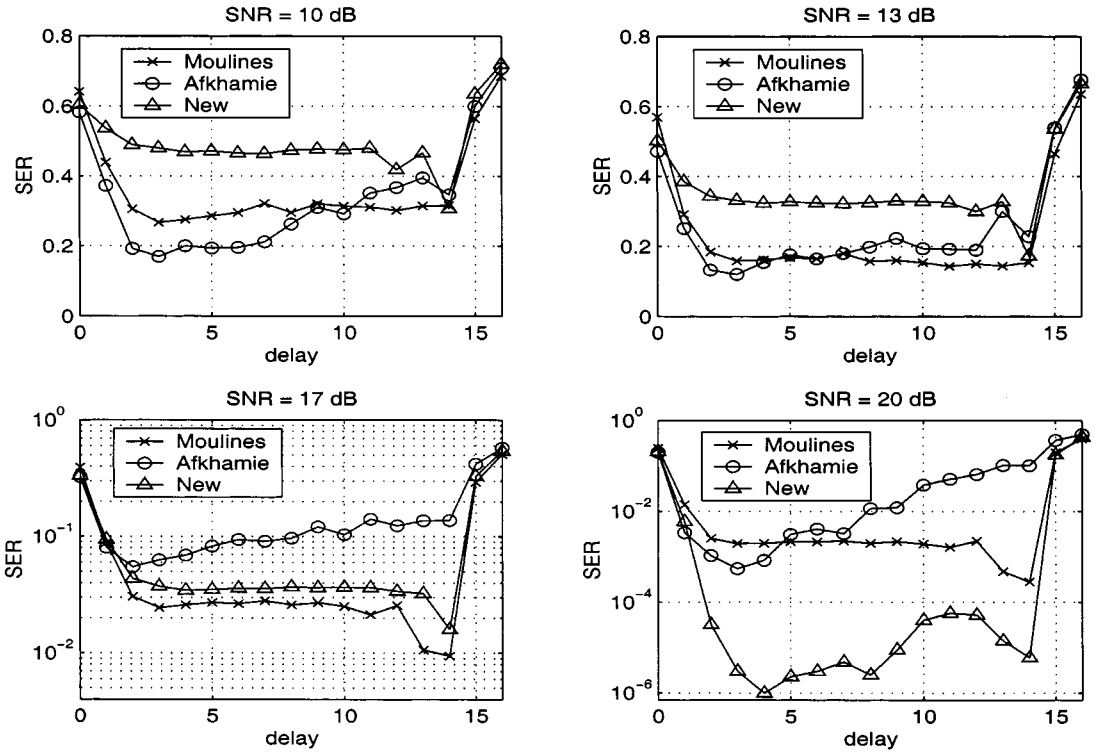

Fig. 4. SER obtained by the length $m = 12$ MMSE equalizers (averaged over 100 independent runs).

in [1], the procedure seems to be not as well behaved numerically. This problem is also evident from Fig. 6, which shows the NRMSE obtained by using the three methods with $m = 12$.

VI. Conclusion

An extension of the algorithm of Tong *et al.* [16] for blind identification of FIR SIMO channels has been developed in order to account for source correlation. The new algorithm estimates the channel from second-order statistics of the observed signal for arbitrary but known transmitted symbol coloring. The computational complexity of the method is comparable with that of the original TXK algorithm designed for white sources, which is considerably less than that of previous approaches. In addition, the method is valid under a mild condition on the correlation. Simulations have shown that the new algorithm compares favorably with other SOS-based methods that are capable of dealing with source coloring.

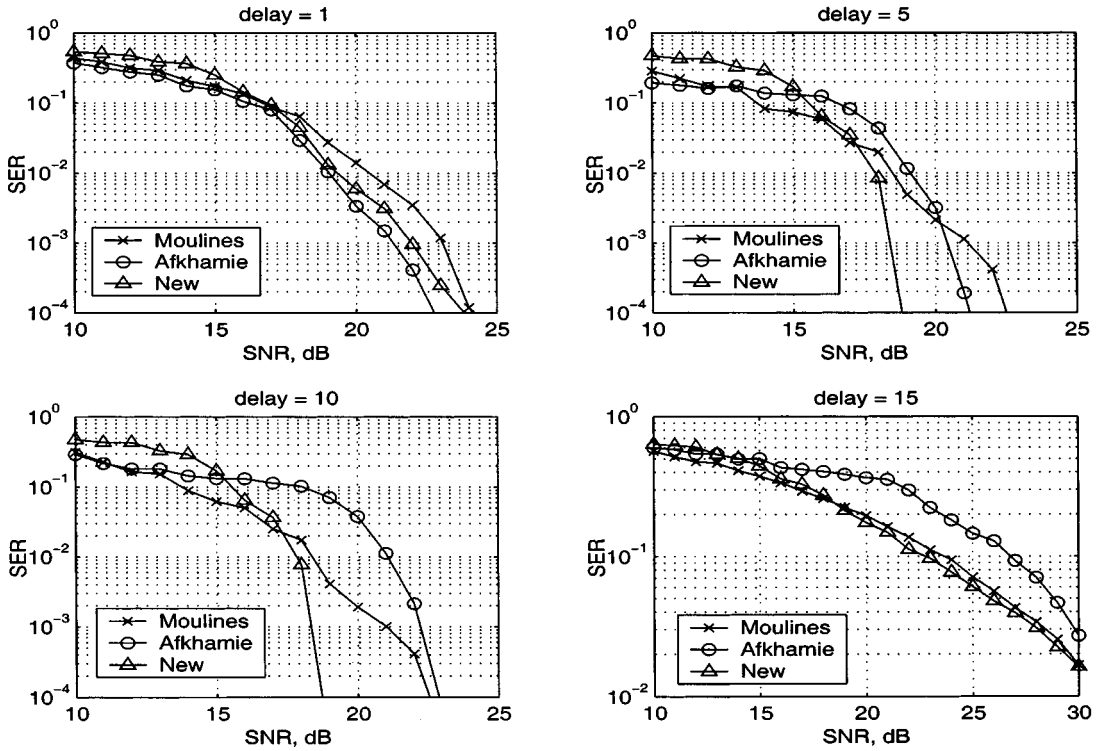

Fig. 5. SER obtained by the length $m = 12$ MMSE equalizers (averaged over 100 independent runs).

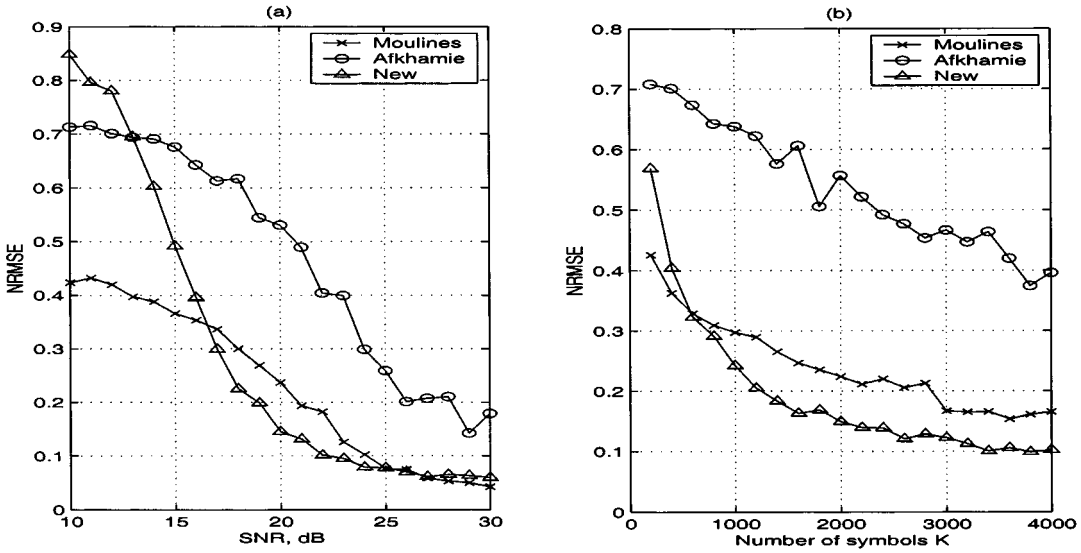

Fig. 6. NRMSE of the estimated channel response using $m = 12$ versus (a) SNR using $K = 2000$ data samples. (b) Number of samples K, for SNR = 20 dB (averaged over 100 runs).

APPENDIX A
PROOF OF LEMMA 1

Suppose first that $\alpha_d \neq 0$. In that case, it suffices to show that $\mathbf{Q} = \mathbf{R}_s(1)\mathbf{D}^{-1}$ is unitary. One has

$$\mathbf{Q}\mathbf{Q}^H = \mathbf{R}_s(1)\mathbf{D}^{-2}\mathbf{R}_s(1)^H \tag{37}$$

and, because of (21), $\mathbf{D}^{-2}$ can be written as

$$\mathbf{D}^{-2} = \mathbf{I} - (1 - |\alpha_d|^{-2})\mathbf{e}_d\mathbf{e}_d^H = \mathbf{I} + \frac{\gamma^2\beta_0^2}{|\alpha_d|^2}\,\mathbf{e}_d\mathbf{e}_d^H.$$

For convenience, let $\mathbf{F} = \mathbf{J} - \mathbf{e}_1\boldsymbol{\alpha}^H$; then, from (16), $\mathbf{R}_s(1) = \mathcal{L}^{-1}\mathbf{F}\mathcal{L}$. One has from (9) and (18)

$$\begin{aligned}\mathbf{Q}\mathbf{Q}^H &= \mathcal{L}^{-1}\mathbf{F}\mathcal{L}\left(\mathbf{I} + \frac{\gamma^2\beta_0^2}{|\alpha_d|^2}\,\mathbf{e}_d\mathbf{e}_d^H\right)\mathcal{L}^H\mathbf{F}^H\mathcal{L}^{-H} \\ &= \mathcal{L}^{-1}\left(\mathbf{F}\mathcal{L}\mathcal{L}^H\mathbf{F}^H + \frac{\gamma^2\beta_0^2}{|\alpha_d|^2}\,\mathbf{F}(\mathcal{L}\mathbf{e}_d)(\mathcal{L}\mathbf{e}_d)^H\mathbf{F}^H\right)\mathcal{L}^{-H} \\ &= \mathcal{L}^{-1}\left(\mathbf{F}\mathcal{R}_s(0)\mathbf{F}^H + \frac{\gamma^2}{|\alpha_d|^2}\,\mathbf{F}\mathbf{e}_d\mathbf{e}_d^H\mathbf{F}^H\right)\mathcal{L}^{-H}. \end{aligned} \tag{38}$$

Now, from (22), one has $\mathbf{F}\mathcal{R}_s(0)\mathbf{F}^H = \mathcal{R}_s(0) - \gamma^2\mathbf{e}_1\mathbf{e}_1^H$. In addition, $\mathbf{F}\mathbf{e}_d = -\alpha_d^*\mathbf{e}_1$. Substituting these in (38), from (9)

$$\begin{aligned}\mathbf{Q}\mathbf{Q}^H &= \mathcal{L}^{-1}\left(\mathcal{R}_s(0) - \gamma^2\mathbf{e}_1\mathbf{e}_1^H + \frac{\gamma^2}{|\alpha_d|^2}|\alpha_d|^2\mathbf{e}_1\mathbf{e}_1^H\right)\mathcal{L}^{-H} \\ &= \mathcal{L}^{-1}\mathcal{R}_s(0)\mathcal{L}^{-H} = \mathbf{I}\end{aligned}$$

so that $\mathbf{Q}$ is unitary.

Now, suppose $\alpha_d = 0$. In that case, we will show that with $\mathbf{u}$, which is the vector given by

$$\mathbf{u} = \gamma\mathcal{L}^{-1}\mathbf{e}_1 \tag{39}$$

the matrix $\mathbf{Q} = \mathbf{R}_s(1) + \mathbf{u}\mathbf{e}_d^H$ satisfies the requirements of the lemma. First, note that the last column of $\mathbf{R}_s(1)$ is the zero vector. Indeed, from (18)

$$\mathbf{R}_s(1)\mathbf{e}_d = \mathcal{L}^{-1}\mathbf{F}\mathcal{L}\mathbf{e}_d = \frac{1}{\beta_0}\mathcal{L}^{-1}\mathbf{F}\mathbf{e}_d = \mathbf{0} \tag{40}$$

since $\mathbf{F}\mathbf{e}_d = -\alpha_d^*\mathbf{e}_1 = \mathbf{0}$. Then

$$\begin{aligned}\mathbf{Q}\mathbf{D} &= (\mathbf{R}_s(1) + \mathbf{u}\mathbf{e}_d^H)(\mathbf{I} - \mathbf{e}_d\mathbf{e}_d^H) \\ &= \mathbf{R}_s(1) - \mathbf{R}_s(1)\mathbf{e}_d\mathbf{e}_d^H + \mathbf{u}\mathbf{e}_d^H - \mathbf{u}(\mathbf{e}_d^H\mathbf{e}_d)\mathbf{e}_d^H \\ &= \mathbf{R}_s(1) - \mathbf{0} + \mathbf{u}\mathbf{e}_d^H - \mathbf{u}\mathbf{e}_d^H = \mathbf{R}_s(1).\end{aligned}$$

Thus, it remains to be shown that $\mathbf{Q}$ is unitary. To do so, observe that in view of (9) and (22)

$$\begin{aligned}\mathbf{R}_s(1)\mathbf{R}_s(1)^H &= \mathcal{L}^{-1}\mathbf{F}\mathcal{R}_s(0)\mathbf{F}^H\mathcal{L}^{-H} \\ &= \mathcal{L}^{-1}(\mathcal{R}_s(0) - \gamma^2\mathbf{e}_1\mathbf{e}_1^H)\mathcal{L}^{-H} \\ &= \mathbf{I} - \gamma^2(\mathcal{L}^{-1}\mathbf{e}_1)(\mathcal{L}^{-1}\mathbf{e}_1)^H = \mathbf{I} - \mathbf{u}\mathbf{u}^H.\end{aligned}$$

Therefore

$$\begin{aligned}\mathbf{Q}\mathbf{Q}^H &= (\mathbf{R}_s(1) + \mathbf{u}\mathbf{e}_d^H)(\mathbf{R}_s(1)^H + \mathbf{e}_d\mathbf{u}^H) \\ &= \mathbf{R}_s(1)\mathbf{R}_s(1)^H + \mathbf{R}_s(1)\mathbf{e}_d\mathbf{u}^H + \mathbf{u}\mathbf{e}_d^H\mathbf{R}_s(1)^H \\ &\quad + \mathbf{u}(\mathbf{e}_d^H\mathbf{e}_d)\mathbf{u}^H \\ &= \mathbf{R}_s(1)\mathbf{R}_s(1)^H + \mathbf{u}\mathbf{u}^H = \mathbf{I}\end{aligned}$$

since, from (40), one has $\mathbf{R}_s(1)\mathbf{e}_d = \mathbf{0}$. Thus, $\mathbf{Q}$ is unitary. ■

Appendix B
Proof of Lemma 2

In order to prove Lemma 2, we will need the following result.

Lemma 3: With $\mathbf{R}$, $\tilde{\mathbf{v}}_i$, and α_d as above, there holds

$$\mathbf{R}^H\tilde{\mathbf{v}}_1 = -\alpha_d\tilde{\mathbf{v}}_d.$$

Proof: Observe that (33) yields

$$\mathbf{R}^H\overline{\mathbf{V}}\begin{bmatrix}1\\ \alpha_1\\ \vdots\\ \alpha_{d-1}\end{bmatrix} = -\alpha_d\overline{\mathbf{v}}_d. \tag{41}$$

In view of (17) and (19), the vector in brackets in (41) is

$$\begin{bmatrix}1\\ \alpha_1\\ \vdots\\ \alpha_{d-1}\end{bmatrix} = \frac{1}{\beta_0}(\mathbf{X}\boldsymbol{\beta} + \alpha_d\mathbf{J}\boldsymbol{\beta}^*) = \frac{1}{\beta_0}(\mathbf{X}\mathcal{L}^{-T} + \alpha_d\mathbf{J}\mathcal{L}^{-H})\mathbf{e}_d. \tag{42}$$

Hence, using (42), (41) successively yields

$$\mathbf{R}^H\overline{\mathbf{V}}(\mathbf{X}\mathcal{L}^{-T} + \alpha_d\mathbf{J}\mathcal{L}^{-H})\mathbf{e}_d = -\alpha_d\beta_0\overline{\mathbf{v}}_d$$

and

$$\mathbf{R}^H\mathbf{V}\mathcal{L}^H\mathbf{X}\mathcal{L}^{-T}\mathbf{e}_d = -\alpha_d(\beta_0\overline{\mathbf{v}}_d + \mathbf{R}^H\overline{\mathbf{V}}\mathbf{J}\mathcal{L}^{-H}\mathbf{e}_d). \tag{43}$$

Now, since $\mathcal{R}_s(0)$ is Hermitian Toeplitz, $\mathbf{X}\mathcal{R}_s(0) = \mathcal{R}_s^*(0)\mathbf{X}$. Because of (9), this yields

$$\mathcal{L}^H\mathbf{X}\mathcal{L}^{-T} = \mathcal{L}^{-1}\mathbf{X}\mathcal{L}^*. \tag{44}$$

Further, from (17) and the fact that the diagonal elements of the lower triangular matrix $\mathcal{L}$, which is the Cholesky factor of $\mathcal{R}_s(0)$, are positive real

$$[\mathbf{0} \quad \cdots \quad \mathbf{0} \quad \overline{\mathbf{v}}_d]\mathcal{L}^{-H}\mathbf{e}_d = [\mathbf{0} \quad \cdots \quad \mathbf{0} \quad \overline{\mathbf{v}}_d]\boldsymbol{\beta}^* = \beta_0\overline{\mathbf{v}}_d. \tag{45}$$

In addition, (32) shows that

$$\mathbf{R}^H\overline{\mathbf{V}}\mathbf{J} = [\overline{\mathbf{v}}_1 \quad \cdots \quad \overline{\mathbf{v}}_{d-1} \quad \mathbf{0}].$$

Thus, because of (44) and (45), (43) reads as

$$\begin{aligned}\mathbf{R}^H\mathbf{V}\mathcal{L}^{-1}\mathbf{X}\mathcal{L}^*\mathbf{e}_d &= -\alpha_d(\beta_0\overline{\mathbf{v}}_d + [\overline{\mathbf{v}}_1 \quad \cdots \quad \overline{\mathbf{v}}_{d-1} \quad \mathbf{0}]\mathcal{L}^{-H}\mathbf{e}_d) \\ &= -\alpha_d\overline{\mathbf{V}}\mathcal{L}^{-H}\mathbf{e}_d.\end{aligned} \tag{46}$$

Now, from (18) and the definitions of $\tilde{\mathbf{V}}$ and $\overline{\mathbf{V}}$, one has

$$\mathbf{R}^H\mathbf{V}\mathcal{L}^{-1}\mathbf{X}\mathcal{L}^*\mathbf{e}_d = \beta_0^{-1}\mathbf{R}^H\mathbf{V}\mathcal{L}^{-1}\mathbf{e}_1 = \beta_0^{-1}\mathbf{R}^H\tilde{\mathbf{V}}\mathbf{e}_1 \tag{47}$$

and

$$-\alpha_d\overline{\mathbf{V}}\mathcal{L}^{-H}\mathbf{e}_d = -\beta_0^{-1}\alpha_d\overline{\mathbf{V}}\mathcal{L}^{-H}\mathcal{L}^{-1}\mathbf{e}_d = -\beta_0^{-1}\alpha_d\tilde{\mathbf{V}}\mathbf{e}_d. \tag{48}$$

In view of (46), (47) equals (48):

$$\beta_0^{-1}\mathbf{R}^H\tilde{\mathbf{V}}\mathbf{e}_1 = -\beta_0^{-1}\alpha_d\tilde{\mathbf{V}}\mathbf{e}_d$$

i.e., $\mathbf{R}^H\tilde{\mathbf{v}}_1 = -\alpha_d\tilde{\mathbf{v}}_d$, which proves the lemma. ■

Now, we can proceed to prove Lemma 2. First, observe that since $\tilde{\mathbf{v}}_1 = \mathbf{V}\mathcal{L}^{-1}\mathbf{e}_1$, the norm of $\tilde{\mathbf{v}}_1$ is the same as that of $\mathcal{L}^{-1}\mathbf{e}_1$. One has $(\mathcal{L}^{-1}\mathbf{e}_1)^H(\mathcal{L}^{-1}\mathbf{e}_1) = \mathbf{e}_1^H\mathcal{R}_s^{-1}(0)\mathbf{e}_1$, which is the (1, 1) element of $\mathcal{R}_s^{-1}(0)$. Since $\mathcal{R}_s(0)$ is Toeplitz, its inverse $\mathcal{R}_s^{-1}(0)$ is symmetric about the antidiagonal so that its (1, 1) and (d, d) elements coincide, but it is easily seen that as $\mathcal{L}$ is lower triangular, because of (17), one has $\mathbf{e}_d^H\mathcal{R}_s^{-1}(0)\mathbf{e}_d = \beta_0^2$. Thus, $\beta_0^{-1}\tilde{\mathbf{v}}_1$ has unit norm. Similarly, the (squared) norm of $\tilde{\mathbf{v}}_d$ is $\tilde{\mathbf{v}}_d^H\tilde{\mathbf{v}}_d = (\mathcal{L}^{-1}\mathbf{e}_d)^H(\mathcal{L}^{-1}\mathbf{e}_d) = \beta_0^2$.

From (30), one has $\mathbf{R}\tilde{\mathbf{v}}_d = -\alpha_d^*\tilde{\mathbf{v}}_1$. Premultiply this by $\mathbf{R}^H$ and use the result from Lemma 3 to obtain

$$\mathbf{R}^H\mathbf{R}\tilde{\mathbf{v}}_d = |\alpha_d|^2\tilde{\mathbf{v}}_d \tag{49}$$

which shows that $\beta_0^{-1}\tilde{\mathbf{v}}_d$ is a *right* singular vector of $\mathbf{R}$ associated with the smallest singular value. However, using (31)

$$\mathbf{R}(\beta_0^{-1}\tilde{\mathbf{v}}_d) = -\alpha_d^*(\beta_0^{-1}\tilde{\mathbf{v}}_1)$$

which must then equal $|\alpha_d|\mathbf{w}$, where $\mathbf{w}$ is a left unit-norm singular vector. This shows that $\beta_0^{-1}\tilde{\mathbf{v}}_1$ is a left unit-norm singular vector of $\mathbf{R}$ associated with its smallest singular value. ■

APPENDIX C
PROOF OF THEOREM 2

Because of Theorem 1, $|\alpha_d| < 1$. Thus, from Lemma 1, the smallest singular value of $\mathbf{R}$ has multiplicity one. Hence, the left singular vectors of $\mathbf{R}$ corresponding to this singular value span a space of dimension one. Because of Lemma 2, the vector $\beta_0^{-1}\tilde{\mathbf{v}}_1$ is one such unit-norm left singular vector of $\mathbf{R}$. Since, by construction, $\hat{\mathbf{v}}_1$ is also a unit-norm left singular vector of $\mathbf{R}$ corresponding to this singular value, it follows that there exists a real θ such that

$$\hat{\mathbf{v}}_1 = \beta_0^{-1}e^{j\theta}\tilde{\mathbf{v}}_1.$$

Because of Step 3 and (30)

$$\hat{\mathbf{V}} = [\,\hat{\mathbf{v}}_1 \quad \cdots \quad \hat{\mathbf{v}}_d\,] = \beta_0^{-1}e^{j\theta}\tilde{\mathbf{V}}.$$

Thus, because of (10), (24), (27), and Step 4

$$\begin{aligned}\hat{\mathcal{H}} &= \beta_0\mathbf{U}_1\mathbf{\Sigma}(\beta_0^{-1}e^{j\theta}\tilde{\mathbf{V}}) \\ &= e^{j\theta}\mathbf{U}_1\mathbf{\Sigma}\mathbf{V}\mathcal{L}^{-1} \\ &= e^{j\theta}\mathbf{H}\mathcal{L}^{-1} \\ &= e^{j\theta}\mathcal{H}.\end{aligned}$$

Further, with $\mathbf{w}_n = 0$ in (2), from (9), (10), (24), (27), and Step 5, one obtains

$$\begin{aligned}\mathcal{G}_{\mathrm{ZF}}^H\mathbf{x}_n &= \mathcal{G}_{\mathrm{ZF}}^H\mathcal{H}\mathbf{s}_n \\ &= e^{-j\theta}\beta_0\mathcal{L}\mathcal{L}^H\frac{\tilde{\mathbf{V}}^H}{\beta_0}\mathbf{\Sigma}^{-1}\mathbf{U}_1^H\mathbf{U}_1\mathbf{\Sigma}\mathbf{V}\mathcal{L}^{-1}\mathbf{s}_n \\ &= e^{-j\theta}\mathcal{L}\mathcal{L}^H\mathcal{L}^{-H}\mathbf{V}^H\mathbf{V}\mathcal{L}^{-1}\mathbf{s}_n \\ &= e^{-j\theta}\mathbf{s}_n.\end{aligned}$$

This concludes the proof. ■

REFERENCES

[1] K. H. Afkhamie and Z.-Q. Luo, "Blind equalization of FIR systems driven by Markov-like input signals," *IEEE Trans. Signal Processing*, vol. 48, pp. 1726–1736, June 2000.

[2] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*. Englewood Cliffs, NJ: Prentice-Hall, 1979.

[3] E. de Carvalho and D. T. M. Slock, "Semi-blind methods for FIR multichannel estimation," in *Signal Processing Advances in Communications*, G. Giannakis, Y. Hua, P. Stoica, and L. Tong, Eds. Englewood Cliffs, NJ: Prentice-Hall, 2000, vol. I.

[4] Z. Ding, "Matrix outer-product decomposition method for blind multiple channel identification," *IEEE Trans. Signal Processing*, vol. 45, pp. 3053–3061, Dec. 1997.

[5] G. Giannakis and J. Mendel, "Identification of nonminimum phase systems using higher order statistics," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 360–377, Mar. 1989.

[6] D. Godard, "Self recovering equalization and carrier tracking in two-dimensional data communication systems," *IEEE Trans. Commun.*, vol. COMM-28, pp. 1867–1875, Nov. 1980.

[7] S. Haykin, *Adaptive Filter Theory*, 3rd ed. Upper Saddle River, NJ: Prentice-Hall, 1996.

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

[9] 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.

[10] T. Kailath, *Linear Systems*. Englewood Cliffs, NJ: Prentice-Hall, 1980.

[11] A. Liavas, P. A. Regalia, and J.-P. Delmas, "Blind channel approximation: Effective channel order determination," *IEEE Trans. Signal Processing*, vol. 47, pp. 3336–3344, Dec. 1999.

[12] J. Mannerkoski and V. Koivunen, "Autocorrelation properties of channel encoded sequences—Applicability to blind equalization," *IEEE Trans. Signal Processing*, vol. 48, pp. 3501–3506, Dec. 2000.

[13] 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.

[14] C. B. Papadias and D. T. M. Slock, "Fractionally spaced equalization of linear polyphase channels and related blind techniques based on multichannel linear prediction," *IEEE Trans. Signal Processing*, vol. 47, pp. 641–654, Mar. 1999.

[15] L. Tong and S. Perreau, "Multichannel blind identification: From subspace to maximum likelihood methods," *Proc. IEEE*, vol. 86, pp. 1951–1968, Oct. 1998.

[16] 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. 340–350, Mar. 1994.

[17] M. K. Tsatsanis and G. B. Giannakis, "Transmitter induced cyclostationarity for blind channel equalization," *IEEE Trans. Signal Processing*, vol. 45, pp. 1785–1794, July 1997.

[18] 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.

[19] Q. Zhao and L. Tong, "Adaptive blind channel estimation by least squares smoothing," *IEEE Trans. Signal Processing*, vol. 47, pp. 3000–3012, Nov. 1999.

Roberto López-Valcarce (M'01) was born in Spain in 1971. He received the Telecommunications Engineer degree from the University of Vigo, Vigo, Spain, in 1995. He received the M.S. and Ph.D. degrees in electrical engineering in 1998 and 2000 respectively, both from the University of Iowa, Iowa City, IA.

From 1995 to 1996, he was a Systems Engineer with Intelsis. He was the recipient of a Fundacin Pedro Barri de la Maza fellowship in 1996 for continuation of studies. Currently, he has been working as an associate researcher for the Communication Technologies Department, University of Vigo. His research interests are in adaptive signal processing and communications.

Soura Dasgupta (M'87–SM'93–F'98) was born in 1959 in Calcutta, India. He received the B.E. degree in electrical engineering from the University of Queensland, Brisbane, Australia, in 1980 and the Ph.D. in systems engineering from the Australian National University, Canberra, in 1985.

He is currently Dean's Diamond Professor of Electrical and Computer Engineering, University of Iowa, Iowa City. In 1981, he was a Junior Research Fellow with the Electronics and Communications Sciences Unit, Indian Statistical Institute, Calcutta. He has held visiting appointments at the University of Notre Dame, Notre Dame, IN, University of Iowa, Université Catholique de Louvain-La-Neuve, Belgium, and the Australian National University. His research interests are in controls, signal processing, and communications.

Dr. Dasgupta served as an Associate Editor of the IEEE TRANSACTIONS ON AUTOMATIC CONTROL between 1988 and 1991. He is a corecipient of the Gullimen Cauer Award for the best paper published in the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS in 1990 and 1991, a past Presidential Faculty Fellow, and an Associate Editor for the IEEE Control Systems Society Conference Editorial Board.