Blind Equalization of Nonlinear Channels From Second-Order Statistics

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

***Abstract*—This paper addresses the blind equalization problem for single-input multiple-output nonlinear channels, based on the second-order statistics (SOS) of the received signal. We consider the class of "linear in the parameters" channels, which can be seen as multiple-input systems in which the additional inputs are nonlinear functions of the signal of interest. These models include (but are not limited to) polynomial approximations of nonlinear systems. Although any SOS-based method can only identify the channel to within a mixing matrix (at best), sufficient conditions are given to ensure that the ambiguity is at a level that still allows for the computation of linear FIR equalizers from the received signal SOS, should such equalizers exist. These conditions involve only statistical characteristics of the input signal and the channel nonlinearities and can therefore be checked *a priori*. Based on these conditions, blind algorithms are developed for the computation of the linear equalizers. Simulation results show that these algorithms compare favorably with previous deterministic methods.**

***Index Terms*—Blind equalization, nonlinear channels, second-order statistics.**

I. Introduction

BLIND methods are of interest in digital communications as they permit channel identification/equalization at the receiver without training signals. The topic of blind equalization of single-input multiple-output (SIMO) linear channels in particular has received considerable attention (see [17] and the references therein), motivated by the facts that these channels can be perfectly equalized if the subchannels are coprime and the equalizer is long enough and that channel estimation/equalization can be performed from the second-order statistics (SOS) of the received signal. (SOS-based methods are often preferred to higher order statistics approaches since they usually require shorter data records).

By contrast, there is relatively little work in the area of blind equalization of nonlinear channels. (Some exceptions are [6], [15], and [19]). This problem, however, is of considerable practical interest since many communication systems, such as digital satellite and radio links [2], high-density magnetic [3], [7], and optical [1] storage channels exhibit significant nonlinearity. Further many physiological systems are also nonlinear [13]. In this paper, we consider nonlinear SIMO channels modeled as

$$y(n) = \sum_{i=1}^{q} \sum_{j=0}^{l_i} h_{ij} s_i(n-j) + z(n) \tag{1}$$

where $s_1(n) \triangleq a(n)$ is the scalar stationary channel input, the terms $s_i(n) = \phi_i(a(n), a(n-1), \ldots)$ for $i = 2, \ldots, q$ are scalar nonlinear causal functions of $a(\cdot)$, h_{ij} are $p \times 1$ coefficient vectors, and $z(n)$ and $y(n)$ are $p \times 1$ signal vectors representing an additive disturbance and the observed channel output, respectively; the number of subchannels is p. The noise $z(\cdot)$ and the signal $a(\cdot)$ are assumed to be independent. This model accommodates, for example, polynomial approximations of nonlinear channels (Volterra models), although the "basis functions" $s_i(\cdot)$ need not be monomials in principle. It is assumed that the functions $\phi_i(\cdot, \cdot, \ldots)$ generating the nonlinearities are known. This multichannel configuration can be obtained at the receiving end by using a sensor array, by fractional sampling the channel output [18], or by injecting cyclic redundancy at the channel input [16].

In this paper, we ask the following question. What information can be obtained about the channel from the SOS of its output? In the linear channel case, it is well known that under certain coprimeness conditions, the channel coefficients can be estimated from SOS up to a multiplicative constant [17]. We will show that this is not true in general for the nonlinear model (1) but that nevertheless, zero-forcing (ZF) linear equalizers for this class of channels can still be designed using only the SOS of $y(\cdot)$, provided certain conditions are satisfied. The fact that linear finite impulse response (FIR) systems can perform ZF equalization of nonlinear SIMO channels under certain conditions was first pointed out in [6], where a blind, deterministic approach for equalizer design was also presented. Although this algorithm is elegant and relatively simple, it has several drawbacks. First, it assumes that an associated channel matrix is full rank and square, claiming that squareness can always be achieved, if necessary, by decreasing the number of channels and increasing the equalizer length. However, a longer equalizer would increase the computational complexity; further, if some channels are to be dropped, it is not possible to ensure *a priori* that the surviving channels satisfy the corresponding full-rank condition, even if the original set did. Thus, the selection of the channels to drop is a difficult problem. The first issue motivating this paper is whether this squareness assumption can be relaxed.

Second, in the event the linear kernel has the same length as another kernel, [6] has to resort to higher order methods to equalize the channel. The theory developed in this paper shows

Manuscript received November 6, 2000; revised August 24, 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. Michail K. Tstasanis.

The authors are with the Department of Electrical and Computer Engineering, University of Iowa, Iowa City, IA 52242-1595 USA (e-mail: valcarce@engineering.uiowa.edu; dasgupta@engineering.uiowa.edu).

Publisher Item Identifier S 1053-587X(01)10514-3.

that this restriction on the linear kernel length is not necessary for SOS methods to apply.

Third, when a kernel other than the linear one has the largest memory, then the techniques of [6] only resolve this largest kernel. In our case, we show that under the right conditions, even the linear kernel is resolvable, despite the violation of this particular length condition.

Fourth, the algorithm from [6] only provides equalizers with zero and maximal delays, whereas in general, the greatest reduction in mean squared error is obtained for some intermediate delay. The algorithms presented here can be used to compute equalizers for all possible delay values.

Fifth, deterministic methods do not require nor exploit knowledge of the statistics of $a(\cdot)$. SOS-based methods exploiting such information when available can be expected to lead to improved performance. Simulation results given here verify this fact for the class of channels (1).

Recently, an SOS-based approach appeared in [15], inspired by the method of [20] for linear channels. However, this method requires that every nonlinear subchannel be linearizable by an FIR Volterra system of known order and memory, which is, in general, not possible, especially if each subchannel is modeled as an FIR Volterra system itself (a common practice).

In principle, (1) could be seen as a linear multiple-input multiple-output (MIMO) system by viewing the signals $s_i(\cdot)$, $2 \leq i \leq q$ as additional inputs. Although SOS-based techniques exist for equalization within such a framework [17], they usually assume that the different inputs are independent (which is no longer true in our setting) and only resolve the inputs to within a mixing matrix. In the current context, as the $s_i(\cdot)$ are functions of $a(\cdot)$, this would mean that only a linear mixture of $a(n)$, $s_2(n)$, $\ldots$, $s_q(n)$ could be obtained. The results of this work show that under right conditions, the structure of the mixing matrix permits obtaining linear ZF equalizers. These conditions are on the statistical properties of the symbols $a(\cdot)$ and the remaining basis functions $s_i(\cdot)$. Therefore, they can be checked *a priori* in order to determine whether a given channel structure can be equalized from SOS.

Two popular SOS-based methods for blind equalization of linear channels are i) the method of [18] and ii) subspace methods [14]. The latter do not exploit the explicit knowledge of input statistics and consequently exhibit inferior performance relative to methods that do exploit such knowledge [11]. Thus, here, we pursue extensions of [18] to the nonlinear channel case.

In our notation, $(\cdot)^*$, $(\cdot)^T$, $(\cdot)^H$ and $(\cdot)^\#$ denote conjugate, transpose, conjugate transpose, and pseudoinverse, respectively. Here

- J_k — $k \times k$ shift matrix with ones in the first subdiagonal and zeros elsewhere
- X_k — $k \times k$ exchange matrix with ones in the antidiagonal and zeros elsewhere
- e_k — vector of all zeros, except for a 1 in the kth position.

We use the direct sum notation $A \oplus B$ for block diagonal matrices:

$$A \oplus B \triangleq \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}.$$

The paper is organized as follows. Section II introduces the matrix-vector formulation of the setup and the problem statement and shows that SOS methods do not suffice to estimate the channel to within a scaling constant. The structure of the mixing matrix inherent to any blind SOS approach is discussed in Section III. Section IV presents a test that can be used to determine whether a given channel is equalizable from SOS. Sufficient conditions on the source statistics that ensure a positive answer from this test constitute our main results and are given in Section V. Algorithms and simulation results are given in Sections VI and VII, respectively.

II. Problem Formulation

The channel input–output relation can be expressed in matrix-vector form as follows. By collecting m successive observations in the vector

$$Y(n) \triangleq [\, y(n)^T \quad y(n-1)^T \quad \cdots \quad y(n-m+1)^T \,]^T$$

one can write

$$Y(n) = \mathcal{F}S(n) + Z(n) \tag{2}$$

where $Z(n)$ and $S(n)$ are the noise vector and the signal regressor, respectively, given by

$$Z(n) \triangleq [\, z(n)^T \quad z(n-1)^T \quad \cdots \quad z(n-m+1)^T \,]^T$$

and $S(n) \triangleq [S_1^T(n) \quad S_2^T(n)]^T$ with

$$S_1^T(n) \triangleq [a(n) \quad a(n-1) \quad \cdots \quad a(n-l_1-m+1)] \tag{3}$$

$$S_2^T(n) \triangleq [s_2(n) \quad \cdots \quad s_2(n-l_2-m+1) \quad | \quad \cdots \quad | \quad s_q(n) \quad \cdots \quad s_q(n-l_q-m+1)] \tag{4}$$

which represent the linear and nonlinear parts of $S(n)$. The channel matrix $\mathcal{F}$ is given by

$$\mathcal{F} \triangleq [\, \mathcal{F}_1 \quad \mathcal{F}_2 \quad \cdots \quad \mathcal{F}_q \,]$$

with every $\mathcal{F}_i$ block Toeplitz:

$$\mathcal{F}_i \triangleq \begin{bmatrix} h_{i0} & \cdots & h_{il_i} & & \\ & \ddots & & \ddots & \\ & & h_{i0} & \cdots & h_{il_i} \end{bmatrix} \quad pm \times (m+l_i).$$

$\mathcal{F}$ has size $pm \times (qm + \sum_{i=1}^{q} l_i)$. For convenience, let $d_1 \triangleq m + l_1$, which is the size of $S_1(n)$, and $d_2 \triangleq (q-1)m + l_2 + \cdots + l_q$, which is the size of $S_2(n)$. Thus, $S(n)$ has size $d_1 + d_2$.

Since the channel nonlinearities may induce nonzero mean $s_i(\cdot)$ terms even if $a(\cdot)$ is zero mean, we will consider covariance matrices rather than autocorrelation matrices. The covariance sequence of the process $Y(\cdot)$ can be written as

$$C_y(k) \triangleq \text{cov}[Y(n),\, Y(n-k)] = \mathcal{F}C_s(k)\mathcal{F}^H + C_z(k) \tag{5}$$

where $C_s(k)$, $C_z(k)$ are the source and noise covariance sequences given by

$$C_s(k) \triangleq \text{cov}[S(n),\, S(n-k)]$$

$$C_z(k) \triangleq \text{cov}[Z(n),\, Z(n-k)]. \tag{6}$$

The following standard assumptions are adopted throughout the paper:

A1) The channel matrix $\mathcal{F}$ is tall and has full column rank.
A2) $z(\cdot)$ is zero-mean white with covariance $\sigma_z^2 I_p$.
A3) The covariance matrix $C_s(0)$ is positive definite.

Assumption **A1** is a "coprimeness" requirement on the subchannels. It ensures the existence of vectors g_δ such that $g_\delta^H \mathcal{F} = e_\delta^H$. Thus, in the noiseless case, for $1 \le \delta \le d_1$, one has $g_\delta^H Y(n) = a(n-\delta+1)$ so that these $pm \times 1$ vectors provide ZF linear equalizers with associated delay $\delta - 1$.

Under **A2**, one has $C_z(k) = \sigma_z^2 J_{pm}^{pk}$. Then, because of **A1**, σ_z^2 can be estimated as the smallest eigenvalue of $C_y(0)$ so that the effect of noise can be removed from $C_y(k)$. Henceforth, we will assume that $C_y(k) = \mathcal{F} C_s(k) \mathcal{F}^H$.

Assumption **A3** constitutes a "persistent excitation" condition on $a(\cdot)$. It allows one to write

$$C_s(0) = QQ^H \qquad \text{with } Q \text{ invertible.} \tag{7}$$

The problem under consideration can be posed as follows.

Blind Equalizability Problem: Let $\tilde{\mathcal{F}}$ be a matrix of the same size as $\mathcal{F}$ such that

$$\tilde{\mathcal{F}} C_s(k) \tilde{\mathcal{F}}^H = \mathcal{F} C_s(k) \mathcal{F}^H, \qquad k = 0, 1, \ldots \overline{k}. \tag{8}$$

We say that $\tilde{\mathcal{F}}$ is *compatible* with the SOS of $Y(\cdot)$ (up to lag $\overline{k}$). Determine conditions under which a ZF equalizer g_δ for any compatible $\tilde{\mathcal{F}}$ is also a ZF equalizer for $\mathcal{F}$

$$g_\delta^H \tilde{\mathcal{F}} = e_\delta^H \qquad \Longrightarrow \qquad g_\delta^H \mathcal{F} = c e_\delta^H, \quad \text{for some } 1 \le \delta \le d_1 \text{ and } c \ne 0. \tag{9}$$

This was solved in [18] for the particular case of linear channels with white inputs, for which if $\tilde{\mathcal{F}}$ is compatible up to lag $\overline{k} = 1$, then $\tilde{\mathcal{F}} = e^{j\theta} \mathcal{F}$ so that (9) holds. This is not necessarily true in our case. For example, suppose that the different terms $s_i(\cdot)$ are uncorrelated: $\mathrm{cov}[s_i(n), s_j(m)] = 0$, $i \ne j$ (e.g., a linear-quadratic channel with iid input). Then, $C_s(k)$ is block diagonal for all k so that any matrix of the form $\tilde{\mathcal{F}} = [e^{j\theta_1}\mathcal{F}_1 \;\; e^{j\theta_2}\mathcal{F}_2 \;\; \cdots \;\; e^{j\theta_q}\mathcal{F}_q]$ is compatible up to any lag. Thus, *in general, $\mathcal{F}$ cannot be identified to within a single scaling constant*. However, this may not be necessary in order for (9) to hold, i.e., a higher level of indeterminacy in the channel parameters could still allow equalization. We explore this issue in the next section.

III. Mixing Matrix

Let Q be a square root of $C_s(0)$ as in (7). Define the *normalized* channel and source covariance matrices, respectively, as

$$F = \mathcal{F} Q, \qquad \overline{C}_s(k) = Q^{-1} C_s(k) Q^{-H}. \tag{10}$$

F has full column rank in view of **A1** and **A3**. Using (10), the matrices $C_y(k)$ become

$$C_y(k) = F \overline{C}_s(k) F^H, \qquad \text{with } \overline{C}_s(0) = I_{d_1+d_2}. \tag{11}$$

Similarly, if $\tilde{\mathcal{F}}$ is compatible up to lag $\overline{k}$, let $\tilde{\mathcal{F}} = \tilde{\mathcal{F}} Q$ so that $\tilde{\mathcal{F}}$ satisfies

$$\tilde{\mathcal{F}} \overline{C}_s(k) \tilde{\mathcal{F}}^H = F \overline{C}_s(k) F^H, \qquad 0 \le k \le \overline{k}. \tag{12}$$

For $k = 0$, (12) gives $\tilde{\mathcal{F}} \tilde{\mathcal{F}}^H = FF^H$. Since F has full column rank, this implies $\tilde{\mathcal{F}} = FP$ for some unitary P. Thus, the corresponding (unnormalized) compatible channel matrix must satisfy

$$\tilde{\mathcal{F}} = \tilde{\mathcal{F}} Q^{-1} = \mathcal{F}(QPQ^{-1}) \tag{13}$$

which shows that any compatible channel matrix is related to the true channel via a mixing matrix of the form $\tilde{P} = QPQ^{-1}$. Observe that although P is unitary, in general, $\tilde{P}$ is not. Let us introduce now the concept of admissibility.

Definition 1 (Admissibility): A $(d_1 + d_2)$-square matrix T is said to be *admissible* if

$$T = \begin{bmatrix} \Lambda & 0 \\ \times & \times \end{bmatrix}, \quad \Lambda,\ d_1 \times d_1 \text{ diagonal invertible} \tag{14}$$

with "$\times$" indicating irrelevant values. Note that if T is admissible and invertible, so is T^{-1}, and that any function of an admissible matrix is admissible.

Observe that if $\tilde{\mathcal{F}} = \mathcal{F}\tilde{P}$ is compatible with $\tilde{P} = QPQ^{-1}$ admissible, then (9) is satisfied. Thus, resolution of the channel matrix to within this ambiguity suffices for equalization purposes.

Our goal is to determine conditions under which the mixing matrix $\tilde{P}$ is ensured to be admissible. To address this issue, we must explore the constraints that the conditions (12) impose on the unitary matrix P. Substituting $\tilde{\mathcal{F}} = FP$ into (12) and using the fact that F has full rank, these constraints can be written as

$$P\overline{C}_s(k) = \overline{C}_s(k) P, \qquad 1 \le k \le \overline{k} \tag{15}$$

i.e., P must commute with the normalized source covariance matrices $\overline{C}_s(1), \ldots, \overline{C}_s(\overline{k})$.

IV. Simplified Equalizability Test

Determining the general form of all unitary matrices P that satisfy (15) requires solving $\overline{k}$ linear sets of equations with quadratic constraints. Fortunately, this problem can be replaced by one of solving $\overline{k}$ linear sets of equations with *linear* constraints. First, recall that any unitary matrix P can be written as $P = e^{jW}$, where W is a Hermitian matrix with eigenvalues in $[0, 2\pi)$ [9]. Second, we have the following result from [10].

Theorem 1: Let W be $(d_1 + d_2)$-square Hermitian and $P = e^{jW}$. Then, P and $\overline{C}_s(k)$ commute if and only if W and $\overline{C}_s(k)$ commute.

Therefore, the problem of determining the set of unitary matrices that commute with $\overline{C}_s(k)$ is equivalent to finding the set of Hermitian matrices that commute with $\overline{C}_s(k)$. Hence, the blind equalizability problem can be broken into these three steps.

1) Select a square root Q of $C_s(0)$.
2) Find all Hermitian matrices W commuting with $\overline{C}_s(k) = Q^{-1} C_s(k) Q^{-H}$ for $1 \le k \le \overline{k}$.
3) Check whether for these matrices W, QWQ^{-1} is admissible. If so, the channel can be equalized using SOS, as QWQ^{-1} admissible implies $Qe^{jW}Q^{-1} = QPQ^{-1}$ admissible.

The utility of Theorem 1 is revealed in that steps 2) and 3) are much easier to solve for Hermitian matrices than for unitary matrices.

We proceed now to determine sufficient conditions on the source statistics and the channel nonlinearities in order to ensure success of the SOS-based equalizability test *a priori*.

V. Main Results

It is useful to consider block lower triangular square roots Q [with block partition corresponding to linear and nonlinear parts of $S(n)$, as in (14)], for the following reason. Suppose that any Hermitian W solving step 2) of our test is block diagonal. Then, $P = e^{jW}$ is block diagonal; therefore, if Q was block lower triangular, then so is $\tilde{P} = QPQ^{-1}$. Having $\tilde{P}$ block lower triangular [i.e., as in (14) but with Λ not necessarily diagonal] is the first step toward admissibility. Its significance is that a linear ZF equalizer g_δ for the compatible matrix $\tilde{\mathcal{F}}$ ($g_\delta^H \tilde{\mathcal{F}} = e_\delta^H$, $1 \le \delta \le d_1$), although not a ZF equalizer for the true channel $\mathcal{F}$, still removes all the nonlinear intersymbol interference (ISI) if $\tilde{P}$ is block lower triangular since $g_\delta^H \mathcal{F} = e_\delta^H \tilde{P}^{-1}$. Once this has been achieved, additional conditions for the removal of the residual linear ISI can be sought.

A. Preliminary Result

The following result ensures the block diagonal property of any Hermitian W commuting with $\overline{C}_s(1)$. The proof is given in Appendix A.

Theorem 2: Assume that there exists a matrix Q such that $C_s(0) = QQ^H$ and

$$\overline{C}_s(1) = Q^{-1} C_s(1) Q^{-H} = \begin{bmatrix} C_{11} & 0 \\ C_{21} & C_{22} \end{bmatrix} \tag{16}$$

with C_{ij} having size $d_i \times d_j$, $i, j \in \{1, 2\}$. Suppose that either i) C_{11}, C_{22} do not share any eigenvalues; or ii) $C_{21} = 0$, and C_{11}, C_{22} do not share any elementary Jordan block in their Jordan decompositions, i.e., Jordan blocks belonging to the same eigenvalues have different sizes. Then, any Hermitian W commuting with $\overline{C}_s(1)$ must be of the form in (17) with W_{ii} $d_i \times d_i$:

$$W = W_{11} \oplus W_{22}. \tag{17}$$

B. Useful Square Root Q

With the result from Theorem 2 in mind, we will focus on block triangular square roots Q and look for conditions under which (16) is satisfied. To this end, let us define the $d_i \times d_j$ matrices

$$\begin{aligned} A_{ij} &\triangleq \text{cov}[S_i(n),\, S_j(n)] \\ B_{ij} &\triangleq \text{cov}[S_i(n),\, S_j(n-1)], \qquad i,\, j \in \{1,\, 2\} \end{aligned} \tag{18}$$

so that the covariance matrices $C_s(0)$, $C_s(1)$ can be written as

$$C_s(0) = \begin{bmatrix} A_{11} & A_{12} \\ A_{12}^H & A_{22} \end{bmatrix}, \qquad C_s(1) = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}$$

with

$$\begin{aligned} w_{1i} &\triangleq \text{cov}[S_i(n-1),\, a(n)] \qquad \text{for } i \in \{1,\, 2\} \\ w_{21} &\triangleq \text{cov}[S_2(n),\, a(n-d_1)] \end{aligned} \tag{19}$$

and the shift structure of $S_1(n)$ yields the following relations:

$$B_{11} = J_{d_1} A_{11} + e_1 w_{11}^H \tag{20}$$

$$= A_{11} J_{d_1} + X_{d_1} w_{11}^* e_{d_1}^H \tag{21}$$

$$B_{12} = J_{d_1} A_{12} + e_1 w_{12}^H \tag{22}$$

$$B_{21} = A_{12}^H J_{d_1} + w_{21} e_{d_1}^H. \tag{23}$$

Define $A_0 \triangleq A_{22} - A_{12}^H A_{11}^{-1} A_{12}$ [the Schur complement of $C_s(0)$ with respect to A_{11}^{-1}], which is positive definite. The following choice of Q will prove particularly useful:

$$Q = \begin{bmatrix} A_{11}^{1/2} & 0 \\ A_{12}^H A_{11}^{-H/2} & A_0^{1/2} \end{bmatrix} \tag{24}$$

with $A_{11}^{1/2}$, $A_0^{1/2}$ square roots of A_{11}, A_0, respectively. Using (24), $\overline{C}_s(1) = Q^{-1} C_s(1) Q^{-H}$ becomes (25), shown at the bottom of the page, where $B_0 \triangleq B_{22} - A_{12}^H A_{11}^{-1} B_{11} A_{11}^{-1} A_{12}$.

C. Block Triangular $\overline{C}_s(1)$

By using (20)–(23), the off-diagonal terms of the middle matrix in (25) can be rewritten as

$$B_{12} - B_{11} A_{11}^{-1} A_{12} = e_1 (w_{12} - A_{12}^H A_{11}^{-1} w_{11})^H \tag{26}$$

$$B_{21} - A_{12}^H A_{11}^{-1} B_{11} = \left(w_{21} - A_{12}^H A_{11}^{-1} X_{d_1} w_{11}^*\right) e_{d_1}^H. \tag{27}$$

Introduce the vectors

$$\alpha \triangleq A_{11}^{-1} w_{11}, \qquad \beta \triangleq A_{11}^{-1} X_{d_1} w_{11}^* \tag{28}$$

which comprise the coefficients of the optimum forward prediction error filter (FPEF) and optimum backward prediction error filter (BPEF), respectively, of order d_1 associated with the process $a(\cdot)$ [12]. These prediction errors are given by

$$\begin{aligned} f(n) &= a(n) - \alpha^H S_1(n-1) \\ b(n) &= a(n-d_1) - \beta^H S_1(n). \end{aligned} \tag{29}$$

One can readily check that

$$\text{cov}[S_2(n-1),\, f(n)] = w_{12} - A_{12}^H \alpha = w_{12} - A_{12}^H A_{11}^{-1} w_{11}$$

$$\text{cov}[S_2(n),\, b(n)] = w_{21} - A_{12}^H \beta = w_{21} - A_{12}^H A_{11}^{-1} X_{d_1} w_{11}^*.$$

$$\overline{C}_s(1) = (A_{11}^{-1/2} \oplus A_0^{-1/2}) \begin{bmatrix} B_{11} & B_{12} - B_{11} A_{11}^{-1} A_{12} \\ B_{21} - A_{12}^H A_{11}^{-1} B_{11} & B_0 - A_{12}^H A_{11}^{-1} B_{12} - B_{21} A_{11}^{-1} A_{12} \end{bmatrix} (A_{11}^{-H/2} \oplus A_0^{-H/2}) \tag{25}$$

Substituting these into (26) and (27), one obtains

$$B_{12} - B_{11}A_{11}^{-1}A_{12} = e_1\text{cov}[S_2(n-1),\, f(n)]^H \tag{30}$$
$$B_{21} - A_{12}^H A_{11}^{-1}B_{11} = \text{cov}[S_2(n),\, b(n)]e_{d_1}^H. \tag{31}$$

Thus, for Q as in (24), $\overline{C}_s(1)$ is block lower triangular iff cov$[S_2(n-1),\, f(n)] = 0$; in that case, we have (32), shown at the bottom of the page, where $v \triangleq$ cov$[S_2(n),\, b(n)]$. We can now state the following result, proved in Appendix B.

Theorem 3: Suppose that the symbol sequence $a(\cdot)$ is an autoregressive (AR) process of order not exceeding d_1 with independent, identically distributed (iid) innovations, i.e., it is generated by means of stable all-pole filtering of an iid process $w(\cdot)$ as follows:

$$a(n) = w(n) - \sum_{k=1}^{d_1} \gamma_k a(n-k). \tag{33}$$

Assume that **A1**–**A3** hold, that

$$\Upsilon \triangleq \text{cov}\left(\begin{bmatrix} S_1(n) \\ a(n-l_1-m)\end{bmatrix},\quad \begin{bmatrix} S_1(n) \\ a(n-l_1-m)\end{bmatrix}\right) > 0 \tag{34}$$

and that the matrices $J_{d_1} + e_1\alpha^H$ and $A_0^{-1}(B_0 - ve_{d_1}^H A_{11}^{-1}A_{12})$ do not have any common eigenvalues. Then, $\overline{C}_s(1)$ is block lower triangular, and any matrix compatible with the SOS of the channel output up to the lag $\overline{k} = 1$ is related to the true channel matrix via an admissible matrix.

The following remarks are in order.

1) The AR condition on the symbols $a(\cdot)$ is sufficient for having $\overline{C}_s(1)$ block lower triangular, but it is by no means necessary. If even, for non-AR symbols, cov$[S_2(n-1),\, f(n)] = 0$ holds, one can use Theorem 2 in order to reduce the blind equalizability problem to an eigenvalue check on the matrices $J_{d_1} + e_1\alpha^H$ and $A_0^{-1}(B_0 - ve_{d_1}^H A_{11}^{-1}A_{12})$.

2) The matrix $A_0^{-1}(B_0 - ve_{d_1}^H A_{11}^{-1}A_{12})$ has a linear prediction interpretation. Let $\tilde{S}_2(n)$ be the prediction error obtained by approximating $S_2(n)$ by a linear function of $S_1(n)$:

$$\tilde{S}_2(n) = S_2(n) - \Gamma^H S_1(n).$$

The value of Γ that minimizes trace$\{$cov$[\tilde{S}_2(n),\, \tilde{S}_2(n)]\}$ is $\Gamma = A_{11}^{-1}A_{12}$, for which one has cov$[\tilde{S}_2(n),\, \tilde{S}_2(n)] = A_0$, and if cov$[S_2(n-1),\, f(n)] = 0$ holds, then cov$[\tilde{S}_2(n),\, \tilde{S}_2(n-1)] = B_0 - ve_{d_1}^H A_{11}^{-1}A_{12}$. Therefore,

$$A_0^{-1}\left(B_0 - ve_{d_1}^H A_{11}^{-1}A_{12}\right) = \text{cov}[\tilde{S}_2(n),\, \tilde{S}_2(n)]^{-1}\text{cov}[\tilde{S}_2(n),\, \tilde{S}_2(n-1)].$$

The problem of isolating conditions under which this matrix and $J_{d_1} + e_1\alpha^H$ do not have common eigenvalues remains open.

3) Theorem 3 covers the important case of iid symbols $a(\cdot)$ (by having $\gamma_1 = \cdots = \gamma_{d_1} = 0$) for which $\alpha = \beta = 0$, $f(n) = a(n)$, $b(n) = a(n-d_1)$, and $A_{11} = \sigma_a^2 I_{d_1}$. More will be said about the iid input case in Theorem 4 and Section V-E.

D. Block Diagonal $\overline{C}_s(1)$

It is clear from (30) and (31) that for our choice of Q, $\overline{C}_s(1)$ is block diagonal iff cov$[S_2(n-1),\, f(n)] = 0$ *and* $v =$ cov$[S_2(n),\, b(n)] = 0$, in which case, the resulting value is

$$\overline{C}_s(1) = \left[A_{11}^{-1/2}\left(J_{d_1} + e_1\alpha^H\right)A_{11}^{1/2}\right] \oplus \left[A_0^{-1/2}B_0A_0^{-H/2}\right]. \tag{35}$$

The theorems below provide examples in which (35) holds. The first one makes the following additional assumptions:

A4) The transmitted symbol sequence $a(\cdot)$ is iid with $E[|a(n)|^2] = \sigma_a^2$.

A5) cov$[S_2(n),\, a(n-d_1)] = 0$.

Observe that under assumptions **A1**–**A5** and taking Q as in (24) with $A_{11}^{1/2} = \sigma_a I_{d_1}$, the matrix $\overline{C}_s(1) = Q^{-1}C_s(1)Q^{-H}$ is block diagonal as in (36), with C $d_2 \times d_2$:

$$\overline{C}_s(1) = J_{d_1} \oplus C. \tag{36}$$

Theorem 4: Under assumptions **A1–A5**, if the Jordan decomposition of C in (36) has no Jordan block of size d_1 associated with the zero eigenvalue, then any matrix compatible with the channel output SOS up to lag 1 is related to the true channel matrix by an admissible matrix.

Proof: Since $\overline{C}_s(1)$ is as in (36), by Theorem 2, any Hermitian W commuting with $\overline{C}_s(1)$ is as in (17). As the characteristic polynomial of J_{d_1} is just λ^{d_1}, which is minimum phase, the result follows in a manner similar to the last part of the proof of Theorem 3. ■

A sufficient, although not necessary, condition for **A5** to hold is that **A4** hold and that the linear kernel have the largest memory. An example where this is not necessary is when $s_i(n)$ for $i > 1$ in (1) are degree-two monomials in $a(\cdot)$, and $a(\cdot)$ is iid real, symmetrically distributed around zero. Then, $C = J_{m+l_2} \oplus \cdots \oplus J_{m+l_q}$, and blind equalizability is obtained, as long as $l_i \neq l_1$ for $i \neq 1$.

Theorem 5: Suppose assumptions **A1**–**A3** hold, the symbols $a(\cdot)$ are Gaussian, and the memory of the nonlinear part of the channel does not exceed that of the linear part, i.e., $S_2(n) = \phi(S_1(n))$, where $\phi(\cdot)$ is a memoryless mapping. In addition, assume that the Jordan decompositions of the matrices $J_{d_1} + e_1\alpha^H$ and $A_0^{-1}B_0$ do not have any common elementary Jordan block and that Υ, which is defined in (34), is positive definite. Then, any matrix compatible with the SOS of the channel output

$$\overline{C}_s(1) = \begin{bmatrix} A_{11}^{-1/2}\left(J_{d_1} + e_1\alpha^H\right)A_{11}^{1/2} & 0 \\ A_0^{-1/2}ve_{d_1}^H A_{11}^{-H/2} & A_0^{-1/2}\left(B_0 - ve_{d_1}^H A_{11}^{-1}A_{12}\right)A_0^{-H/2}\end{bmatrix} \tag{32}$$

up to the lag $\overline{k} = 1$ is related to the true channel matrix via an admissible matrix.

Proof: By the orthogonality principle, $\text{cov}[S_1(n-1), f(n)] = 0$ and $\text{cov}[S_1(n), b(n)] = 0$ so that $(S_1(n-1), f(n))$ are uncorrelated random variables, and so are $(S_1(n), b(n))$. Since $a(\cdot)$ is Gaussian, so are $S_1(\cdot)$, $f(\cdot)$, and $b(\cdot)$. Therefore, the random variables $(\phi(S_1(n-1)), f(n))$ and $(\phi(S_1(n)), b(n))$ are independent so that $\text{cov}[S_2(n-1), f(n)] = \text{cov}[S_2(n), b(n)] = 0$, i.e., $v = 0$. Thus, with Q as in (24), $\overline{C}_s(1)$ is block diagonal as in (35) with (2, 2) block $A_0^{-1/2} B_0 A_0^{1/2}$, which is similar to $A_0^{-1} B_0$. Since the Jordan forms of $J_{d_1} + e_0 \alpha^H$ and $A_0^{-1} B_0$ do not share any elementary Jordan blocks, by Theorem 2, any Hermitian W commuting with $\overline{C}_s(1)$ is block diagonal. Admissibility of QWQ^{-1} follows, as in the last part of the proof of Theorem 3. ■

Remarks:

1) Gaussianity of the symbols $a(\cdot)$ is sufficient for having $\overline{C}_s(1)$ block diagonal, but it is not necessary. In general, as long as $\text{cov}[S_2(n-1), f(n)] = \text{cov}[S_2(n), b(n)] = 0$ is satisfied, Theorem 2 allows one to reduce the blind equalizability problem to a check on the Jordan structures of $J_{d_1} + e_1 \alpha^H$ and $A_0^{-1} B_0$. This is the case if, for example, the symbols and the nonlinear terms are uncorrelated, i.e., if $\text{cov}[a(n), s_i(n-k)] = 0$ for all k and for $i = 2, \ldots, q$.
2) If the conditions of Theorem 5 are satisfied, the color of the symbols $a(\cdot)$ does not affect blind equalizability. It is conceivable, however, that for Gaussian symbols, a particular choice of the symbol autocorrelation could result in the two diagonal blocks of $\overline{C}_s(1)$ sharing a Jordan block.
3) An example of a communications system in which the symbols are approximately Gaussian is the orthogonal frequency division multiplexing (OFDM) scheme [4], in which the original symbol stream is divided in blocks that undergo an inverse discrete Fourier transform (IDFT) operation prior to transmission. If the original symbols were zero-mean i.i.d., then the IDFT output is asymptotically (i.e., as the block size tends to infinity) white Gaussian [5]. Constellation shaping is another technique that produces a Gaussian-like symbol sequence [21].

E. Result for iid Symbols

For iid $a(\cdot)$, we present an equalizability result that bypasses the need for a check on the eigenstructure of the covariance matrices at the price of imposing more stringent conditions on the relative memory lengths of the channel.

Theorem 6: Under Assumptions **A1–A3**, suppose that the symbols $a(\cdot)$ are iid and that the memory of the linear part of the channel is *strictly greater* than that of the nonlinear part, i.e.,

$$S_2(n) = \phi(a(n), \ldots, a(n-d_1+2)) \quad \text{with } \phi(\cdot, \ldots, \cdot) \text{ a memoryless mapping.} \tag{37}$$

Then, the ZF equalizers of delays 0 and $d_1 - 1$ obtained for any channel matrix compatible with the SOS of the received signal up to lag $\overline{k} = d_1 - 1$ are also ZF equalizers for the true channel. That is, (9) holds with $\delta = 1$ and $\delta = d_1$.

Proof: In view of the equalizability test of Section IV, we will look at the structure of any Hermitian W commuting with $\overline{C}_s(d_1 - 1)$. The iid assumption and stationarity yield

$$\text{cov}[S_1(n), S_1(n-d_1+1)] = \sigma_a^2 e_{d_1} e_1^H \tag{38}$$

$$\text{cov}[S_1(n), S_2(n-d_1+1)] = e_{d_1} e_1^H A_{12} \tag{39}$$

$$\text{cov}[S_2(n), S_1(n-d_1+1)] = 0 \tag{40}$$

$$\text{cov}[S_2(n), S_2(n-d_1+1)] = 0 \tag{41}$$

with A_{12} defined in (18), and $\sigma_a^2 \triangleq E[|a(n)|^2]$. Therefore

$$C_s(d_1 - 1) = \begin{bmatrix} \sigma_a^2 e_{d_1} e_1^H & e_{d_1} e_1^H A_{12} \\ 0 & 0 \end{bmatrix}. \tag{42}$$

Take Q as in (24) with $A_{11}^{1/2} = \sigma_a I_{d_1}$. Noting that $e_{d_1}^H A_{12} = \text{cov}[a(n-d_1+1), S_2(n)] = 0$ due to the iid and memory assumptions, it follows from (42) that $\overline{C}_s(d_1-1) = Q^{-1} C_s(d_1-1) Q^{-H}$ has only one nonzero element

$$\overline{C}_s(d_1 - 1) = (e_{d_1} e_1^H) \oplus 0_{d_2 \times d_2}. \tag{43}$$

Now, let $W = [W_{ij}]_{i,j=1,2}$ be a Hermitian matrix commuting with $\overline{C}_s(d_1 - 1)$, with each W_{ij} $d_i \times d_j$. From (43), we must then have

$$W_{11} e_{d_1} e_1^H = e_{d_1} e_1^H W_{11}, \quad e_{d_1} e_1^H W_{12} = 0, \quad W_{12}^H e_{d_1} e_1^H = 0.$$

These imply that with θ a scalar

$$W_{11} = \theta \oplus \overline{W}_{11} \oplus \theta, \qquad W_{12}^H = \begin{bmatrix} 0 & \overline{W}_{12}^H & 0 \end{bmatrix} \tag{44}$$

where $\overline{W}_{11}$ is $(d_1-2) \times (d_1-2)$ Hermitian, and $\overline{W}_{12}$ is $(d_1 - 2) \times d_2$. Then, one can check that

$$QWQ^{-1} = \begin{bmatrix} W_{11} - \dfrac{1}{\sigma_a} W_{12} A_0^{-1/2} A_{12}^H & \sigma_a W_{12} A_0^{-1/2} \\ \times & \times \end{bmatrix}. \tag{45}$$

Therefore, in view of (44), the first and d_1th rows of QWQ^{-1} are just θe_1^H and $\theta e_{d_1}^H$, respectively. Thus, the first and d_1th rows of $\mathrm{e}^{-jQWQ^{-1}}$ are $\mathrm{e}^{-j\theta} e_1^H$ and $\mathrm{e}^{-j\theta} e_{d_1}^H$. Now, recall from (13) that any $\tilde{\mathcal{F}}$ that is compatible must satisfy $\mathcal{F} = \tilde{\mathcal{F}} \mathrm{e}^{-jQWQ^{-1}}$. Thus, for $\delta = 1$ and $\delta = d_1$, (9) holds. ■

In this section, we have presented conditions on the symbol statistics and channel nonlinearities that suffice for blind equalizability. In particular, under the specified conditions on the input statistics, every channel estimate that is compatible with the SOS must share all its equalizers with the "true" channel. Thus, the intrinsic ambiguity in the channel estimation procedure poses no practical difficulty as any equalizer obtained on the basis of the channel estimate necessarily equalizes the true channel. Nonetheless, we have yet to describe how the equalizers can be found from the channel output SOS. We consider this issue in the next section.

VI. Algorithm Development

The previous sections demonstrated that blind equalizers can be found from the channel output SOS under a variety of settings. In this section, we provide algorithms for determining these equalizers for the important iid input cases covered by Theorems 4 and 6.

A. Setting of Theorem 4

Suppose the conditions of Theorem 4 hold. As in [18], we start with a singular value decomposition (SVD) of $C_y(0)$

$$C_y(0) = [\,U_1 \quad U_2\,] \begin{bmatrix} \Sigma^2 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} U_1^H \\ U_2^H \end{bmatrix} \tag{46}$$

where $\Sigma^2 = \operatorname{diag}(\sigma_1^2, \ldots, \sigma_{d_1+d_2}^2)$, and U_1 has $d_1 + d_2$ columns. Recall that with F, which is the normalized channel matrix defined in (10), one has $C_y(0) = FF^H$. Since F has full column rank, it follows that $F = U_1 \Sigma V$ for some unitary $(d_1 + d_2) \times (d_1 + d_2)$ matrix V. Let us partition V as $V = [V_1 \;\; V_2]$, where each V_i has d_i columns $i = 1, 2$.

Observe that the first d_1 rows of the pseudoinverse $\mathcal{F}^\# = QF^\#$ provide the ZF equalizers of delays 0 through $d_1 - 1$. Using the facts that $F = U_1 \Sigma V$ and that $[I_{d_1} \;\; 0]Q = \sigma_a [I_{d_1} \;\; 0]$, it can be checked that the ZF equalizers are

$$[\,I_{d_1} \quad 0\,]\,\mathcal{F}^\# = \sigma_a V_1^H \Sigma^{-1} U_1^H. \tag{47}$$

Therefore, for equalization purposes, it suffices to estimate V_1; the remaining columns of V are of no interest to us. Now, it can be checked that

$$R \triangleq \Sigma^{-1} U_1^H C_y(1) U_1 \Sigma^{-1} = V \overline{C}_s(1) V^H. \tag{48}$$

Our goal is to estimate V_1 from R using the structure of $\overline{C}_s(1)$ shown in (36). Note that C in (36) is known to us and that (36) and (48) imply

$$RV_1 = V_1 J_{d_1}, \qquad R^H V_1 = V_1 J_{d_1}^H. \tag{49}$$

These "Jordan chains" show that V_1 can be estimated from either its first or its last column. If we partition V_1 columnwise as $V_1 = [v_{1,1} \;\; \cdots \;\; v_{1,d_1}]$, (49) reads as

$$Rv_{1,d_1} = 0, \qquad Rv_{1,i-1} = v_{1,i} \quad i = d_1, \ldots, 3, 2 \tag{50}$$

$$R^H v_{1,1} = 0, \qquad R^H v_{1,j} = v_{1,j-1} \quad j = 2, 3, \ldots, d_1. \tag{51}$$

Thus, it suffices to obtain an estimate of $v_{1,1}$ or v_{1,d_1}. In the original algorithm of [18] for linear channels, one had $\overline{C}_s(1) = J_{d_1}$, i.e., the block C in (36) was absent. This allowed for $(v_{1,1}, v_{1,d_1})$ to be taken as a "singular pair" associated with the smallest singular value of the matrix R. This would work in our case, provided that C in (36) is nonsingular. However, we would like to allow for singular C as well since this is usually the case in practice.

To do so, we extract the information about V_1 in R through several of its powers. Note that

$$R^k = V_1 J_{d_1}^k V_1^H + V_2 C^k V_2^H. \tag{52}$$

For $k = d_1 - 1$, one has $J_{d_1}^{d_1-1} = e_{d_1} e_1^H$, whereas $J_{d_1}^k = 0$ for $k \geq d_1$. Suppose for the moment that $V_2 C^{d_1-1} V_2^H$ has been found. Then, $R^{d_1-1} - V_2 C^{d_1-1} V_2^H = V_1 e_{d_1} e_1^H V_1^H = v_{1,d_1} v_{1,1}^H$. Hence, $(v_{1,1}, v_{1,d_1})$ could then be estimated, up to a constant of the form $e^{j\theta}$, as the singular pair associated with the largest singular value of $R^{d_1-1} - V_2 C^{d_1-1} V_2^H$. Thus, we must determine $V_2 C^{d_1-1} V_2^H$. Observe that C can be written as

$$C = T(\hat{Y} \oplus Y \oplus Z)T^{-1} \tag{53}$$

where

$\hat{Y}$	direct sum of shift matrices J_i with $i < d_1$;
Y	direct sum of matrices J_i with $i > d_1$;
Z	square nonsingular.

Partitioning

$$T = [\,\hat{T}_1 \quad T_1 \quad T_2\,],$$
$$T^{-H} = \left[\, \left(\hat{T}_1^\#\right)^H \quad \left(T_1^\#\right)^H \quad \left(T_2^\#\right)^H \,\right]$$

$\hat{T}_1$, T_1, and T_2 having the same number of columns as $\hat{Y}$, Y and Z, respectively, one can express

$$V_2 C V_2^H = \hat{\Gamma} + \Gamma + \Delta \tag{54}$$

where

$$\hat{\Gamma} \triangleq V_2 \hat{T}_1 \hat{Y} \hat{T}_1^\# V_2^H, \qquad \Gamma \triangleq V_2 T_1 Y T_1^\# V_2^H$$
$$\Delta \triangleq V_2 T_2 Z T_2^\# V_2^H. \tag{55}$$

Moreover, one has $V_2 C^k V_2^H = \hat{\Gamma}^k + \Gamma^k + \Delta^k$. Since $\hat{Y}$ is a direct sum of shift matrices of size smaller than d_1, $\hat{\Gamma}^{d_1-1} = 0$ so that the matrix of interest reduces to $V_2 C^{d_1-1} V_2^H = \Gamma^{d_1-1} + \Delta^{d_1-1}$.

Denote $\mathcal{J}_{rs}$ as the r-fold direct sum of J_s, e.g., $\mathcal{J}_{3s} = J_s \oplus J_s \oplus J_s$. Letting $s_1 > s_2 > \cdots > s_u > d_1$, one can write

$$Y = \mathcal{J}_{r_1 s_1} \oplus \cdots \oplus \mathcal{J}_{r_u s_u} \tag{56}$$

for some r_i, s_i, u. Likewise, partition $T_1 = [T_{11} \quad \cdots \quad T_{1u}]$, with T_{1i} having the same number of columns as $\mathcal{J}_{r_i s_i}$. Note that the matrices $\Gamma_i \triangleq V_2 T_{1i} \mathcal{J}_{r_i s_i} T_{1i}^\# V_2^H$ satisfy $\Gamma^k = \sum_{i=1}^u \Gamma_i^k$ $\forall k$, and

$$\Gamma_i^k = V_2 T_{1i} \mathcal{J}_{r_i s_i}^k T_{1i}^\# V_2^H = \begin{cases} 0, & \text{if } k > s_i - 1 \\ V_2 T_{1i} \left(\bigoplus_{j=1}^{r_i} e_{s_i} e_1^H \right) T_{1i}^\# V_2^H, & \text{if } k = s_i - 1. \end{cases} \tag{57}$$

We propose the following "peeling" algorithm for estimating $\Gamma_1, \ldots, \Gamma_u$ and Δ^{d_1-1}.

For $i = 1$ to u do:
  Step 1: Find $\Delta^{s_i-1} = V_2 T_2 Z^{s_i-1} T_2^\# V_2^H$.
  Step 2: Find $\Gamma_i = V_2 T_{1i} \mathcal{J}_{r_i s_i} T_{1i}^\# V_2^H$.
end for;
Step 3: Find $\Delta^{d_1-1} = V_2 T_2 Z^{d_1-1} T_2^\# V_2^H$.

At the end, one obtains $V_2 C^{d_1-1} V_2^H = \sum_{i=1}^{u} \Gamma_i^{d_1-1} + \Delta^{d_1-1}$. We now show how these steps can be accomplished.

Steps 1 and 3: With $\overline{R} \triangleq V_1 J_{d_1} V_1^H$, in the ith iteration of Step 1, the matrix

$$R_i \triangleq R - \sum_{j=1}^{i-1} \Gamma_j = \overline{R} + \hat{\Gamma} + \sum_{j=i}^{u} \Gamma_j + \Delta \tag{58}$$

is available ($R_1 \triangleq R$). Similarly, at Step 3, $R_{u+1} = R - \Gamma = \overline{R} + \hat{\Gamma} + \Delta$ is available. One has

$$k \geq s_i \quad \Longrightarrow \quad R_i^k = \left(V_2 T_2 Z T_2^{\#} V_2^H\right)^k = \Delta^k. \tag{59}$$

Let $\rho(\lambda) = \lambda^t + \rho_1 \lambda^{t-1} + \cdots + \rho_t$ be the characteristic polynomial of the $t \times t$ matrix Z. As Z is nonsingular, $\rho_t \neq 0$. By the Cayley–Hamilton theorem, $\rho(Z) = 0$ so that $Z^{s_i-1}\rho(Z) = 0$, i.e.,

$$Z^{s_i-1} = -\frac{1}{\rho_t}\left(Z^{s_i+t-1} + \rho_1 Z^{s_i+t-2} + \cdots + \rho_{t-1} Z^{s_i}\right)$$

which in view of (59) implies

$$\begin{aligned}\Delta^{s_i-1} &= -\frac{1}{\rho_t}\left(\Delta^{s_i+t-1} + \rho_1 \Delta^{s_i+t-2} + \cdots + \rho_{t-1}\Delta^{s_i}\right)\\ &= -\frac{1}{\rho_t}\left(R_i^{s_i+t-1} + \rho_1 R_i^{s_i+t-2} + \cdots + \rho_{t-1} R_i^{s_i}\right).\end{aligned} \tag{60}$$

Similarly, at Step 3, one computes

$$\Delta^{d_1-1} = -\frac{1}{\rho_t}\left(R_{u+1}^{d_1+t-1} + \rho_1 R_{u+1}^{d_1+t-2} + \cdots + \rho_{t-1} R_{u+1}^{d_1}\right).$$

Step 2: Observe that $R_i^{s_i-1} = \Gamma_i^{s_i-1} + \Delta^{s_i-1}$. Therefore, at the ith iteration, $\Gamma_i^{s_i-1} = R_i^{s_i-1} - \Delta^{s_i-1}$ is available. For notational convenience, let $\Phi \triangleq V_2 T_{1i}$, $\Psi^H \triangleq T_{1i}^{\#} V_2^H$ so that $\Gamma_i = \Phi \mathcal{J}_{r_i s_i} \Psi^H$. Note that $\Psi^H \Phi = I$, i.e., $\Psi^H = \Phi^{\#}$. Let

$$\begin{aligned}R^{\dagger} &\triangleq V\left(I_{d_1} \oplus T\right)\left(J_{d_1}^H \oplus \hat{Y}^H \oplus Y^H \oplus Z^{-1}\right)\left(I_{d_1} \oplus T^{-1}\right) V^H\\ &= V_1 J_{d_1}^H V_1^H + (V_2 \hat{T}_1)\hat{Y}^H\left(\hat{T}_1^{\#} V_2^H\right)\\ &\quad + (V_2 T_1) Y^H \left(T_1^{\#} V_2^H\right) + (V_2 T_2) Z^{-1}\left(T_2^{\#} V_2^H\right).\end{aligned} \tag{61}$$

There exists a known permutation matrix Ω such that $R^{\dagger} = \Omega R^{\#}$; thus, $R^{\dagger}$ is available. In addition

$$\begin{aligned}&V_1^H \Phi = 0, \quad \left(T_2^{\#} V_2^H\right)\Phi = 0, \quad \left(T_{1i}^{\#} V_2^H\right)\Phi = I\\ &\left(T_{1j}^{\#} V_2^H\right)\Phi = 0 \qquad \text{for } j \neq i.\end{aligned} \tag{62}$$

Therefore, from (61) and (62)

$$R^{\dagger}\Phi = \Phi \mathcal{J}_{r_i s_i}^H. \tag{63}$$

Similarly, one finds that

$$(R^{\dagger})^H \Psi = \Psi \mathcal{J}_{r_i s_i}. \tag{64}$$

For convenience, let $r = r_i$ and $s = s_i$. Partition

$$\begin{aligned}\Phi &= [\,\Phi_1 \ \ \cdots \ \ \Phi_r\,]\\ &= [\,\phi_{11} \ \ \cdots \ \ \phi_{1s} \ \ | \ \ \cdots \ \ | \ \ \phi_{r1} \ \ \cdots \ \ \phi_{rs}\,]\end{aligned} \tag{65}$$

$$\begin{aligned}\Psi &= [\,\Psi_1 \ \ \cdots \ \ \Psi_r\,]\\ &= [\,\psi_{11} \ \ \cdots \ \ \psi_{1s} \ \ | \ \ \cdots \ \ | \ \ \psi_{r1} \ \ \cdots \ \ \psi_{rs}\,]\end{aligned} \tag{66}$$

and note from (63) and (64) the following Jordan chain relations for each $j = 1, \ldots, r$:

$$R^{\dagger}\phi_{jl} = \phi_{j,l-1}, \quad (R^{\dagger})^H \psi_{j,l-1} = \psi_{jl}, \qquad l = 2, \ldots, s. \tag{67}$$

In view of (67), the matrix $\Gamma_i = \Phi \mathcal{J}_{r_i s_i} \Psi^H$ can be written as

$$\begin{aligned}\Gamma_i &= \sum_{j=1}^{r} \Phi_j J_s \Psi_j^H\\ &= \sum_{j=1}^{r} [\,\phi_{j2} \ \ \cdots \ \ \phi_{js}\,][\,\psi_{j1} \ \ \cdots \ \ \psi_{j,s-1}\,]^H\\ &= \sum_{j=1}^{r}\sum_{k=1}^{s-1} (R^{\dagger})^{s-k-1}\phi_{js}\psi_{j1}^H (R^{\dagger})^{k-1}\\ &= \sum_{k=1}^{s-1} (R^{\dagger})^{s-k-1}\left(\sum_{j=1}^{r}\phi_{js}\psi_{j1}^H\right)(R^{\dagger})^{k-1}\\ &= \sum_{k=1}^{s-1} (R^{\dagger})^{s-k-1}\Gamma_i^{s-1}(R^{\dagger})^{k-1}\end{aligned} \tag{68}$$

since from (57), it holds that $\Gamma_i^{s-1} = \sum_{j=1}^{r} \phi_{js}\psi_{j1}^H$. This shows how to find Γ_i from Γ_i^{s-1} and $R^{\dagger}$.

B. *Setting of Theorem 6*

Suppose the conditions of Theorem 6 hold, and therefore according to Theorem 6 the equalizers of delays 0 and $d_1 - 1$ can be obtained from SOS. The algorithm below actually provides equalizers for the intermediate delays as well.

We start by performing an SVD of $C_y(0)$ as in (46). The matrix R is then constructed as in (48), which still satisfies $R = V\overline{C}_s(1)V^H$. In addition, the matrix $\tilde{R}$ is also constructed as per

$$\tilde{R} \triangleq \Sigma^{-1} U_1^H C_y(d_1 - 1) U_1 \Sigma^{-1}. \tag{69}$$

Again our goal is to estimate the columns of V_1. One can check that $\tilde{R}$ satisfies now $\tilde{R} = V\,\overline{C}_s(d_1 - 1)V^H$, which in view of (43) gives $\tilde{R} = v_{1,d_1} v_{1,1}^H$. Hence the vectors $v_{1,1}$, v_{1,d_1} can be obtained from $\tilde{R}$ up to a constant $e^{j\theta}$. The remaining columns of V_1 are recovered via the Jordan chains (50), (51) or some combination of them. The ZF equalizers are then obtained via (47).

VII. Simulation Results

We present the results obtained by the proposed algorithm with four numerical examples. For illustration purposes, when computing the error rates, the phase ambiguity $e^{j\theta}$ inherent to the method has been removed. Averages were computed based on 100 independent runs. For simplicity, and to allow square channel matrices for comparison with the algorithm from [6], we have not performed denoising of the covariance matrices.

Example 1: First, we consider the real nonlinear channel from [6, ex. 1], whose expression is

$$y(n) = \sum_{j=0}^{2} h_{1j} a(n-j) + \sum_{j=0}^{1} h_{2j} s_2(n-j) + z(n)$$

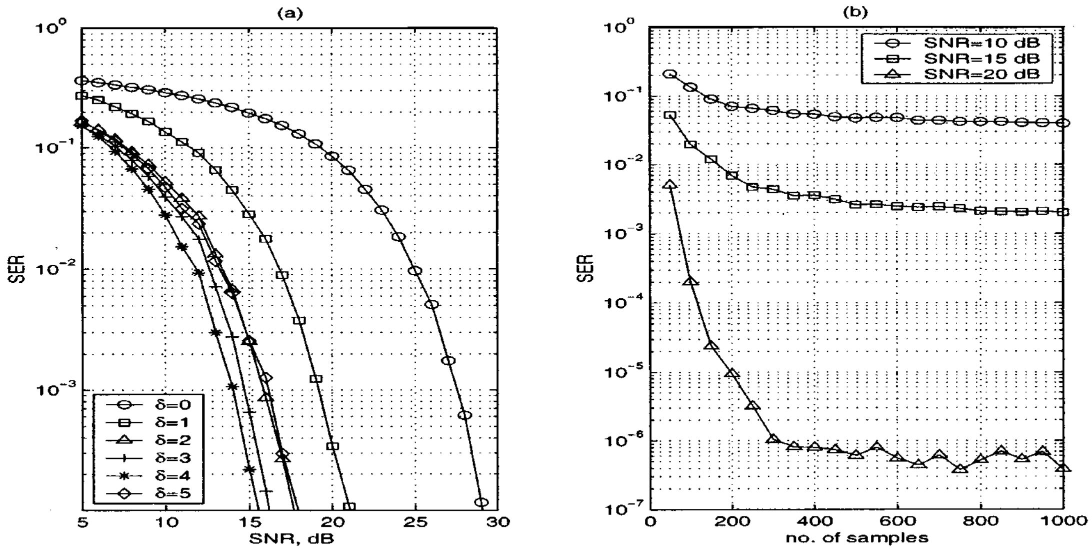

Fig. 1. Performance of the new algorithm for the nonlinear channel in Example 1, equalizer order $m = 4$. (a) SER versus SNR, 500 symbols. (b) SER versus sample size, equalization delay $\delta = 5$.

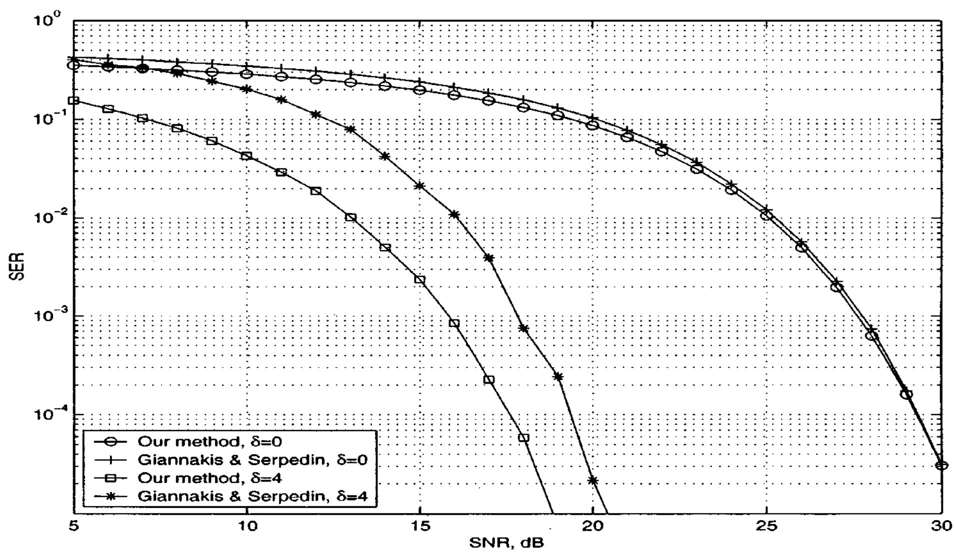

Fig. 2. Nonlinear channel in Example 1. Five hundred symbols, equalizer order $m = 3$.

where $s_2(n) \triangleq a(n)a(n-1)$. The input $a(\cdot)$ is iid BPSK with variance $\sigma_a^2 = 1$, and the number of subchannels is $p = 3$. The noise $z(\cdot)$ is zero-mean white Gaussian with variance σ_z^2. The normalized covariance matrix for this channel is $\overline{C}_s(1) = J_{d_1} \oplus J_{d_2}$, with $d_1 = m + 2$, $d_2 = m + 1$. First, we tested the performance of the equalizers of order $m = 4$ using 500 symbols for covariance estimation for different values of the signal-to-noise ratio (SNR), which is defined as

$$\mathrm{SNR} = 10 \log_{10} \frac{\sigma_a^2 \sum_{j=0}^{l_1} \|h_{1j}\|^2}{\sigma_z^2}.$$

Fig. 1 shows the symbol error rate (SER) versus SNR for the different equalization delays. It is seen that the delay $\delta = 0$ yields the poorest performance. The best results are obtained for $\delta = 4$. The SER as a function of the number of symbols for SNR $= 10$, 15, and 20 dB for the equalizer with associated delay $\delta = 5$ is also shown.

We have compared the performance of our proposed algorithm with that of Giannakis and Serpedin [6] using an equalizer of order $m = 3$ to obtain a square channel matrix $\mathcal{F}$. A drawback of the algorithm from [6] is that it only provides equalizers with minimal and maximal delays—in this case $\delta = 0$ and $\delta = 4$—whereas, as seen in Fig. 1, in general, the best performance is attained for some intermediate delay. Fig. 2 shows the SER as a function of the SNR using 500 symbols. It is seen that the performance of the two algorithms is very close for $\delta = 0$. For $\delta = 4$, however, the new algorithm clearly outperforms the method from [6].

Example 2: The second channel that we consider is the complex channel from [6, ex. 3]:

$$y(n) = \sum_{j=0}^{3} h_{1j} a(n-j) + \sum_{j=0}^{1} h_{2j} s_2(n-j) + z(n)$$

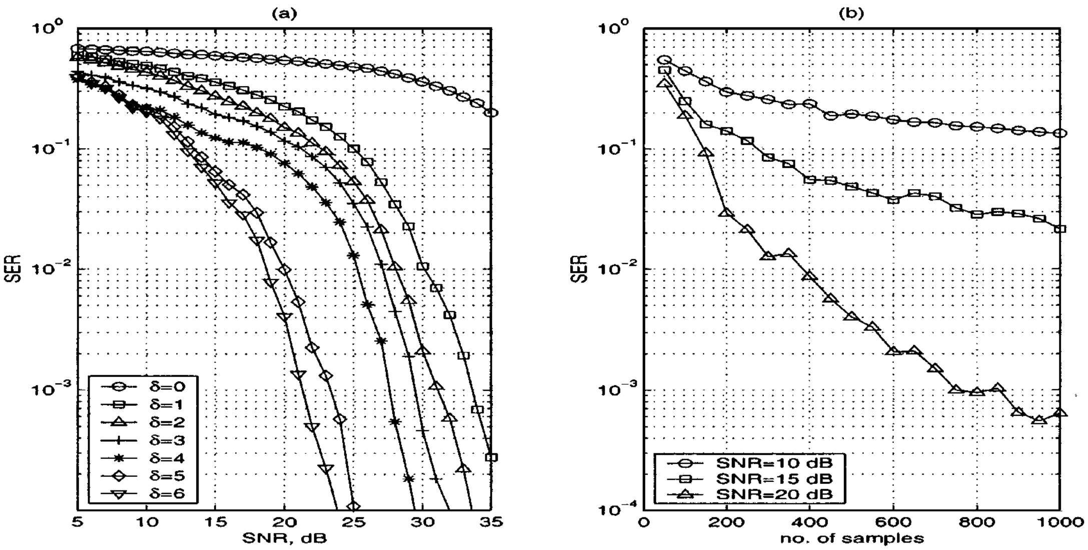

Fig. 3. Performance of the new algorithm for the nonlinear channel in Example 2, equalizer order $m = 4$. (a) SER versus SNR, 500 symbols. (b) SER versus sample size, equalization delay $\delta = 6$.

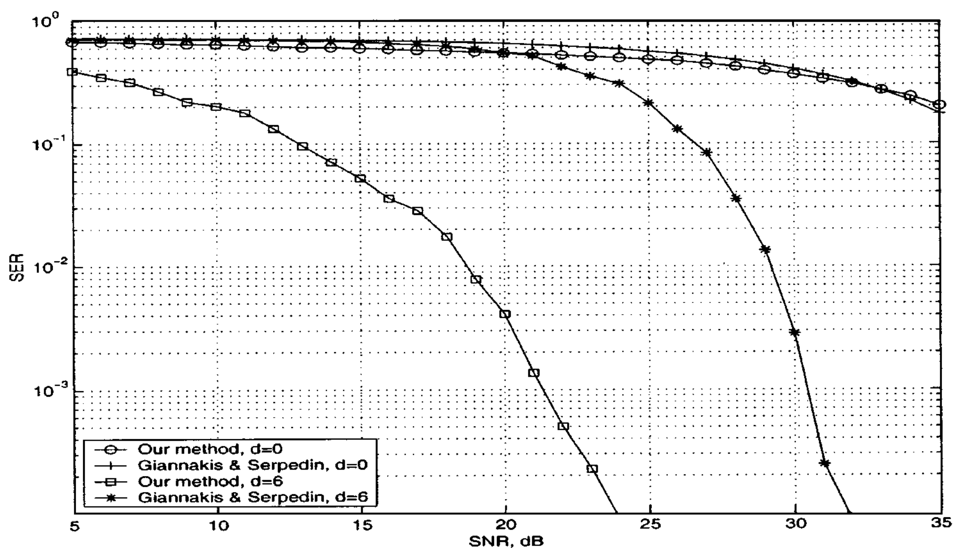

Fig. 4. Nonlinear channel in Example 2. Five hundred symbols, equalizer order $m = 4$.

where now, $s_2(n) \triangleq a(n)a(n-1)a^*(n-2)$. The iid symbols $a(\cdot)$ are drawn from a QPSK constellation $\{\pm 1 \pm j\}$ with equal probabilities so that $\sigma_a^2 = 2$. The number of subchannels is $p = 3$. For this channel, one has $\overline{C}_s(1) = \mathrm{diag}(J_{d_1}, J_{d_2})$ with $d_1 = m + 3$, $d_2 = m + 1$. Fig. 3 shows the performance of the equalizers with order $m = 4$ for different values of the delay, SNR, and sample size. In this case, performance improves as the equalization delay is increased.

Fig. 4 compares the performance of the equalizers of order $m = 4$ obtained with our algorithm and with the method from [6], both using 500 samples. As was the case for the channel in Example 1, for $\delta = 0$, both algorithms show similar performance, whereas for the maximum delay, the proposed method presents a clear advantage. Fig. 5 shows typical scatter plots of the subchannel outputs and the equalized signal, with 500 samples, $m = 4$, $\delta = 6$, and SNR $= 30$ dB.

Example 3: The third channel that we consider is given by

$$y(n) = \sum_{j=0}^{1} h_{1j} a(n-j) + \sum_{j=0}^{1} h_{2j} s_2(n-j) + z(n)$$

where the nonlinear term is $s_2(n) \triangleq a^2(n)a^2(n-1)$. There are $p = 3$ subchannels given by

$$h_{10} = \begin{bmatrix} 1 \\ 0.2 \\ 0.4 \end{bmatrix}, \quad h_{11} = \begin{bmatrix} -0.5 \\ -0.3 \\ 1 \end{bmatrix}, \quad h_{20} = \begin{bmatrix} 0.15 \\ 0.15 \\ 0.5 \end{bmatrix}$$

$$h_{21} = \begin{bmatrix} -0.2 \\ -0.4 \\ 0.2 \end{bmatrix}.$$

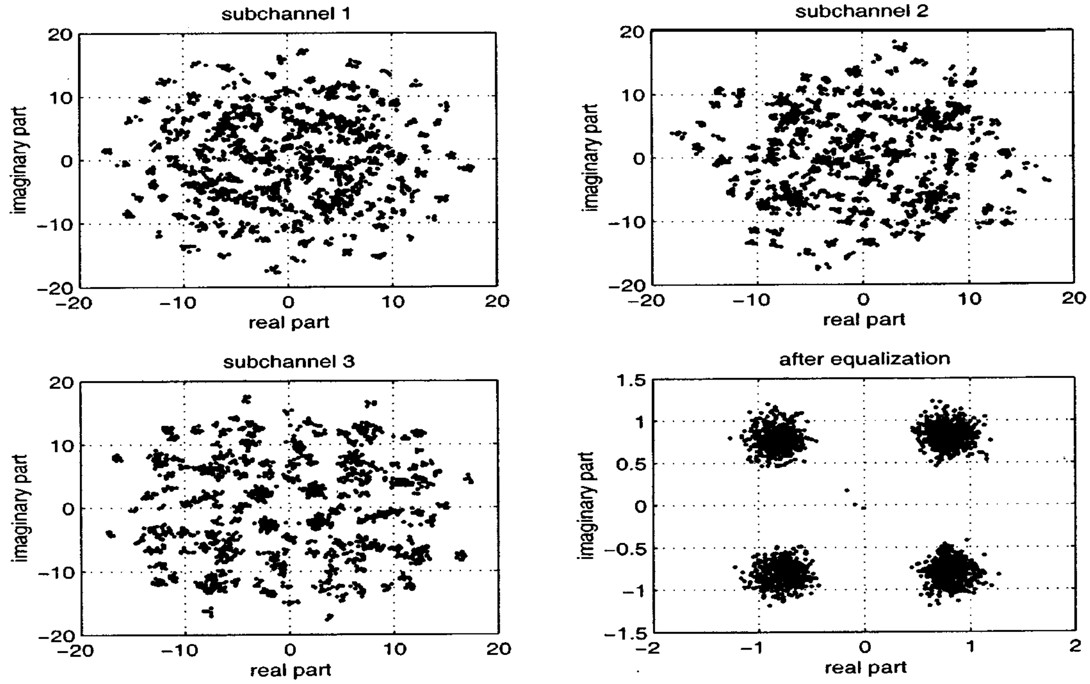

Fig. 5. Performance of the new algorithm for the nonlinear channel in Example 2: Scatter plots. Five hundred symbols, equalizer order $m = 4$, delay $\delta = 6$, SNR $= 30$ dB.

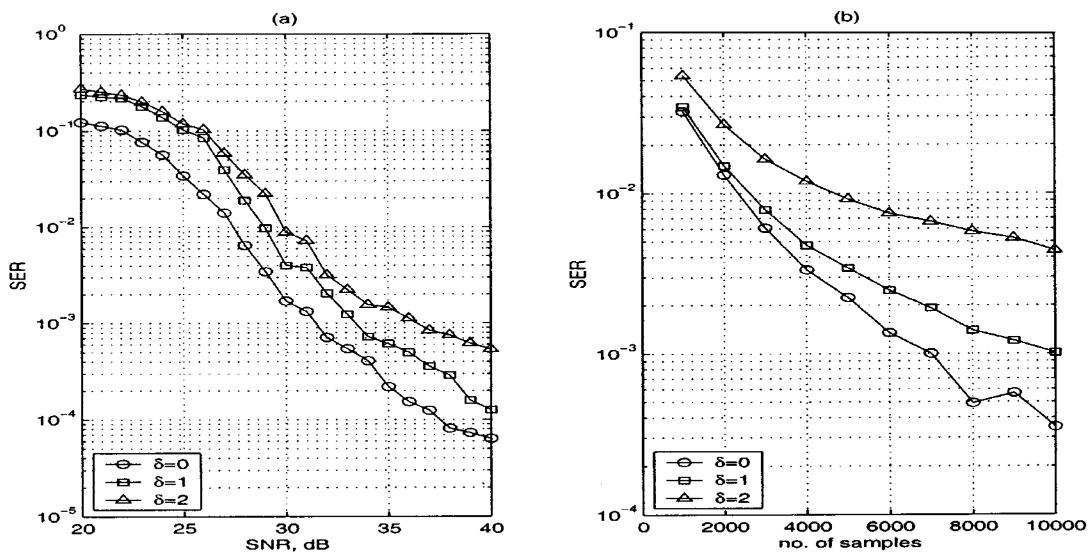

Fig. 6. Performance of the new algorithm for the nonlinear channel in Example 3. Equalizer order $m = 2$. (a) SER versus SNR, 5000 symbols. (b) SER versus sample size, SNR $= 30$ dB.

The input symbols are iid, drawn from a four-PAM constellation $\{\pm(1/3), \pm 1\}$ with probabilities $P(a(n) = -1/3) = P(a(n) = 1/3) = 0.1$, $P(a(n) = -1) = P(a(n) = 1) = 0.4$. For an equalizer order $m = 2$, $\overline{C}_s(1)$ can be shown to be similar to $J_3 \oplus C$ with C given by

$$C = \begin{bmatrix} 0 & 0 & 0.1643 \\ 1 & 0 & -0.3593 \\ 0 & 1 & 0.6216 \end{bmatrix}.$$

Observe that in this case, the linear and nonlinear kernels have the same memory length ($l_0 = l_1 = 1$). However, since all the eigenvalues of C are nonzero, the proposed algorithm can still be used to compute the equalizers. Fig. 6 shows SER versus SNR and sample size. In this case, the best performance is obtained by the equalizer with zero delay.

Observe how, in order to obtain good performance, the algorithm requires considerably more symbols and higher SNR than in the previous examples. A possible explanation is as follows. The singular values of the matrix $\overline{C}_s(1)$ are $\{1, 1, 1, 1, 0.1643, 0\}$; since the blocks associated with the linear and nonlinear kernels have the same size, the algorithm relies in the separation between the two smallest singular values of $\overline{C}_s(1)$, namely, zero (the "linear" singular value) and 0.1643

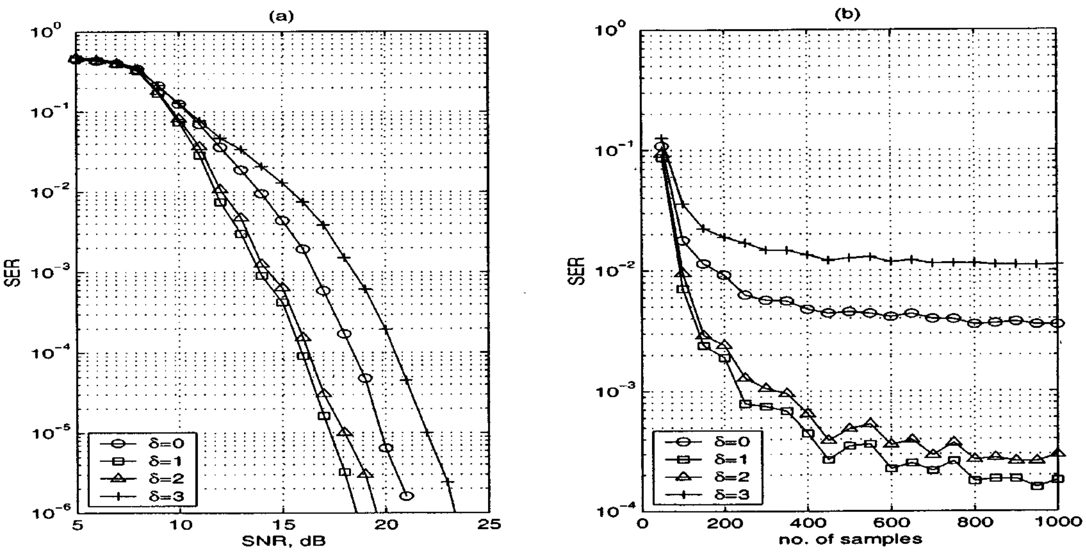

Fig. 7. Performance of the new algorithm for the nonlinear channel in Example 4, equalizer order $m = 3$. (a) SER versus SNR, 500 symbols. (b) SER versus sample size, SNR $= 15$ dB.

(the "nonlinear" singular value). The closer these two numbers are, the more sensitive the algorithm becomes to the effects of noise and finite sample size, as observed in the simulation results.

Example 4: The last channel that we consider has two nonlinear kernels, both longer than the linear one

$$y(n) = \sum_{j=0}^{1} h_{1j} a(n-j) + \sum_{j=0}^{2} h_{2j} s_2(n-j) + \sum_{j=0}^{2} h_{3j} s_3(n-j) + z(n)$$

with $s_2(n) \triangleq a(n)a(n-1)$ and $s_3(n) \triangleq a(n)a(n-2)$. The input symbols are iid, BPSK, and equiprobable, with $\sigma_a^2 = 1$. The number of subchannels is $p = 5$. The channel coefficients are shown in the equation at the bottom of the page. The normalized covariance matrix for this setting is $\overline{C}_s(1) = J_{m+1} \oplus J_{m+2} \oplus J_{m+2}$. Fig. 7 shows the results obtained for the equalizers of order $m = 3$.

VIII. Conclusion

Blind equalization of nonlinear single-input multiple-output channels has been considered. Our approach is based on the second-order statistics of the received signal. We have shown that while SOS do not suffice in general for blind channel identification, under the right conditions, they do enable the determination of an equalizer. A wide range of sufficient conditions on the statistics of the transmitted symbols and the channel nonlinearities ensuring blind equalizability has been presented. Specifically, under the right conditions on the input statistics, SOS-based equalization is possible even if there are kernels with the same length as the linear one.

A procedure has been given in order to compute the equalizers for the special but important case of iid symbols. The algorithm is capable of finding the equalizers for all possible equalization delays, performing better than previous deterministic approaches. This can be justified intuitively since our algorithm explicitly exploits knowledge of the source statistics.

As with most SOS-based methods, the algorithm for computing the equalizers is computationally involved. However, this must be balanced against the need for working with far longer data records required by higher order statistics-based methods. Issues for future work is the extension of this algorithm in order to cover the broader class of channels for which the conditions presented in this work ensure blind equalizability from SOS. Extensions that explicitly exploit cyclostationary nature of channel inputs should also be investigated.

Appendix A
Proof of Theorem 2

Let $C_{11} = SMS^{-1}$, $C_{22} = TNT^{-1}$ be Jordan decompositions of C_{11}, C_{22}. Assume that these have no common eigen-

$$[h_{10} \quad h_{11} \quad h_{20} \quad h_{21} \quad h_{22} \quad h_{30} \quad h_{31} \quad h_{32}] = \begin{bmatrix} 1 & -0.5 & 0.2 & -0.4 & 0.3 & 0.3 & 0.4 & 0.6 \\ 0.2 & -0.5 & 0 & -0.4 & -0.2 & -0.2 & -0.7 & 0.8 \\ -0.3 & 1 & -0.2 & 0.5 & 0.4 & 0.1 & -0.4 & -0.6 \\ -0.3 & 0.7 & 1 & -0.5 & -0.2 & 0.4 & 0.4 & 0.3 \\ 0.8 & 0.4 & 0.1 & -0.5 & 0.2 & -0.4 & 0.1 & -0.2 \end{bmatrix}$$

values. Then, the matrix equation $C_{22}TL - TLC_{11} = C_{21}$ has a solution L [9, p. 414] so that

$$\overline{C}_s(1) = \begin{bmatrix} S & 0 \\ -TLS & T \end{bmatrix} \begin{bmatrix} M & 0 \\ 0 & N \end{bmatrix} \begin{bmatrix} S^{-1} & 0 \\ L & T^{-1} \end{bmatrix} \tag{70}$$

constitutes a Jordan decomposition of $\overline{C}_s(1)$. Now, let

$$W = \begin{bmatrix} W_{11} & W_{12} \\ W_{12}^H & W_{22} \end{bmatrix}, \qquad W_{ij} \text{ of size } d_i \times d_j \tag{71}$$

be a Hermitian matrix commuting with $\overline{C}_s(1)$. Then, W must be of the form [9, pp. 417 and 418]

$$W = \begin{bmatrix} S & 0 \\ -TLS & T \end{bmatrix} \begin{bmatrix} \times & G \\ H^H & \times \end{bmatrix} \begin{bmatrix} S^{-1} & 0 \\ L & T^{-1} \end{bmatrix} \tag{72}$$

where the blocks marked "$\times$" are of no concern. Note that $W_{12} = SGT^{-1}$, but since C_{11}, C_{22} have no common eigenvalues, it follows from [9, pp. 417 and 418] that $G = 0$, and hence, $W_{12} = 0$, which proves the first part of the theorem.

Before proving the second part, let us introduce the following definition.

Definition 2 (Set $\mathcal{U}$): An $n \times m$ matrix is said to belong to the set $\mathcal{U}$ if it is of the form $[U \;\; 0]$ when $n < m$, $[0 \;\; U^H]^H$ when $n > m$, or $\overline{U}$ when $n = m$, with U any square upper triangular matrix and $\overline{U}$ any square upper triangular matrix with zeros on the diagonal.

The following fact is readily verified.

Lemma 1: The set $\mathcal{U}$ is closed under addition and multiplication.

Now, assume $C_{21} = 0$ and that C_{11}, C_{22} do not have common Jordan blocks. Then, (70) with $L = 0$ is a Jordan decomposition of $\overline{C}_s(1)$. Let W, as in (71), be a Hermitian matrix commuting with $\overline{C}_s(1)$. Then, W satisfies (72) with $L = 0$. Let $M = M_1 \oplus \cdots \oplus M_t$, $N = N_1 \oplus \cdots \oplus N_c$, with $M_i = \alpha_i I_{r_i} + J_{r_i}^H$, $N_j = \lambda_j I_{s_j} + J_{s_j}^H$ elementary Jordan blocks, and partition $G = [G_{ij}]_{1\le i\le t}^{1\le j\le c}$, $H = [H_{ij}]_{1\le i\le t}^{1\le j\le c}$ accordingly, where each G_{ij}, H_{ij} have size $r_i \times s_j$. By assumption, if $\alpha_i = \lambda_j$, then $r_i \ne s_j$. Therefore, from [9, pp. 417 and 418], it follows that all G_{ij}, H_{ij}^H are in $\mathcal{U}$.

Observe that $W_{12} = SGT^{-1}$ and $W_{12}^H = TH^H S^{-1}$. Then, $GH^H = S^{-1}W_{12}W_{12}^H S$ is similar to the Hermitian positive semidefinite matrix $W_{12}W_{12}^H$. If one shows that GH^H is nilpotent, then all the eigenvalues of $W_{12}W_{12}^H$ will be zero, and thus, $W_{12} = 0$. This we proceed to show.

Write $D \triangleq GH^H = [D_{ij}]_{1\le i,j\le t}$, with each D_{ij} $r_i \times r_j$. Then, from Lemma 1, each D_{ij} is in $\mathcal{U}$. Without loss of generality, assume $r_i \ge r_{i+1}$. We prove the nilpotence of D by induction on t.

Clearly, when $t = 1$, D_{11} being square and in $\mathcal{U}$ is zero diagonal upper triangular, all its eigenvalues are zero so that D is nilpotent. Now, suppose nilpotence holds for $t = \tau - 1$. Then, for $t = \tau$, write

$$D = \begin{bmatrix} \overline{D}_0 & \overline{D}_1 \\ \overline{D}_2 & D_{\tau\tau} \end{bmatrix}$$

where $\overline{D}_1 \triangleq [D_{1\tau}^H \cdots D_{\tau-1,\tau}^H]^H$, and $\overline{D}_2 \triangleq [D_{\tau 1} \cdots D_{\tau,\tau-1}]$. Then

$$\begin{aligned} \det(\lambda I - D) &= \det(\lambda I - D_{\tau\tau}) \\ &\quad \cdot \det\left[(\lambda I - \overline{D}_0) - \overline{D}_1(\lambda I - D_{\tau\tau})^{-1}\overline{D}_2\right] \\ &= \lambda^{r_\tau} \det\left\{\lambda I - \left[\overline{D}_0 + \overline{D}_1(\lambda I - D_{\tau\tau})^{-1}\overline{D}_2\right]\right\}. \end{aligned}$$

As $\lambda I - D_{\tau\tau}$ is $r_\tau \times r_\tau$ upper triangular, and each $D_{\tau i}$, $1 \le i < \tau$ is $r_\tau \times r_i$ and in $\mathcal{U}$, it follows that $(\lambda I - D_{\tau\tau})^{-1}\overline{D}_2$ is $r_\tau \times r_i$ and in $\mathcal{U}$. Thus, $\overline{D}_1(\lambda I - D_{\tau\tau})^{-1}\overline{D}_2$, and hence, $\overline{D}_0 + \overline{D}_1(\lambda I - D_{\tau\tau})^{-1}\overline{D}_2$ can also be partitioned into $r_i \times r_j$ blocks, $1 \le i, j < \tau$, all in $\mathcal{U}$. Thus, by the induction hypothesis

$$\det\left\{\lambda I - \left[\overline{D}_0 + \overline{D}_1(\lambda I - D_{\tau\tau})^{-1}\overline{D}_2\right]\right\} = \lambda^{r_1 + \cdots + r_{\tau-1}}.$$

Hence, the result. ∎

Appendix B
Proof of Theorem 3

If $a(\cdot)$ is generated via (33), then the forward prediction error is just $f(n) = w(n)$, which by assumption is iid. Since $S_2(n-1)$ is a function of $\{f(k),\ k < n\}$, this means that the random variables $S_2(n-1)$ and $f(n)$ are independent, and therefore, $\mathrm{cov}[S_2(n-1), f(n)] = 0$. Thus, with Q as in (24), $\overline{C}_s(1)$ is lower triangular as in (32). We can apply the result of Theorem 2 to conclude that if the diagonal blocks of $\overline{C}_s(1)$, which are similar to $J_{d_1} + e_1\alpha^H$ and $A_0^{-1}(B_0 - ve_{d_1}^H A_{11}^{-1}A_{12})$, do not share any eigenvalues, then any Hermitian W commuting with $\overline{C}_s(1)$ is block diagonal as in (17). Hence, QWQ^{-1} is block lower triangular with (1, 1) block $A_{11}^{1/2}W_{11}A_{11}^{-1/2}$. It suffices to show that $W_{11} = \theta I_{d_1}$ for some scalar θ.

Let $\lambda_1, \ldots, \lambda_s$ be the distinct eigenvalues of $J_{d_1} + e_1\alpha^H$ with multiplicities $m_1, \ldots, m_s$ ($m_1 + \cdots + m_s = d_1$). Note that these are the zeros of the FPEF of order d_1. Since $\Upsilon > 0$, the FPEF is minimum phase [12] i.e., $|\lambda_i| < 1$ for $1 \le i \le s$. Since $J_{d_1} + e_1\alpha^H$ is a companion matrix, it has a Jordan decomposition

$$J_{d_1} + e_1\alpha^H = KNK^{-1} = K(N_1 \oplus \cdots \oplus N_s)K^{-1}$$

such that $N_i = \lambda I_{m_i} + J_{m_i}^H$, i.e., there is only one Jordan block per distinct eigenvalue [9]. In addition, K is a generalized Vandermonde matrix given by $K = [K_1 \;\; \cdots \;\; K_s]$, where $K_i = [K_{i,1} \;\; \cdots \;\; K_{i,m_i}]$, and

$$K_{i,k} = \frac{1}{(k-1)!}\left\{\frac{\mathrm{d}^{k-1}}{\mathrm{d}z^{k-1}}\left[z^{d_1-1} \;\; \cdots \;\; z \;\; 1\right]^T\right\}_{z=\lambda_i} \tag{73}$$

(see, e.g., [9, pp. 69 and 70]). Now, since W_{11} and $A_{11}^{-1/2}KNK^{-1}A_{11}^{1/2}$ commute, one must have

$$W_{11} = A_{11}^{-1/2}K(Y_1 \oplus \cdots \oplus Y_s)K^{-1}A_{11}^{1/2} \tag{74}$$

with each Y_i $m_i \times m_i$ upper triangular Toeplitz [9, pp. 416–418]. Since W_{11} is Hermitian, it is diagonalizable with real eigenvalues; in view of (74), this must also be true for $Y_1, \ldots, Y_s$. Since these are upper triangular Toeplitz, this in turn gives $Y_i = \theta_i I_{m_i}$, $1 \le i \le s$ for some real scalars θ_i. We will show that all the θ_i are equal. $\theta_1 = \cdots = \theta_s \triangleq \theta$, yielding $W_{11} = \theta I_{d_1}$ as desired.

Let $\Theta \triangleq \theta_1 I_{m_1} \oplus \cdots \oplus \theta_s I_{m_s}$. Since $W_{11} = W_{11}^H$, one has

$$A_{11}^{-1/2} K \Theta K^{-1} A_{11}^{1/2} = A_{11}^{H/2} K^{-H} \Theta K^H A_{11}^{-H/2}$$
$$\Leftrightarrow \quad (K^H A_{11}^{-1} K)\Theta = \Theta(K^H A_{11}^{-1} K)$$

which reads as $(\theta_i - \theta_j)\,(K_i^H A_{11}^{-1} K_j) = 0$ for $1 \le i,\, j \le s$. It suffices to show that $K_i^H A_{11}^{-1} K_j$ has at least one nonzero element for every $i,\, j$. In particular, from (73), the (1,1) element of $K_i^H A_{11}^{-1} K_j$ is given by $K_{i,1}^H A_{11}^{-1} K_{j,1} = P(\lambda_i,\, \lambda_j)$, where the bivariate polynomial $P(z,\, w)$ is defined as

$$P(z,\, w) \triangleq \left[(z^*)^{d_1 - 1} \ \ \cdots \ \ z^* \ \ 1\right] A_{11}^{-1} \left[w^{d_1 - 1} \ \ \cdots \ \ w \ \ 1\right]^T.$$

Using the Christoffel–Darboux formula [8], the polynomial $P(z,\, w)$ can also be written as

$$P(z,\, w) = \frac{\alpha(1/z^*)\alpha^*(1/w^*) - \beta(1/z^*)\beta^*(1/w^*)}{\gamma^2(1 - z^* w)}$$

where γ is a real constant, $\alpha(z) = z^{-d_1} \det[zI - (J + e_1 \alpha^H)]$, and $\beta(z) = z^{-d_1}\alpha^*(1/z^*)$. [Specifically, $\alpha(z)$ and $\beta(z)$ are the transfer functions of the FPEF and BPEF of order d_1 for the process $a(\cdot)$, and γ^2 is the variance of the corresponding prediction errors.]

Since both $\lambda_i,\, \lambda_j$ are roots of $\alpha(z)$, one has $\beta(1/\lambda_i^*) = \beta^*(1/\lambda_j^*) = 0$ so that

$$P(\lambda_i,\, \lambda_j) = \frac{\alpha(1/\lambda_i^*)\alpha^*(1/\lambda_j^*)}{\gamma^2(1 - \lambda_i^* \lambda_j)}. \tag{75}$$

Now, all roots of $\alpha(z)$ lie strictly inside the unit circle, whereas both $1/\lambda_i^*$ and $1/\lambda_j^*$ lie strictly outside the unit circle. Therefore, in view of (75), $P(\lambda_i,\, \lambda_j)$ cannot be zero. ∎

References

[1] L. Agarossi, S. Bellini, A. Canella, and P. Migliorati, "A Volterra model for the high density optical disc," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 3, Seattle, WA, pp. 1605–1608.

[2] S. Benedetto, E. Biglieri, and V. Castellani, *Digital Transmission Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1987.

[3] E. Biglieri, E. Chiaberto, G. P. Maccone, and E. Viterbo, "Compensation of nonlinearities in high-density magnetic recording channels," *IEEE Trans. Magn.*, vol. 30, pp. 5079–5086, Nov. 1994.

[4] J. A. C. Bingham, "Multicarrier modulation for data transmission: An idea whose time has come," *IEEE Commun. Mag.*, vol. 28, pp. 5–14, May 1990.

[5] D. Brillinger, *Time Series: Data Analysis and Theory*. San Francisco, CA: Holden-Day, 1981.

[6] G. B. Giannakis and E. Serpedin, "Linear multichannel blind equalizers of nonlinear FIR Volterra channels," *IEEE Trans. Signal Processing*, vol. 45, pp. 67–81, Jan. 1997.

[7] R. Hermann, "Volterra modeling of digital magnetic saturation recording channels," *IEEE Trans. Magn.*, vol. 26, pp. 2125–2127, Sept. 1990.

[8] T. Kailath, A. Vieira, and M. Morf, "Inverses of Toeplitz operators, innovations, and orthogonal polynomials," *SIAM Rev.*, vol. 20, pp. 106–119, Jan. 1978.

[9] P. Lancaster and M. Tismenetsky, *The Theory of Matrices*. New York: Academic, 1985.

[10] R. López-Valcarce and S. Dasgupta, "Some properties of the matrix exponential," *IEEE Trans. Circuits Syst. II*, vol. 48, pp. 213–215, Feb. 2001.

[11] ——, "Blind channel equalization with colored sources based on second-order statistics," *IEEE Trans. Signal Processing*, vol. 49, pp. 2050–2059, Sept. 2001.

[12] J. Makhoul, "Linear prediction: A tutorial review," *Proc. IEEE*, vol. 63, no. 4, pp. 561–580, Apr. 1975.

[13] V. Z. Marmarelis, *Advanced Methods of Physiological System Modeling*. New York: Plenum, 1994.

[14] 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.

[15] G. M. Raz and B. D. Van Veen, "Blind equalization and identification of nonlinear and IIR systems—A least squares approach," *IEEE Trans. Signal Processing*, vol. 48, pp. 192–200, Jan. 2000.

[16] A. Scaglione, G. B. Giannakis, and S. Barbarossa, "Redundant filterbank precoders and equalizers, Parts I and II," *IEEE Trans. Signal Processing*, vol. 47, pp. 1988–2022, July 1999.

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

[18] 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.

[19] M. Tsatsanis and H. Cirpan, "Blind identification of nonlinear channels excited by discrete alphabet inputs," in *Proc. IEEE Signal Process. Workshop Statist. Signal Array Process.*, vol. 1, Corfu, Greece, pp. 176–179.

[20] 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.

[21] E. Zervas, J. Proakis, and V. Eyuboglu, "Effects of constellation shaping on blind equalization," *SPIE vol. 1565 Adaptive Signal Process.*, pp. 178–187, Aug. 1991.

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

From 1995 to 1996, he was a systems engineer with Intelsis Sistemas Inteligentes. He was the recipient of a Fundación Pedro Barrié de la Maza fellowship for continuation of studies. He has been working as an a research associate 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, St. Lucia, Australia, in 1980 and the Ph.D. degree 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 neural networks.

Between 1988 and 1991, he served as an Associate Editor for the IEEE TRANSACTIONS ON AUTOMATIC CONTROL. 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.