Linear Multichannel Blind Equalizers of Nonlinear FIR Volterra Channels

Georgios B. Giannakis, *Senior Member, IEEE,* and Erchin Serpedin, *Student Member, IEEE*

***Abstract*— Truncated Volterra expansions model nonlinear systems encountered with satellite communications, magnetic recording channels, and physiological processes. A general approach for blind deconvolution of single-input multiple-output Volterra finite impulse response (FIR) systems is presented. It is shown that such *nonlinear* systems can be blindly equalized using only *linear* FIR filters. The approach requires that the Volterra kernels satisfy a certain coprimeness condition and that the input possesses a minimal persistence-of-excitation order. No other special conditions are imposed on the kernel transfer functions or on the input signal, which may be deterministic or random with unknown statistics. The proposed algorithms are corroborated with simulation examples.**

I. Introduction

IDENTIFICATION of nonlinear systems is of considerable practical interest, since many real-life systems exhibit nonlinear characteristics. Examples of such systems are encountered in satellite and microwave channels with nonlinear amplifiers [17], underwater and magnetic recording channels [3], [11], and physiological modeling [21].

In digital communications, blind equalization approaches are important for the following reasons: No training input and no interruption of the transmission are necessary to equalize the channel. Therefore, for channels exhibiting multipath phenomena, changing characteristics, or high data rates, blind methods are attractive. Satellite communication channels are modeled as a cascade of a linear filter (the uplink channel), followed by a zero memory nonlinearity of polynomial type and by a linear filter (the downlink channel) [2, pp. 533–541]. Although the zero-order memory nonlinearity appears to be time invariant in satellite links (and so there is no need to blindly estimate it), the uplink and downlink linear channels are time varying in mobile communications. In this case, a training sequence has to be sent periodically to update the channel coefficients. Blind identification and equalization of such channels is potentially useful since no training sequence needs to be transmitted and, hence, there is no reduction in the effective data rate. Identification of nonlinear dynamics is also a subject of interest in biomedical research, since many physiological signals undergo nonlinear transformations. For example, the auditory nervous system includes memoryless nonlinearities [21, pp. 65–66], and the response of photoreceptors is modeled as a Volterra series expansion [21, pp. 81–90]. Blind identification of such systems is attractive in cases where the design of the experiment (input sequence) may be difficult, or the input to the system is not accessible.

So far, mostly input/output-based (I/O-based) system identification methods have been developed for nonlinear channels (see e.g., [29]), while the blind scenario has not been addressed in its generality. Only methods that assume that the channels and the input signals satisfy special (and often restrictive) conditions have been developed [26]–[28]. For example, the model adopted in [27] consists of two linear subsystems separated by a polynomial-type zero-memory nonlinearity (the LTI–ZMNL–LTI model), which represents a particular case of a Volterra filter with factorizable kernels. In addition, the input sequence is required to be circularly symmetric, the first subsystem can be fully identified only if it is of minimum phase, and the identification of linear subsystems is based on the higher order output polyspectrum. Also, the zero-memory nonlinear subsystem cannot be identified. This limits the use of these algorithms for blind equalization of general nonlinear channels.

The present paper describes a general approach for blind deconvolution (equalization) and identification of nonlinear single-input multiple-output (SIMO) FIR Volterra systems. Although impossible with a single output, multiple outputs make it possible to deconvolve blindly multiple FIR Volterra channels. The approach requires only that a generalized Sylvester resultant, constructed from the channel coefficients, has maximum column rank and that the input signal possesses a certain persistence-of-excitation order—a requirement also encountered with I/O-based methods. The input is allowed to be deterministic or random with unknown color or distribution, the estimation approach is not based on higher order statistics of the input/output signals, and the channel can be any FIR Volterra channel, which satisfies a certain coprimeness condition. Surprisingly, it is shown that *nonlinear* FIR Volterra channels can be perfectly and blindly equalized using *linear* FIR equalizers.

The proposed blind deconvolution and identification method of FIR nonlinear Volterra channels exploits the temporal and/or spatial diversity offered in the form of multichannel output time series. The latter is obtained by oversampling the continuous output of a single sensor at a rate faster than the

Manuscript received December 22, 1995; revised August 21, 1996. Parts of this paper were presented at the 30th Conference on Information Sciences and Systems, Princeton, NJ, March 20–22, 1996, and at the 8th IEEE Signal Processing Workshop on Statistical Signal and Array Processing, Corfu, Greece, June 24–26, 1996.

The authors are with the Department of Electrical Engineering, University of Virginia, Charlottesville, VA 22903-2442 USA (e-mail: georgios@virginia.edu).

Publisher Item Identifier S 1053-587X(97)00537-0.

symbol rate and/or by sampling at the symbol rate the output of a sensor array. Diversity is also exploited in [8], [23], [31], and [33] for blind identification and equalization of *linear* time-invariant FIR channels, and the present work generalizes these ideas to the case of *nonlinear* FIR Volterra models.

The organization of this paper is the following. In Section II, a short description of the I/O-based identification methods for nonlinear Volterra systems is presented first. Second, it is shown how time and space diversity are introduced in the nonlinear framework. Third, a *linear* multi-input multi-output (MIMO) interpretation for the *nonlinear* SIMO FIR Volterra channels is described, over which the present approach is built up. Section III presents basic results concerning the existence and uniqueness of blind linear deconvolvers of Volterra channels. A general approach for deriving blind linear FIR zero-forcing deconvolvers (equalizers) is described in Section IV. Simulations are presented in Section V, and last, comments and concluding remarks are made in Section VI.

II. Preliminaries and Problem Statement

After a brief review of I/O Volterra identification methods, we show how by oversampling the continuous output of a single sensor or by (over)sampling data of an antenna array, an equivalent SIMO nonlinear channel is obtained. The nonlinear SIMO Volterra channel is then viewed as a linear MIMO channel with specifically related inputs. At the end of Section II, we state the problem.

A. I/O-Based Methods for Nonlinear Volterra Systems

Consider a general nonlinear time-invariant system described by $\hat{x}(n) = f[s(n), \cdots, s(n-L)]$. Its sampled Pth-order truncated Taylor expansion has the form

$$x(n) = \sum_{p=1}^{P} \sum_{l_1, \cdots, l_p=0}^{L_p} h_p(l_1, \cdots, l_p) \cdot s(n-l_1)\cdots s(n-l_p) + v(n) \tag{1}$$

where $v(n) = x(n) - \hat{x}(n)$ describes unmodeled dynamics and additive noise. Considering the vectors $\mathbf{s}_p(n)$ and $\mathbf{h}_p(n)$, which have as their entries $s(n-l_1)\cdots s(n-l_p)$, and, respectively, $h_p(l_1, \cdots, l_p)$, for $p = 1, \cdots, P$ and $0 \le l_1, \cdots, l_p \le L_p$, (1) can be rewritten as

$$x(n) = \underbrace{[\mathbf{s}_1'(n) \ \cdots \ \mathbf{s}_P'(n)]}_{:=\mathbf{s}_{1:P}} \underbrace{\begin{bmatrix} \mathbf{h}_1 \\ \vdots \\ \mathbf{h}_P \end{bmatrix}}_{:=\mathbf{h}_{1:P}} + v(n) \tag{2}$$

where prime denotes transpose. The linear-in-the-kernels (1) can be viewed as a regression problem. For $n \in [0, N-1]$, (2) can be solved using the least-squares (LS) approach in the time or frequency domain [13], [14], [29], provided that the input higher order moment matrices involved are invertible [25]. With $\mathbf{x}_{0:N-1}' := [x(0)\cdots x(N-1)]$, the standard LS solution is given by

$$\hat{\mathbf{h}}_{1:P} = (\mathbf{S}_{1:P}'\,\mathbf{S}_{1:P})^{-1}\,\mathbf{S}_{1:P}'\,\mathbf{x}_{0:N-1} \tag{3}$$

where

$$\mathbf{S}_{1:P} = \begin{bmatrix} \mathbf{s}_{1:P}(0) \\ \vdots \\ \mathbf{s}_{1:P}(N-1) \end{bmatrix}.$$

Note that $\mathbf{S}_{1:P}$ must be full rank in order to guarantee invertibility of the higher order matrix $(\mathbf{S}_{1:P}'\mathbf{S}_{1:P})$ in (3). Such a condition is met if the input is sufficiently rich in amplitudes and frequencies, and is referred to as persistence-of-excitation condition (see [24], [25], and the references therein). Computationally efficient orthogonal [15] and adaptive [18], [22] solutions have been proposed for I/O Volterra identification. In [16], closed-form expressions in the frequency domain for the kernels are reported, without orthogonalizing the Wiener functionals [29], but using Gaussian inputs. In all these methods both input and output are required to solve (3). Our focus herein is the blind set-up when $s(n)$ is not available.

B. Time and Space Diversity

Our approach for solving the blind problem exploits additional information provided by time or spatial diversity. Similar to the linear case [23], [31], in the nonlinear case, time and space diversity become available by oversampling the continuous output of a single sensor, and/or by considering sampled outputs of a sensor array. Both possibilities can be cast and treated in the common framework of SIMO channels. We now show how by oversampling (by a factor M) the continuous output of a single sensor, it is possible to obtain a set of M discrete subchannel outputs $x^{(m)}(n)$, $m = 1, 2, \cdots, M$.

Consider the output of a Pth-order baseband continuous-time Volterra channel given by

$$x_c(t) = v_c(t) + \sum_{p=1}^{P} \sum_{l_1, \cdots, l_p=0}^{L_p} \cdot h_{pc}(t-l_1T, \cdots, t-l_pT) s(l_1)\cdots s(l_p)$$

where T is the symbol period, and subscript c denotes continuous time. This truncated Volterra model has been proposed in [2, pp. 58–61 and p. 541] as a baseband model for a bandpass nonlinear channel, and in [11] as a model for the nonlinearities encountered in a magnetic saturation recording channel. We assume perfect synchronization (see e.g., [2, p. 292], for some options for carrier and clock synchronization). Oversampling by a rate of M/T yields

$$\begin{aligned} x(n) &:= x_c(t)|_{t=nT/M} \\ &= v(n) + \sum_{p=1}^{P} \sum_{l_1, \cdots, l_p=0}^{L_p} h_p(n-l_1M, \cdots, n-l_pM) \cdot s(l_1)\cdots s(l_p) \end{aligned}$$

where $h_p(l_1, \cdots, l_p) := h_{pc}(l_1T/M, \cdots, l_pT/M)$ and $v(n) := v_c(nT/M)$. Mimicking the derivation for linear channels (e.g., [6]), it follows easily that time series $x(n)$ is cyclostationary with period M. But upon defining the subprocesses $x^{(m)}(n) := x(nM+m-1)$, $m = 1, \cdots, M$, the

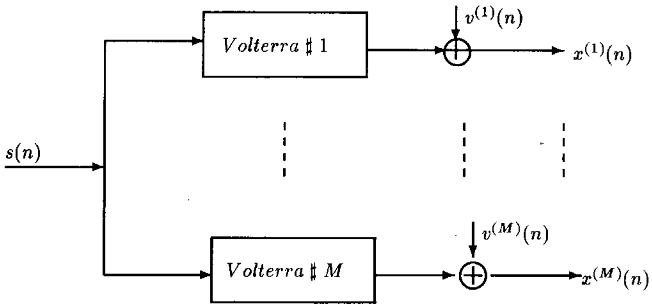

Fig. 1. SIMO nonlinear Volterra channel.

M-channel process $\mathbf{x}'(n) := [x^{(1)}(n) \cdots x^{(M)}(n)]$, becomes stationary, and for $n = 0, 1, \cdots, N-1$ is given by

$$\mathbf{x}(n) = \mathbf{v}(n) + \sum_{p=1}^{P} \sum_{l_1, \cdots, l_p=0}^{L_p} \mathbf{h}_p(l_1, \cdots, l_p) \cdot s(n-l_1) \cdots s(n-l_p) \quad (4)$$

where

i) Lower (upper) bold is used for vectors (matrices).

ii) $M \times 1$ vector $\mathbf{h}_p$ corresponding to the pth-order kernel is defined similar to $\mathbf{x}$

$$\mathbf{h}'_p(l_1, \cdots, l_p) := [h_p^{(1)}(l_1, \cdots, l_p) \cdots h_p^{(M)}(l_1, \cdots, l_p)]$$

with $h_p^{(m)}(l_1, \cdots, l_p) := h_p(l_1 M + m - 1, \cdots, l_p M + m - 1)$ denoting the pth-order kernel of the mth channel.

iii) The inaccessible scalar input $s(n)$ is allowed to be either deterministic or a sample of a random process with unknown distribution.

iv) The range of $(l_1, \cdots, l_p)$ is chosen such that $h_p^{(m)}(l_1, \cdots, l_p)$ is defined over its nonredundant region $0 \le l_1 \le \cdots \le l_p \le L_p$. Note that, as usual, the Volterra kernel is assumed to be symmetric without loss of generality (w.l.o.g.) [29, pp. 41–43 and p. 80], which explains why the Volterra kernels are defined over their nonredundant regions.

v) $\mathbf{v}(n)$ is additive white Gaussian noise (AWGN) $\mathbf{v}'(n) := [v^{(1)}(n) \cdots v^{(M)}(n)]$ with $v^{(m)}(n) := v(nM + m - 1)$.

The structure of a SIMO nonlinear channel is depicted in Fig. 1.

C. Linear MIMO Interpretation

We view the p-dimensional kernel $\mathbf{h}'_p(l_1, \cdots, l_p)$ as a collection of linear (one-dimensional) kernels defined as

$$\mathbf{h}'_{p, i_1:i_{p-1}}(l) := [h_p^{(1)}(l, l+i_1, \cdots, l+i_{p-1}) \cdots h_p^{(M)}(l, \cdots, l+i_{p-1})] \quad (5)$$

where $0 \le i_1 \le \cdots \le i_{p-1} \le L_p$, and $l = 0, 1, \cdots, L_p - i_{p-1}$. In order to compactify notation we, henceforth, use $i_1 : i_{p-1}$ to denote the set $(i_1, \cdots, i_{p-1})$ (for $p = 2$, we consider $i_1 : i_1 = i_1$). Similarly, we define the signals

$$s_{p, i_1:i_{p-1}}(l) := s(l)s(l-i_1) \cdots s(l-i_{p-1}) \quad (6)$$

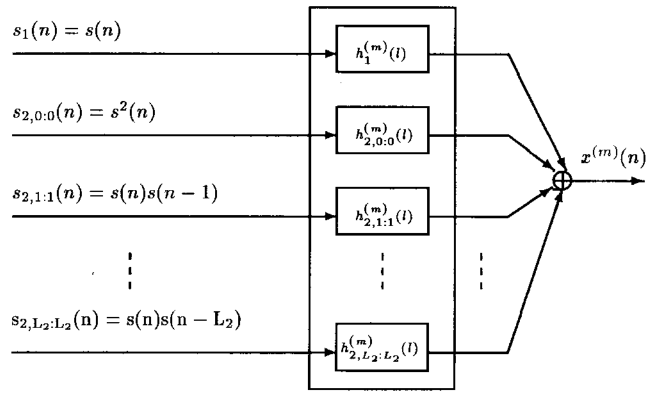

Fig. 2. Structure of the mth subchannel ($P = 2$).

with $0 = i_0 \le i_1 \le \cdots \le i_{p-1} \le L_p$, and denote $\mathbf{h}_1 := \mathbf{h}_{1, i_0:i_0}$, and $s_1 := s_{1, i_0:i_0}$. Using the change of variables $l_p = l + i_{p-1}$, for $p = 1, \cdots, P$, $i_0 = 0$, and definitions (5) and (6), we can rewrite (4) as

$$\mathbf{x}'(n) = \mathbf{v}'(n) + \sum_{p=1}^{P} \sum_{0 \le i_1 \le \cdots \le i_{p-1} \le L_p} \cdot \sum_{l=0}^{L_p - i_{p-1}} \mathbf{h}'_{p, i_1:i_{p-1}}(l) s_{p, i_1:i_{p-1}}(n-l). \quad (7)$$

Equation (7) allows us to view a nonlinear SIMO channel as a linear MIMO channel whose inputs are related (cf., (6)). For example, when $P = 2$, (7) can be rewritten as $\mathbf{x}'(n) = (\mathbf{h}'_1 \star s_1)(n) + \sum_{i_1=0}^{L_2} (\mathbf{h}'_{2, i_1:i_1} \star s_{2, i_1:i_1})(n) + \mathbf{v}'(n)$; i.e., a sum of multichannel linear filters (see Fig. 2). For $P = 2$, we have $s_1(n) = s(n)$, $s_{2, 0:0}(n) = s^2(n), \cdots, s_{2, L_2:L_2}(n) = s(n)s(n-L_2)$; note also that $s_1(n)$ and $s_{2, 0:0}(n)$ are related via $s_{2, 0:0}(n) = s_1^2(n)$.

D. Problem Statement

Given the M-channel system output $\{\mathbf{x}(n)\}_{n=0}^{N-1}$ satisfying (7), we want to blindly deconvolve the system; i.e., we wish to recover both the input sequence $s(n)$ as well as the channel kernels $h_p^{(m)}(l_1, \cdots, l_p)$, $p = 1, \cdots, P, m = 1, \cdots, M$, from knowledge of the received data $\mathbf{x}(n)$ only. Specifically, we seek linear FIR equalizers $\{\mathbf{g}_1^{(d)}(k)\}_{k=0}^{K}$ of order K and delay d, which in the noise-free case satisfy the so called zero-forcing (or perfect equalization) condition

$$\sum_{k=0}^{K} \mathbf{x}'(n-k) \mathbf{g}_1^{(d)}(k) = s(n-d) \quad (8)$$

where the delay (shift) d takes values in $0 \le d \le L_1 + K$ and is nonidentifiable from output data only. Similar to the linear FIR equalizer $\{\mathbf{g}_1^{(d)}(k)\}_{k=0}^{K}$, which deconvolves the linear kernel, we introduce the Kth-order and d-delay FIR equalizer $\{\mathbf{g}_{p, i_1:i_{p-1}}^{(d)}(k)\}_{k=0}^{K}$, which deconvolves the pth-order kernel $\mathbf{h}_{p, i_1:i_{p-1}}(l)$ via [cf., (8)]

$$\sum_{k=0}^{K} \mathbf{x}'(n-k) \mathbf{g}_{p, i_1:i_{p-1}}^{(d)}(k) = s(n-d)s(n-d-i_1) \cdots s(n-d-i_{p-1}) \quad (9)$$

where the $(p-1)$-tuple $(i_1, \cdots, i_{p-1})$ and delay d satisfy $0 \le i_1 \le \cdots \le i_{p-1} \le L_p$ and $0 \le d \le L_p + K - i_{p-1}$, respectively.

When possible to obtain $\mathbf{g}_1^{(d)}$ uniquely, (8) suffices to recover the desired $s(n)$ in the noise-free case. However, because the outputs of $\mathbf{g}_1^{(d)}$ and $\mathbf{g}_{p, i_1:i_{p-1}}^{(d)}$ are related, relation (9) will be helpful when $\mathbf{g}_1^{(d)}$ is not uniquely identifiable from (8). Such identifiability issues will be dealt with for the noise-free case in the ensuing sections, deferring the noisy case to the end of Sections IV and V.

III. Existence-Uniqueness of Linear Equalizers

To study existence and uniqueness of the linear FIR equalizers $\mathbf{g}_{p, i_1:i_{p-1}}^{(d)}$, $p = 1, \cdots, P$, it will be helpful to cast (7) in a matrix form. Toward this objective, define the $(L_1+K+1) \times M(K+1)$, and respectively, $(L_p+K+1-i_{p-1}) \times M(K+1)$ block Toeplitz matrices $\mathbf{H}_1$ and $\mathbf{H}_{p, i_1:i_{p-1}}$, as shown in (9a), shown at the bottom of the page. Define also $(L_1+K+1) \times 1$ and $(L_p+K+1-i_{p-1}) \times 1$ vectors $\mathbf{s}_1(n)$ and $\mathbf{s}_{p, i_1:i_{p-1}}(n)$ through the relations

$$\begin{aligned}\mathbf{s}_1'(n) &:= [s(n) \quad \cdots \quad s(n-L_1-K)],\\ \mathbf{s}_{p, i_1:i_{p-1}}'(n) &:= [s_{p, i_1:i_{p-1}}(n) \cdots s_{p, i_1:i_{p-1}}\\ &\qquad \cdot (n-L_p-K+i_{p-1})]\end{aligned}$$

where $0 \le i_1 \le \cdots \le i_{p-1} \le L_p$, and $s_{p, i_1:i_{p-1}}(n)$ as in (6).

The noise-free input-output relation (7) can now be rewritten in a matrix form as

$$\mathbf{X} = \mathbf{S}\mathbf{H} \tag{10}$$

where the $(N-K) \times M(K+1)$ block Hankel matrix $\mathbf{X}$ is given by

$$\mathbf{X} := \begin{bmatrix} \mathbf{x}'(N-1) & \cdots & \mathbf{x}'(N-1-K) \\ \vdots & \vdots & \vdots \\ \mathbf{x}'(K) & \cdots & \mathbf{x}'(0) \end{bmatrix}$$

the $(N-K) \times D(L_1, \cdots, L_P, K)$ block Hankel input matrix $\mathbf{S}$, and the $D(L_1, \cdots, L_P, K) \times M(K+1)$ block Toeplitz channel matrix $\mathbf{H}$ are given respectively by

$$\mathbf{S} := \begin{bmatrix} \mathbf{s}_1'(N-1) & \cdots & \mathbf{s}_{P, L_P:L_P}'(N-1) \\ \mathbf{s}_1'(N-2) & \cdots & \mathbf{s}_{P, L_P:L_P}'(N-2) \\ \vdots & \vdots & \vdots \\ \mathbf{s}_1'(K) & \cdots & \mathbf{s}_{P, L_P:L_P}'(K) \end{bmatrix} \tag{11}$$

$$\mathbf{H} := \begin{bmatrix} \mathbf{H}_1 \\ \mathbf{H}_{2, 0:0} \\ \vdots \\ \mathbf{H}_{P, L_P:L_P} \end{bmatrix}.$$

The common dimension between $\mathbf{S}$ and $\mathbf{H}$, denoted by $D(L_1, \cdots, L_P, K)$, depends, in general, on $L_1, \cdots, L_P$ and K. Since the number of distinct sequences $(i_1, \cdots, i_{p-1})$ that satisfy $0 \le i_1 \le \cdots \le i_{p-1} = i$ is equal to $\binom{i+p-2}{p-2}$ [20, p. 17], and the matrix $\mathbf{H}_{p, i_1:i_{p-1}}$ has $L_p + K + 1 - i_{p-1}$ rows, it follows that

$$\begin{aligned}D(L_1, \cdots, L_P, K) &= L_1 + K + 1 + \sum_{p=2}^{P} \sum_{i=0}^{L_p}\\ &\quad \cdot \binom{i+p-2}{p-2}(L_p + K + 1 - i).\end{aligned} \tag{12}$$

For $P = 2$, we obtain $D(L_1, L_2, K) = (L_2+2)(K+1) + (L_2^2 + L_2 + 2L_1)/2$.

We adopt the following assumptions:

(a0.1) $\mathbf{v}(n) = \mathbf{0}$, i.e., we consider first noise-free data (see the end of Sections IV and V for the noisy case).

(a0.2) $N \ge \max\{K + d + 2M(K+1), K + P^P\}$. This requirement is introduced to assure more equations (data) than unknowns in subsequent equations (22) and (30). This condition is easily met in practice by collecting sufficient number of samples N (note $K, d, M, P \ll N$).

(a1.1) Rank$(\mathbf{H}) = D(L_1, \cdots, L_P, K)$, which implies that there are no common zeros among the one dimensional kernel transfer functions $\{H_1^{(m)}(z)$, $H_{2,0:0}^{(m)}(z)$, $H_{2,1:1}^{(m)}(z)$, $\cdots$, $H_{P, L_P:L_P}^{(m)}(z)\}$ across all M channels, where $H_{p, i_1:i_{p-1}}^{(m)}(z)$ denotes the Z-transform of sequence $h_{p, i_1:i_{p-1}}^{(m)}(l)$, $l = 0, \cdots, L_p - i_{p-1}$. An alternative characterization of

$$\mathbf{H}_1 := \begin{bmatrix} \mathbf{h}_1'(0) & \cdots & \mathbf{0}' \\ \vdots & \ddots & \vdots \\ \mathbf{h}_1'(L_1) & \cdots & \mathbf{h}_1'(L_1 - K) \\ \vdots & \ddots & \vdots \\ \mathbf{0}' & \cdots & \mathbf{h}_1'(L_1) \end{bmatrix},$$

$$\mathbf{H}_{p, i_1:i_{p-1}} := \begin{bmatrix} \mathbf{h}_{p, i_1:i_{p-1}}'(0) & \cdots & \mathbf{0}' \\ \vdots & \ddots & \vdots \\ \mathbf{h}_{p, i_1:i_{p-1}}'(L_p - i_{p-1}) & \cdots & \mathbf{h}_{p, i_1:i_{p-1}}'(L_p - i_{p-1} - K) \\ \vdots & \ddots & \vdots \\ \mathbf{0}' & \cdots & \mathbf{h}_{p, i_1:i_{p-1}}'(L_p - i_{p-1}) \end{bmatrix}. \tag{9a}$$

(a1.1) in terms of the rank of a generalized Sylvester resultant is provided in Appendix A. Also, note that (a1.1) implies that $\mathbf{H}$ is a wide (or fat) matrix; i.e., $(M, L_1, \cdots, L_P, K)$ obey

$$M(K+1) \geq D(L_1, \cdots, L_P, K). \tag{13}$$

(a1.2) $\mathbf{H}$ is square; i.e., (13) is satisfied with equality.

(a2.1) with $L := \max\{L_1, \cdots, L_P\}$, input $s(n)$ is such that the matrix $\mathcal{S}^{(0, L+K)}$ defined through[1]

$$\mathcal{S}^{(0, L+K)} := [\mathbf{S}(L+K+1 : N-K, :) \\ \cdot \mathbf{S}(1 : N-2K-L, :)] \tag{14}$$

has maximum column rank; i.e.,

$$\text{rank}[\mathcal{S}^{(0, L+K)}] = 2D(L_1, \cdots, L_P, K) - C$$

where C equals the number of pairs of identical columns of $\mathcal{S}^{(0, L+K)}$.

(a2.2) matrix $\overline{\mathbf{S}}$ in (15) is full column rank

$$\overline{\mathbf{S}} := \begin{bmatrix} s(N-1) & s^2(N-1) & \cdots & s^{P^2}(N-1) \\ s(N-2) & s^2(N-2) & \cdots & s^{P^2}(N-2) \\ \vdots & \vdots & \ddots & \vdots \\ s(K) & s^2(K) & \cdots & s^{P^2}(K) \end{bmatrix}. \tag{15}$$

Due to the structure of $\mathcal{S}^{(0, L+K)}$, assumption (a2.1) implies that matrix $\mathbf{S}$ in (11) is full column rank; i.e., the input $s(n)$ is persistently exciting (p.e.) of order $\rho_s \geq D(L_1, \cdots, L_P)$. White noise is p.e. of any order, but $D(L_1, \cdots, L_P, K)$ modes in the spectrum of $s(n)$ may not guarantee p.e. as in the linear case; $s(n)$ must also have sufficient amplitude/phase levels [24], [25]. Note that if, e.g., $s(n) = 0, 1$, matrices $\mathcal{S}^{(0, L+K)}$ and $\mathbf{S}$ are rank deficient because $s(n) = s^2(n) = \cdots = s^P(n)$. As a result, in (4) the kernels $\mathbf{h}_p(0, \cdots, 0)$, $p = 1, \cdots, P$, can be combined. Hence, a violation of the p.e. condition does not allow identification of all kernels, but only of their sum. In Section IV, we will see that the blind identification algorithm of linear equalizers requires, in general, an additional p.e. condition for the input sequence $s(n)$, namely, (a2.2). The next proposition establishes a characterization of (a2.2).

Proposition 1: A necessary and sufficient condition for $\overline{\mathbf{S}}$ to be full column rank is that the input signal $s(n)$, for $n = K, \cdots, N-1$, takes at least P^2 distinct nonzero values.

Proof: Proposition 1 follows from the Vandermonde structure of $\overline{\mathbf{S}}$. Indeed, if $s(n)$, for $n = K, \cdots, N-1$ takes at most $P^2 - 1$ distinct values, then $\overline{\mathbf{S}}$ will have at most $P^2 - 1$ distinct rows. So, $\overline{\mathbf{S}}$ will have rank less than P^2; i.e., $\overline{\mathbf{S}}$ will be column rank deficient. □

In Appendix B, Assumption (a2.1) is shown to hold with probability one, provided that input signal constellation has at least $P+1$ distinct points and N is large enough.

[1] In writing (14), we adopted Matlab's notation $\mathbf{X}(i_1 : i_2, j_1 : j_2)$ to denote the submatrix of $\mathbf{X}$ formed by the i_1 through i_2 rows and the j_1 through j_2 columns of $\mathbf{X}$.

We want to show next that Assumption (a1.2) is not overly restrictive. Note that given $L_1, \cdots, L_P$, one can choose M and K such that (13) holds. From (12), it follows that for fixed $L_1, \cdots, L_P$, the dimension $D(L_1, \cdots, L_P, K)$ depends linearly on K. We can write $D(L_1, \cdots, L_P, K) = \alpha(L_1, \cdots, L_P)K + D_0(L_1, \cdots, L_P)$. In order for (a1.2) to hold, we need equality in (13); i.e., $M(K+1) = \alpha(L_1, \cdots, L_P)K + D_0(L_1, \cdots, L_P)$, or $(K+1)[M - \alpha(L_1, \cdots, L_P)] = D_0(L_1, \cdots, L_P) - \alpha(L_1, \cdots, L_P)$. Hence, if M and K are chosen such that the previous relation holds, then (a1.2) is satisfied. There is no loss of generality in assuming equality instead of strict inequality. In all cases when strict inequality holds, $M(K+1) > D(L_1, \cdots, L_P, K)$, we can obtain the equality of (a1.2), by decreasing M (the number of antennas) and by varying K (the linear equalizer order). For $M = 1$, (13) is not satisfied, which indicates the vital role of diversity in our blind approach.

Considering the case of a second-order system ($P = 2$), relation (13) implies that the minimum number of channels required is $M_{\min} = L_2 + 3$, which depends on the memory of the nonlinearity. Hence, for a memoryless quadratic nonlinearity, we require $M_{\min} = 3$ channels and a minimum equalizer order $K_{\min} = L_1 - 1$ (recall that for linear channels $M_{\min} = 2$ and $K_{\min} = L_1 - 1$). Although the number of antennas (and thus complexity) increases in the nonlinear case, the possibility to equalize nonlinear channels with linear FIR filters is very attractive, since the stability of inverse Volterra systems is difficult to check.

Having stated and characterized some of the assumptions for the multichannel model (10), we now turn to existence and uniqueness issues of the multichannel equalizers introduced in (8) and (9). Consider (8) for $n = N-1, \cdots, K$, and define $\mathbf{g}_1^{(d)} := [\mathbf{g}_1^{(d)}(0)' \cdots \mathbf{g}_1^{(d)}(K)']'$ to obtain the matrix equation

$$\mathbf{X}\mathbf{g}_1^{(d)} = \mathbf{S}(:, d+1) \Leftrightarrow \mathbf{S}\,\mathbf{H}\,\mathbf{g}_1^{(d)} = \mathbf{S}(:, d+1). \tag{16}$$

Under (a2.1), (16) holds if and only if $\mathbf{H}\,\mathbf{g}_1^{(d)} = \mathbf{e}_{d+1}$, where $\mathbf{e}_{d+1}$ is a $D(L_1, \cdots, L_P, K) \times 1$ vector with unity in its $(d+1)$st entry and zero elsewhere. From (16), we deduce:

Proposition 2: Let $\mathbf{H}$ be given and assume that (a0.1)–(a1.1) hold true. The pseudoinverse $\mathbf{H}^\dagger$ exists, it is unique, it coincides with $\mathbf{H}^{-1}$ under (a1.2), and its columns constitute the linear FIR equalizers of the channel $\mathbf{H}$.

For arbitrary second-order systems and for systems with orders $P > 2$, which satisfy $L_1 = \cdots = L_P = L$, it is possible to blindly determine the orders L_1, L_2 and, respectively, L. To show the order determination approach, we first establish the following:

Proposition 3: Under (a0)–(a2.1), matrix $\mathbf{X}$ in (10) has

$$\rho_{\mathbf{X}} := \text{rank}(\mathbf{X}) = D(L_1, \cdots, L_P, K). \tag{17}$$

Proof: Relation (17) follows from (10) and Sylvester's theorem:

$$\text{rank}(\mathbf{S}) + \text{rank}(\mathbf{H}) - D(L_1, \cdots, L_P, K) \leq \text{rank}(\mathbf{X}) \\ \leq \min\{\text{rank}(\mathbf{S}), \text{rank}(\mathbf{H})\}.$$ □

Consider now the case of a linear-quadratic system, and let $(\overline{L}_1, \overline{L}_2)$ be known upper bounds for (L_1, L_2), i.e., $\overline{L}_p \geq L_p$, $p = 1, 2$. With M given, choose distinct $\overline{K}_1, \overline{K}_2$ such that for each $p = 1, 2$ the triplet $(\overline{L}_1, \overline{L}_2, \overline{K}_p)$ satisfies (13). With $\overline{K}_p$ in place of K, form matrices $\overline{\mathbf{X}}_p$, $p = 1, 2$, as in (10). From Proposition 3, it follows that

$$\rho_{\overline{\mathbf{X}}_p} := \mathrm{rank}(\overline{\mathbf{X}}_p) = D(L_1, L_2, \overline{K}_p) \tag{18}$$

for $p = 1, 2$. Knowing $\overline{K}_1, \overline{K}_2$, and evaluating $\rho_{\overline{\mathbf{X}}_p}$'s from the singular value decomposition (SVD) of $\overline{\mathbf{X}}_p$, we deduce from (12) and (18) a system of equations that yields (L_1, L_2). The orders L_1, L_2 are given by $L_2 = (\rho_{\overline{\mathbf{X}}_1} - \rho_{\overline{\mathbf{X}}_2})/(\overline{K}_1 - \overline{K}_2) - 2$, and $L_1 = \rho_{\overline{\mathbf{X}}_1} - (L_2 + 2)(\overline{K}_1 + 1) - (L_2^2 + L_2)/2$. For the case when $L_1 = \cdots = L_P = L$, the equation that determines L is similar to (18). For $P > 2$, (18) does not allow the determination of all orders $L_1, \cdots, L_P$. This is due to the fact that the dependence of $D(L_1, \cdots, L_P, K)$ on K, for fixed $L_1, \cdots, L_P$, is linear [see (12)]. Hence, relation (18) provides only two independent equations for $p = 1, \cdots, P$, which permit determination of only one or two unknowns.

In general, the question of determining the orders $L_1, \cdots, L_P$, as well as P, blindly, is an open problem. The orders of nonlinearities encountered with satellite communications and magnetic recording channels are less than or equal to 7, and, respectively, 3. In general, these orders have been determined experimentally [2, p. 543 and 566], [11, p. 2126]. From now on we assume that $L_1, \cdots, L_P$ and P are known and choose K to satisfy (a1.2) for a given M.

IV. Direct Blind Equalizers

In this section, we propose a method for estimating directly the linear multichannel equalizer from knowledge of the vector output $\mathbf{x}(n)$ only. First, we focus on Volterra systems for which there is only one kernel with maximum memory $L := \max\{L_1, \cdots, L_P\}$, and then on models having more than one kernel with maximum memory.

A. One Kernel with Maximum Memory

Substitute $n = n + d$ into (8), to obtain

$$\sum_{k=0}^{K} \mathbf{x}'(n + d - k)\mathbf{g}_1^{(d)}(k) = s(n)$$

and rewrite (8) with $d = 0$ as

$$\sum_{k=0}^{K} \mathbf{x}'(n - k)\mathbf{g}_1^{(0)}(k) = s(n).$$

Eliminating $s(n)$ from the previous two equations we find

$$\sum_{k=0}^{K} \mathbf{x}'(n - k)\mathbf{g}_1^{(0)}(k) = \sum_{k=0}^{K} \mathbf{x}'(n + d - k)\mathbf{g}_1^{(d)}(k). \tag{19}$$

From the block Hankel structure of $\mathbf{S}$ [see (11)], relation (19), for $n = N - d - 1, \cdots, K$, can be written as

$$\mathbf{X}(d+1 : N-K, :)\mathbf{g}_1^{(0)} = \mathbf{X}(1 : N-K-d, :)\mathbf{g}_1^{(d)} \tag{20}$$

where $\mathbf{g}_1^{(0)} := [\mathbf{g}_1^{(0)}(0)' \cdots \mathbf{g}_1^{(0)}(K)']'$, and similarly for $\mathbf{g}_1^{(d)}$. Upon defining

$$\mathcal{X}^{(0,d)} := [\mathbf{X}(d+1 : N-K, :) \quad -\mathbf{X}(1 : N-K-d, :)] \tag{21}$$

we obtain from (20)

$$\mathcal{X}^{(0,d)}\boldsymbol{g}_1^{(0,d)} = \mathbf{0} \tag{22}$$

where $\boldsymbol{g}_1^{(0,d)} := [\mathbf{g}_1^{(0)'}\ \mathbf{g}_1^{(d)'}]'$. The pair of equalizers in $\boldsymbol{g}_1^{(0,d)}$ belongs to the null space $\mathcal{N}[\mathcal{X}^{(0,d)}]$. Equation (20) relates first order equalizers of delays $(d_1, d_2) = (0, d)$. It is possible to start from (9) and derive relations similar to (20) among pth-order equalizers of different delays. Following the notational convention in (9), it follows that $\boldsymbol{g}_p^{(0,d)} := [\mathbf{g}_{p,0:0}^{(0)'}\ \mathbf{g}_{p,0:0}^{(d)'}]'$ also satisfies (22), and thus belongs to $\mathcal{N}[\mathcal{X}^{(0,d)}]$, for any $p = 1, \cdots, P$. Identifiability of the vector $\boldsymbol{g}_1^{(0,d)}$ from (22) depends on the nullity of $\mathcal{X}^{(0,d)}$. Specifically, if $\dim\{\mathcal{N}[\mathcal{X}^{(0,d)}]\} = 1$, for some $d \in [0, L + K]$, then $\boldsymbol{g}_1^{(0,d)}$ will be uniquely identifiable (within a scale) as the null eigenvector of $\mathcal{X}^{(0,d)}$ in (22).

Two factors will determine the nullity of $\mathcal{X}^{(0,d)}$: i) how many kernels in (7) achieve the maximum order (or memory) $L := \max\{L_1, \cdots, L_P\}$; and ii) the delay d adopted in (22). In this subsection, we suppose that L is attained by only one kernel, say the pth-order one; i.e., $L_p = L$, and $L_i < L_p$, for $\forall i \in \{1, \cdots, p-1, p+1, \cdots, P\}$. It turns out that when the delay d increases, the dimension of the null space of $\mathcal{X}^{(0,d)}$ decreases. In order to show this, let us consider the maximum delay $d = L + K$ and decompose the matrix in (22) as $\mathcal{X}^{(0,L+K)} = \mathcal{S}^{(0,L+K)}\mathcal{H}_2$, or in block form [see also (21) and (10)] as follows:

$$\mathcal{X}^{(0,L+K)} = [\mathbf{S}(L+K+1 : N-K, :)\mathbf{S}(1 : N-2K-L, :)] \times \begin{bmatrix} \mathbf{H} & \mathbf{0} \\ \mathbf{0} & -\mathbf{H} \end{bmatrix}. \tag{23}$$

Under (a1.1), $2D(L_1, \cdots, L_P, K) \times 2M(K+1)$ matrix $\mathcal{H}_2$ in (23) has full row rank, and thus the rank$[\mathcal{X}^{(0,L+K)}]$ depends upon the rank of $\mathcal{S}^{(0,L+K)}$. To find rank$[\mathcal{S}^{(0,L+K)}]$ let us zoom in $\mathcal{S}^{(0,L+K)}$ which, based on (11), can be written as shown in (24) at the bottom of the page.

$$\mathcal{S}^{(0,L+K)} = \begin{bmatrix} \mathbf{s}_1'(N-L-K-1) & \cdots & \mathbf{s}_{P,L_P:L_P}'(N-L-K-1) & \mathbf{s}_1'(N-1) & \cdots & \mathbf{s}_{P,L_P:L_P}'(N-1) \\ \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\ \mathbf{s}_1'(K) & \cdots & \mathbf{s}_{P,L_P:L_P}'(K) & \mathbf{s}_1'(2K+L) & \cdots & \mathbf{s}_{P,L_P:L_P}'(2K+L) \end{bmatrix}. \tag{24}$$

From (24), it turns out that $\mathcal{S}^{(0,\,L+K)}$ has only two identical columns ($[s_{p,\,0:0}\cdots s_{p,\,0:0}(K)]'$ is common to both $\mathbf{S}(L+K+1:N-K,:)$ and $\mathbf{S}(1:N-2K-L,:)$), while the other columns are, in general, independent of each other. From (a2.1), it follows that $\dim\{\mathcal{N}[\mathcal{S}^{(0,\,L+K)}]\}=1$; hence, $\dim\{\mathcal{N}[\mathcal{X}^{(0,\,L+K)}]\}=1$. When delay d decreases, the number of pairs of identical columns in $\mathcal{S}^{(0,\,d)}$ increases; hence, the dimension of the null space of $\mathcal{X}^{(0,\,d)}$ increases.

Theorem 1: Suppose that (a0.1)–(a2.1) are satisfied. If $L=\max\{L_1,\cdots,L_P\}$ is attained by a single L_p, where $p\in\{1,\cdots,P\}$, then $\mathbf{g}_{p,\,0:0}^{(0)}$ and $\mathbf{g}_{p,\,0:0}^{(L+K)}$ can be uniquely identified (within a scale) from (22) as the null eigenvector of $\mathcal{X}^{(0,\,L+K)}$.

For $P=2$, it follows from Theorem 1, that if $L_1>L_2$, only a single SVD (for (22)) solves the blind equalization problem. Two questions arise at this point: When does $L_1>L_2$ hold in practice, and what if $L_1=L_2$? Condition $L_1>L_2$ requires memory domination of the linear part, which is expected in some practical cases. In magnetic recording applications, we have $L_1=L_2=L=2$ or 3, but $s(n)=0,1$; hence, $s(n)=s^2(n)$, which allows us to combine the quadratic kernel $\mathbf{h}_{2,\,0:0}(l)$ (of order L) with the linear one $\mathbf{h}_1(l)$, leaving the remaining kernels $\mathbf{h}_{2,\,i_1:i_1}(l)$ with orders $L-i_1$, $i_1\in[1,\,L]$. In this case too, Theorem 1 applies because there is a single kernel attaining the maximum order L.

Now consider the situation when the linear kernel is absent (i.e., homogeneous model) for a second-order Volterra channel. We again find $\dim\{\mathcal{N}[\mathcal{X}_{(0,\,L+K)}]\}=1$, and the equalizer $\mathbf{g}_{2,\,0:0}^{(0)}$ is uniquely identifiable although of limited value since its output $s^2(n)$ can be used to recover $s(n)$ only when the sign ambiguity is not a problem (e.g., when there is no 180° phase shift ambiguity in the input sequence).

Upon defining

$$\mathbf{X}^{(d)}:=\mathbf{X}(d+1:N-K,\,:)$$
$$\mathbf{X}^{(0,\,d)}:=\mathbf{X}(1:N-K-d,\,:)$$

and $\mathcal{X}^{(0,\,d)}:=[\mathbf{X}^{(d)}\;-\mathbf{X}^{(0,\,d)}]$, we infer from (22)

$$\mathcal{X}^{(0,\,d)}\boldsymbol{g}_1^{(0,\,d)}=[\mathbf{X}^{(d)}\quad -\mathbf{X}^{(0,\,d)}]\begin{bmatrix}\mathbf{g}_1^{(0)}\\ \mathbf{g}_1^{(d)}\end{bmatrix}=\mathbf{0}.$$

It is possible to collect all pairs of equalizers corresponding to all possible shifts in a vector and solve

$$\begin{bmatrix}\mathbf{X}^{(1)} & -\mathbf{X}^{(0,\,1)} & \cdots & \mathbf{0}\\ \vdots & \vdots & \vdots & \vdots\\ \mathbf{X}^{(L+K)} & \mathbf{0} & \cdots & -\mathbf{X}^{(0,\,L+K)}\end{bmatrix}\cdot\begin{bmatrix}\mathbf{g}_1^{(0)}\\ \vdots\\ \mathbf{g}_1^{(L+K)}\end{bmatrix}=\mathbf{0}. \tag{25}$$

Under the assumptions of Theorem 1, it can also be shown that (25) has a unique solution. Simulations have shown that in terms of accuracy of estimates, both estimation procedures (22) and (25) have similar performance. In terms of computational effort, solving (22) requires less flops. In additi38

, the same p.e. condition for the input sequence $s(n)$ is required by both approaches.

B. Many Kernels with Maximum Memory

In this subsection, we treat the case when $\dim\{\mathcal{N}[\mathcal{X}^{(0,\,L+K)}\}>1$, first for the case of second-order nonlinear channels, and then for nonlinearities of arbitrary order. We will see that complexity increases, but we can still identify multichannel FIR equalizers, provided an additional p.e. condition, namely (a2.2), is also satisfied by the input signal $s(n)$. Henceforth, w.l.o.g. we suppose that: $L_1=L_2=\cdots=L_P=L$. Using (a2.1) and following the same steps used to derive (22), we find that the pairs of equalizers $\boldsymbol{g}_p^{(0,\,L+K)}$, $p=1,\cdots,P$, also, satisfy (22); hence, $\dim\{\mathcal{N}[\mathcal{X}^{(0,\,L+K)}]\}=P$. Suppose that the null space $\mathcal{N}[\mathcal{X}^{(0,\,L+K)}]$ is spanned by the columns of the $2M(K+1)\times P$ matrix $\mathbf{U}$, which can be easily obtained by performing an SVD on $\mathcal{X}^{(0,\,L+K)}$. Considering the SVD $\mathcal{X}^{(0,\,L+K)}=\mathbf{V}_1\boldsymbol{\Sigma}\mathbf{V}_2'$, matrix $\mathbf{U}$ is given by the columns of $\mathbf{V}_2$ corresponding to the null singular values [9, p. 18]. Also, consider that the pairs of equalizers $\boldsymbol{g}_p^{(0,\,L+K)}$, $p=1,\cdots,P$, are given by the columns of the matrix

$$\mathbf{G}=[\boldsymbol{g}_1^{(0,\,L+K)}\quad \boldsymbol{g}_2^{(0,L+K)}\;\cdots\;\boldsymbol{g}_P^{(0,\,L+K)}].$$

Based on (a2.1), it follows from (22) that $\mathcal{R}(\mathbf{G})=\mathcal{R}(\mathbf{U})=\mathcal{N}[\mathcal{X}^{(0,\,L+K)}]$, where $\mathcal{R}$ denotes the range space of a matrix. Since $\mathbf{G}$ and $\mathbf{U}$ are full column rank and both span the same space, there exists a nonsingular $P\times P$ matrix $\boldsymbol{\Lambda}$ such that

$$\mathbf{G}=\mathbf{U}\,\boldsymbol{\Lambda}. \tag{26}$$

Considering only the first $M(K+1)$ rows of (26), we obtain

$$\mathbf{G}^{(0)}=\mathbf{U}[1:M(K+1),\,:]\boldsymbol{\Lambda} \tag{27}$$

where

$$\mathbf{G}^{(0)}=[\mathbf{g}_1^{(0)}\quad \mathbf{g}_{2,\,0:0}^{(0)}\;\cdots\;\mathbf{g}_{P,\,0:0}^{(0)}].$$

Since $\mathbf{U}$ is available from the data matrix $\mathcal{X}^{(0,\,L+K)}$, our goal is to identify $\boldsymbol{\Lambda}$. Identification of the pth column $\boldsymbol{\Lambda}(:,\,p)$ yields the pth column of $\mathbf{G}^{(0)}$ (see (27)); i.e., the pth-order equalizer $\mathbf{g}_{p,\,0:0}^{(0)}$. In order to find $\boldsymbol{\Lambda}$, we take into account the dependence between the outputs of the equalizers corresponding to the first and pth-order kernels. Equality $[s(n)]^p=s^p(n)$ can be rewritten as

$$\left[\sum_{k=0}^{K}\mathbf{x}'(n-k)\mathbf{g}_1^{(0)}(k)\right]^p=\sum_{k=0}^{K}\mathbf{x}'(n-k)\mathbf{g}_{p,\,0:0}^{(0)}(k)$$

or, equivalently,

$$[\mathbf{X}(N-n,\,:)\mathbf{g}_1^{(0)}]^p=\mathbf{X}(N-n,\,:)\mathbf{g}_{p,\,0:0}^{(0)}. \tag{28}$$

Using (10) and (27), we can rewrite (28) as

$$\begin{aligned}&\{\mathbf{S}(N-n,\,:)\mathbf{H}\,\mathbf{U}[1:M(K+1),\,:]\,\boldsymbol{\Lambda}(:,\,1)\}^p\\ &\quad=\mathbf{S}(N-n,\,:)\mathbf{H}\,\mathbf{U}[1:M(K+1),\,:]\,\boldsymbol{\Lambda}(:,\,p).\end{aligned} \tag{29}$$

Denote the k-fold Kronecker product of a matrix $\mathbf{W}$ with itself by $\mathbf{W}^{[k]}=\mathbf{W}\otimes\cdots\otimes\mathbf{W}$, and recall that $(\mathbf{W}_1\;\mathbf{W}_2)^{[k]}=\mathbf{W}_1^{[k]}\;\mathbf{W}_2^{[k]}$. Considering relation (29) for all $n=N-1,\cdots,K$, we obtain

$$\mathbf{A}_p\,\boldsymbol{\Lambda}^{[p]}(:,\,1)=\mathbf{B}\,\boldsymbol{\Lambda}(:,\,p) \tag{30}$$

where $\mathbf{A}_p$ and $\mathbf{B}$ are defined, respectively, as

$$\mathbf{A}_p := \begin{bmatrix} \mathbf{S}^{[p]}(1,:) \\ \mathbf{S}^{[p]}(2,:) \\ \vdots \\ \mathbf{S}^{[p]}(N-K,:) \end{bmatrix} \cdot \mathbf{H}^{[p]}\mathbf{U}[1:M(K+1),:]^{[p]} = \begin{bmatrix} \mathbf{X}^{[p]}(1,:) \\ \mathbf{X}^{[p]}(2,:) \\ \vdots \\ \mathbf{X}^{[p]}(N-K,:) \end{bmatrix} \cdot \mathbf{U}[1:M(K+1),:]^{[p]} \tag{31}$$

$$\mathbf{B} := \begin{bmatrix} \mathbf{S}(1,:) \\ \mathbf{S}(2,:) \\ \vdots \\ \mathbf{S}(N-K,:) \end{bmatrix} \cdot \mathbf{H}\mathbf{U}[1:M(K+1),:] = \begin{bmatrix} \mathbf{X}(1,:) \\ \mathbf{X}(2,:) \\ \vdots \\ \mathbf{X}(N-K,:) \end{bmatrix} \cdot \mathbf{U}[1:M(K+1),:] \tag{32}$$

where in deriving the second equalities in (31) and (32), we used (10) and the previously mentioned property of the Kronecker product. Since matrices $\mathbf{U}$, $\mathbf{A}_p$, $\mathbf{B}$ can be obtained from the data, we will show subsequently how to identify $\mathbf{\Lambda}$ from (30). From (27) we have $\mathbf{U}[1:M(K+1),:] = \mathbf{G}^{(0)}\,\mathbf{\Delta}$, where $\mathbf{\Delta} = \mathbf{\Lambda}^{-1}$. Note that the first row of $\mathbf{A}_p$ can be rewritten as

$$\mathbf{A}_p(1,:) = \{\mathbf{S}(1,:)\mathbf{H}\mathbf{U}_{[1:M(K+1),:]}\}^{[p]} = [\mathbf{S}(1,:)\mathbf{H}\mathbf{G}^{(0)}\,\mathbf{\Delta}]^{[p]}.$$

Also note that $\mathbf{S}(1,:)\mathbf{H}\ \mathbf{G}^{(0)} = \mathbf{s}_{1p}(N-1)$, where by definition

$$\mathbf{s}_{1p}(n) := [s(n)\ s^2(n)\ \cdots\ s^p(n)] \qquad \forall n.$$

It turns out that $\mathbf{A}_p(1,:) = [\mathbf{s}_{1p}(N-1)\,\mathbf{\Delta}]^{[p]} = \mathbf{s}_{1p}^{[p]}(N-1)\,\mathbf{\Delta}^{[p]}$. Similar decompositions can be obtained for the other rows of $\mathbf{A}_p$. So we have

$$\mathbf{A}_p = \mathbf{S}_p\,\mathbf{\Delta}^{[p]} \tag{33}$$

where

$$\mathbf{S}_p := \begin{bmatrix} \mathbf{s}_{1p}^{[p]}(N-1) \\ \mathbf{s}_{1p}^{[p]}(N-2) \\ \vdots \\ \mathbf{s}_{1p}^{[p]}(K) \end{bmatrix}.$$

Similarly, we factorize $\mathbf{B}$ as

$$\mathbf{B} = \mathbf{S_1\Delta} \tag{34}$$

where

$$\mathbf{S}_1 := \begin{bmatrix} \mathbf{s}_{1p}(N-1) \\ \mathbf{s}_{1p}(N-2) \\ \vdots \\ \mathbf{s}_{1p}(K) \end{bmatrix}.$$

Factorizations (33) and (34) will be useful in the ensuing subsections where we propose two methods for establishing identifiability of $\mathbf{\Lambda}$ from (30). Our first method applies only to linear-quadratic channels and is presented next.

1) Second-Order Nonlinearities: For $p = P = 2$, (30) can be rewritten as

$$[\mathbf{B}\quad \mathbf{A}_2]\begin{bmatrix} \mathbf{\Lambda}(:,2) \\ -\mathbf{\Lambda}^{[2]}(:,1) \end{bmatrix} = \mathbf{0}. \tag{35}$$

Matrix $[\mathbf{B}\ \mathbf{A}_2]$ is rank deficient because the Kronecker product has introduced redundant entries that must be removed before solving (35) for $\mathbf{\Lambda}$. Define as $\hat{\mathbf{A}}_2$ the matrix obtained from $\mathbf{A}_2$ by removing its third column (i.e., by eliminating the redundancies from $\mathbf{A}_2$ introduced by the Kronecker product), and similarly, let $\hat{\mathbf{\Lambda}}^{[2]}(:,1)$ denote the vector obtained from $\mathbf{\Lambda}^{[2]}(:,1)$ by removing its third entry. An equivalent form for (35) is

$$\mathbf{D}\begin{bmatrix} \mathbf{\Lambda}(:,2) \\ -\hat{\mathbf{\Lambda}}^{[2]}(:,1) \end{bmatrix} = \mathbf{0} \tag{36}$$

where $\mathbf{D} := [\mathbf{B}\ \hat{\mathbf{A}}_2]$. To establish identifiability, we wish to show that $\dim[\mathcal{N}(\mathbf{D})] = 1$. Considering (33) with $p = 2$, it follows that $\hat{\mathbf{A}}_2$ can be factorized as

$$\hat{\mathbf{A}}_2 = \begin{bmatrix} s^2(N-1) & s^3(N-1) & s^4(N-1) \\ \vdots & \vdots & \vdots \\ s^2(K) & s^3(K) & s^4(K) \end{bmatrix} \times \begin{bmatrix} \delta_{11}^2 & \delta_{11}\,\delta_{12} & \delta_{12}^2 \\ 2\,\delta_{11}\,\delta_{21} & \delta_{11}\,\delta_{22} + \delta_{12}\,\delta_{21} & 2\,\delta_{12}\,\delta_{22} \\ \delta_{21}^2 & \delta_{21}\,\delta_{22} & \delta_{22}^2 \end{bmatrix} \tag{37}$$

with $\mathbf{\Delta} = [\delta_{ij}], 1 \le i, j \le 2$. From (34) and (37), we have the following factorization for $\mathbf{D}$:

$$\mathbf{D} = \overline{\mathbf{S}}\ \overline{\mathbf{\Delta}} \tag{38}$$

where $\overline{\mathbf{S}}$ is the Vandermonde matrix defined in (15) for $P = 2$, and $\overline{\mathbf{\Delta}}$ is given by

$$\overline{\mathbf{\Delta}} := \begin{bmatrix} \delta_{11} & \delta_{12} & 0 & 0 & 0 \\ \delta_{21} & \delta_{22} & \delta_{11}^2 & \delta_{11}\,\delta_{12} & \delta_{12}^2 \\ 0 & 0 & 2\,\delta_{11}\,\delta_{21} & \delta_{11}\,\delta_{22} + \delta_{12}\,\delta_{21} & 2\,\delta_{12}\,\delta_{22} \\ 0 & 0 & \delta_{21}^2 & \delta_{21}\,\delta_{22} & \delta_{22}^2 \end{bmatrix}.$$

Since $\mathbf{\Delta}$ is an invertible matrix, it follows easily that $\operatorname{rank}(\overline{\mathbf{\Delta}}) = 4$. Also, since $\overline{\mathbf{S}}$ is full-column rank by (a2.2), it follows from (38) that $\dim[\mathcal{R}(\mathbf{D})] = 4$. Hence, $\dim[\mathcal{N}(\mathbf{D})] = 1$. The latter implies that $[\mathbf{\Lambda}'(:,2)\ -\hat{\mathbf{\Lambda}}'^{[2]}(:,1)]$ can be uniquely determined as the null eigenvector of $\mathbf{D}$. Having determined $\mathbf{\Lambda}$, we can go back to (27) to recover the desired matrix of equalizers $\mathbf{G}^{(0)}$.

For $P > 2$, a similar construction leads to $\dim[\mathcal{N}(\mathbf{D})] > 1$, which does not allow unique determination of $[\mathbf{\Lambda}'(:,P)\ -\hat{\mathbf{\Lambda}}'^{[P]}(:,1)]$ from (36). In what follows, we show a general solution for determining $\mathbf{\Lambda}$, which is valid for nonlinearities of any order (including the case $P = 2$).

C. Nonlinearities of Arbitrary Order

In this subsection, we will determine first the Pth-order equalizer (i.e., $\mathbf{G}^{(0)}(:, P)$) which, according to (27), requires identification of $\mathbf{\Lambda}(:, P)$. The latter will be recovered from (30) as the intersection of the range spaces of $\mathbf{A}_P$ and $\mathbf{B}$. To show this, consider the notation

$$\tilde{\mathbf{s}}_p := [s^p(N-1) \cdots s^p(K)]', \quad p = 1, \cdots, P. \tag{39}$$

We establish first a relation between the range spaces of matrices $\mathbf{A}_p$ and $\mathbf{B}$ in (33) and (34).

Proposition 4: Under (a0.1)–(a2.2), the following relations hold:

$$\dim[\mathcal{R}(\mathbf{A}_p) \cap \mathcal{R}(\mathbf{B})] = P - p + 1,$$
$$\mathcal{R}(\mathbf{A}_p) \cap \mathcal{R}(\mathbf{B}) = \text{span}[\tilde{\mathbf{s}}_p, \cdots, \tilde{\mathbf{s}}_P].$$

Proof: Since $\mathbf{\Delta}$ is full rank, it follows from (33) and (34) that $\mathcal{R}(\mathbf{A}_p) = \mathcal{R}(\mathbf{S}_p)$, and $\mathcal{R}(\mathbf{B}) = \mathcal{R}(\mathbf{S}_1)$. From (a2.2), $\overline{\mathbf{S}}$ is full column rank. Hence, $\dim[\mathcal{R}(\mathbf{S}_p) \cap \mathcal{R}(\mathbf{S}_1)] = P - p + 1$, and this intersection is spanned by the vectors $\tilde{\mathbf{s}}_p, \cdots, \tilde{\mathbf{s}}_P$, which establishes Proposition 4. □

For $p = P$, we have from Proposition 4:

$$\mathcal{R}(\mathbf{A}_P) \cap \mathcal{R}(\mathbf{B}) = \text{span}[\tilde{\mathbf{s}}_P]. \tag{40}$$

From (40), an approach for determining $\mathbf{\Lambda}^{[P]}(:, 1)$ and $\mathbf{\Lambda}(:, P)$ can be derived. The steps are listed below.

Step 1) Find the common vector $\mathbf{r}_{A_P, B}$ that spans $\mathcal{R}(\mathbf{A}_P) \cap \mathcal{R}(\mathbf{B})$. This step requires two QR-factorizations (for finding two orthonormal basis for $\mathcal{R}(\mathbf{A}_P)$ and $\mathcal{R}(\mathbf{B})$) and one SVD (for determining the common vector, see [9, pp. 429–430], [5], and [32]).

Step 2) Solve $\mathbf{B\Lambda}(:, P) = \mathbf{r}_{A_P, B}$ for $\mathbf{\Lambda}(:, P)$, and use (27) to compute the equalizer corresponding to the Pth-order kernel using $\mathbf{G}^{(0)}(:, P) = \mathbf{U}(1 : M(K+1), :)\mathbf{\Lambda}(:, P)$.

Step 3) Solve equation

$$\mathbf{A}_P \mathbf{\Lambda}^{[P]}(:, 1) = \mathbf{r}_{A_P, B} \tag{41}$$

for $\mathbf{\Lambda}^{[P]}(:, 1)$. In general, matrix $\mathbf{A}_P$ in (41) is rank deficient, even after the elimination of the redundancies introduced by the Kronecker product. A way to overcome this is to consider the equation obtained by taking the Pth root of both sides of (41). This alternative is presented in the next step.

Step 4) Solve $\mathbf{B\Lambda}(:, 1) = \mathbf{r}_{A_P, B}^{1/P}$, for $\mathbf{\Lambda}(:, 1)$. This equation can be solved as a standard LS equation. The solution will be correct provided that there is no ambiguity in taking the Pth root of $\mathbf{r}_{A_P, B}$. Having computed $\mathbf{\Lambda}(:, 1)$, we obtain $\mathbf{g}_1^{(0)}$ by using (27).

We have thus established the following result.

Theorem 2: If (a0.1)–(a2.2) holds true, then the equalizers corresponding to the Pth-order kernel and to all possible delays can be uniquely identified (within a scale factor).

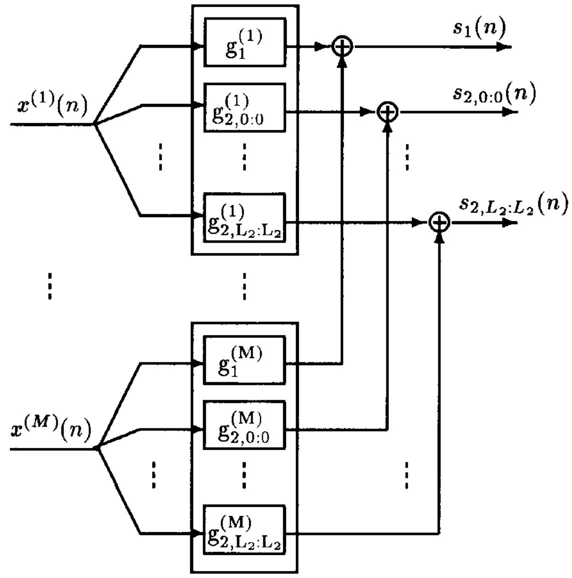

Fig. 3. Structure of the deconvolver ($P = 2$).

Proof: From the above derivations, it follows that $\mathbf{g}_{P,0:0}^{(0)}$ can be uniquely determined. Using a similar relation to (20) for Pth-order equalizers, it follows that we can retrieve the equalizers corresponding to the Pth-order kernel and to all possible delays. □

Note also that Step 4 determines the equalizer $\mathbf{g}_1^{(0)}$ provided that there is no ambiguity in taking the Pth-root of $\mathbf{r}_{A_P, B}(= c\tilde{\mathbf{s}}_P$, where c denotes a constant); i.e., no ambiguity exists in recovering uniquely $s(n)$ from $s^P(n)$. Such unambiguous recovery is guaranteed, for example, in the case of a PSK input signal with Q constellation points such that (Q, P) are coprime; $s(n)$ can be uniquely recovered from $s^P(n)$ since a rotation of $s(n)$ by a factor P, brings the constellation points of $s^P(n)$ at distinct locations. Assuming that we can find uniquely $s(n)$ from $s^P(n)$, the determination of all other equalizers is possible since the problem reduces to an I/O identification problem. However, there are cases when $s(n)$ cannot be recovered uniquely from $s^P(n)$; e.g., a PSK input signal with (Q, P) not coprime.

In this case, we proceed to recover uniquely (up to a constant) the equalizer $\mathbf{g}_{P-1,0:0}^{(0)}$. Knowledge of equalizers corresponding to the $(P-1)$st and Pth-order kernels implies access to their outputs $s^{P-1}(n)$, and, respectively, $s^P(n)$. Selecting samples $s(n) \neq 0$ and taking the ratio of these two outputs, we obtain the input sequence $s(n) = s^P(n)/s^{P-1}(n)$. Hence, the problem is solved if we can find $\mathbf{g}_{P-1,0:0}^{(0)}$. Toward this goal, we apply again Proposition 4 for $p = P - 1$, to infer that

$$\mathcal{R}(\mathbf{A}_{P-1}) \cap \mathcal{R}(\mathbf{B}) = \text{span}[\tilde{\mathbf{s}}_{P-1}, \tilde{\mathbf{s}}_P].$$

Consider $[\mathbf{r}_{A_{P-1},B}^{(1)}, \mathbf{r}_{A_{P-1},B}^{(2)}]$ to be a basis for $\mathcal{R}(\mathbf{A}_{P-1}) \cap \mathcal{R}(\mathbf{B})$. It follows that there exists a 2×2 nonsingular matrix $\mathbf{F} = [f_{ij}]$ such that

$$[\mathbf{r}_{A_{P-1},B}^{(1)}, \mathbf{r}_{A_{P-1},B}^{(2)}] = [\tilde{\mathbf{s}}_{P-1}, \tilde{\mathbf{s}}_P]\mathbf{F}. \tag{42}$$

Define the vectors $\hat{\mathbf{g}}_{P-1,\,0:0}^{(0)}$ and $\hat{\hat{\mathbf{g}}}_{P-1,\,0:0}^{(0)}$ through the relations

$$\mathbf{X}\hat{\mathbf{g}}_{P-1,\,0:0}^{(0)} = \mathbf{r}_{A_{P-1},\,B}^{(1)} \tag{43}$$

$$\mathbf{X}\hat{\hat{\mathbf{g}}}_{P-1,\,0:0}^{(0)} = \mathbf{r}_{A_{P-1},\,B}^{(2)}. \tag{44}$$

Using the relation $[s^{P-1}(n)]^P = [s^P(n)]^{P-1}$ between the outputs of equalizers $\mathbf{g}_{P-1,\,0:0}^{(0)}$ and $\mathbf{g}_{P,\,0:0}^{(0)}$, it is shown in Appendix C that the equalizer $\mathbf{g}_{P-1,\,0:0}^{(0)}$ can be found as a linear combination of $\hat{\mathbf{g}}_{P-1,\,0:0}^{(0)}$ and $\hat{\hat{\mathbf{g}}}_{P-1,\,0:0}^{(0)}$.

By equalizing the channel with $\mathbf{g}_{P-1,\,0:0}^{(0)}$, and, respectively, $\mathbf{g}_{P,\,0:0}^{(0)}$, we obtain sequences $s^{P-1}(n)$ and $s^P(n)$, which uniquely identify $s(n)\ \forall\ n$. We summarize these results in the following theorem.

Theorem 3: If (a0.1)–(a2.2) hold true, then the equalizers corresponding to all kernels and all possible delays can be uniquely identified (within a scale factor).

D. Noisy Case and Kernel Identification

Now we consider briefly the noisy case together with the blind identification of Volterra kernels. Assume that the AWGN $\mathbf{v}(n)$ is present in (4). For high SNR's, the entire analysis carries over here *mutatis mutandis*. In this case, only slight modifications have to be adopted in order to establish identifiability and, thus, feasibility of the algorithms developed. $\mathcal{R}[\mathbf{G}^{(0)}]$ has to be estimated from the noisy null space of $\mathcal{X}^{(0,\,L+K)}$, by considering the SVD of $\mathcal{X}^{(0,\,L+K)}$. Similarly, a basis for $\mathcal{R}(\mathbf{A}_p) \cap \mathcal{R}(\mathbf{B})$, for $p = P-1,\, P$, is obtained by considering the pairs of principal vectors corresponding to the smallest angles between $\mathcal{R}(\mathbf{A}_p)$ and $\mathcal{R}(\mathbf{B})$ [5]. For low SNR, the performance of the proposed algorithm may be significantly affected by noise, since the proposed method does not take into account the noise statistics. Also, the computation of a basis for $\mathcal{R}(\mathbf{A}_p) \cap \mathcal{R}(\mathbf{B})$ may be very sensitive to noise perturbations. In this case, a suboptimal solution may consist in using a combination of the proposed method with a subspace-based method [23], [30]. Using the present approach an estimate of $\mathbf{G}^{(0)}$ can be found. With a subspace-based method, we can estimate $\mathcal{R}(\mathbf{H}')$. From an estimate for $\mathbf{G}^{(0)}$, an estimate for $\mathbf{h}_{p,\,i_1:i_p}(0)$, $p = 1, \cdots, P$, can be obtained. The rest of channel coefficients $\mathbf{h}_{p,\,i_1:i_p}(k)$, $p = 1, \cdots, P$, $k = 1, \cdots, L$ can be obtained using a subspace fitting approach [30].

For the proposed method, once all the equalizers corresponding to all possible delays are available, we can align their outputs and average them in order to obtain an averaged estimate of the input via (cf., (8))

$$s(n) = \frac{1}{L+K+1}\sum_{d=0}^{L+K}\left[\sum_{k=0}^{K}\mathbf{x}'(n+d-k)\,\mathbf{g}_1^{(d)}(k)\right].$$

Averaging may improve the equalization performance but thorough analysis is due before definitive conclusions can be reached.

If blind channel identification is the objective, the estimated equalizers can be used to recover $s(n)$, from which $\mathbf{H}$ can be obtained by solving (7), using (batch or recursive) linear regression methods or by simply inverting the matrix whose columns are the equalizers, since $\mathbf{H}\ \mathbf{g}_1^{(d)} = \mathbf{e}_{d+1}$ (see (16)). Finally, note that the equalizer can be implemented as a set of linear FIR filter banks. Fig. 3 depicts the structure of the multichannel equalizer for a second-order system. This structure of the multichannel deconvolver is derived by considering a graphical interpretation of (8) and (9). In order not to complicate the notation, only the equalizers of zero delay are presented, and the superscript for each equalizer refers to the channel to which the equalizer is associated with and not to the delay. The zero delay is not represented in Fig. 3.

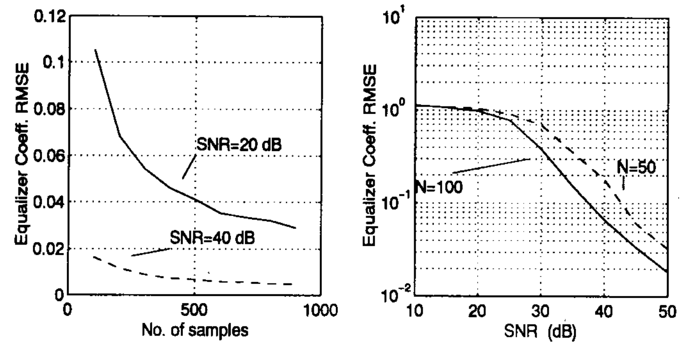

Fig. 4. RMSE curves for Example 1.

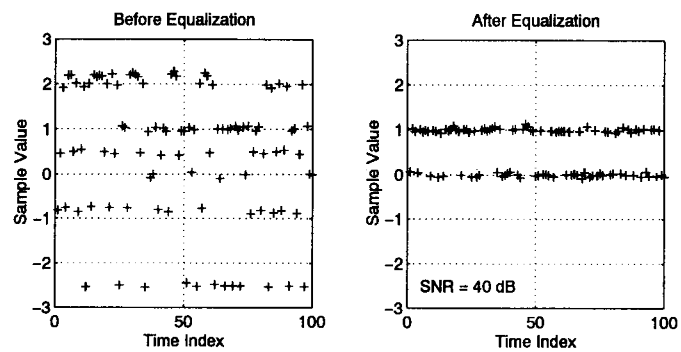

Fig. 5. Eye-patterns for Example 1.

V. Simulations

To illustrate the proposed algorithms and study their performance in noise, we resorted to simulations summarized in the following five examples.

Example 1. Blind Equalization of a Magnetic Recording Channel with $\dim\{\mathcal{N}[\mathcal{X}^{(0,\,L+K)}]\} = 1$*:* We generated two-level pulse amplitude modulated (PAM) independent identically distributed (i.i.d.) data ($s(n) = 0,\, 1$) and passed them through $M = 3$ FIR channels ($m = 1,\, 2,\, 3$) to obtain the data

$$x^{(m)}(n) = \sum_{l=0}^{2} h_1^{(m)}(l)s(n-l) + \sum_{l=0}^{1} h_{2,\,1}^{(m)}(l) \cdot s(n-1-l)s(n-l) + v^{(m)}(n).$$

The impulse response vectors were $\mathbf{h}_1(0) = [1,\, 0.5,\, 2]'$, $\mathbf{h}_1(1) = [-2.5,\, 3,\, 0]'$, $\mathbf{h}_1(2) = [1,\, 5,\, 2]'$, $\mathbf{h}_{2,\,1}(0) = [2,\, 0.3,\, -0.7]'$, $\mathbf{h}_{2,\,1}(1) = [0.7,\, 1.2,\, 3]'$. Such a channel has form similar to that used in magnetic recording models [3], [11]. Theorem 1 applies to this channel ($L_1 = 2 > L_{2,\,1} = 1$),

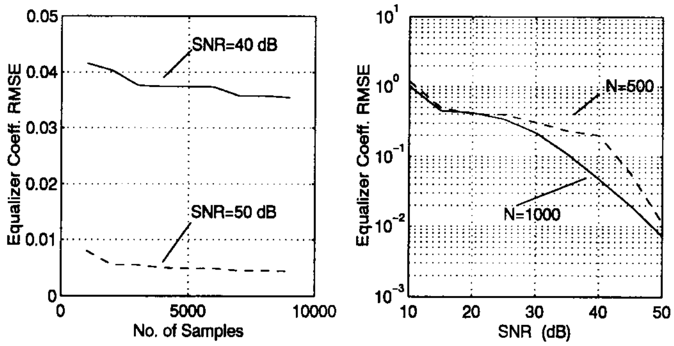

Fig. 6. RMSE curves for Example 2.

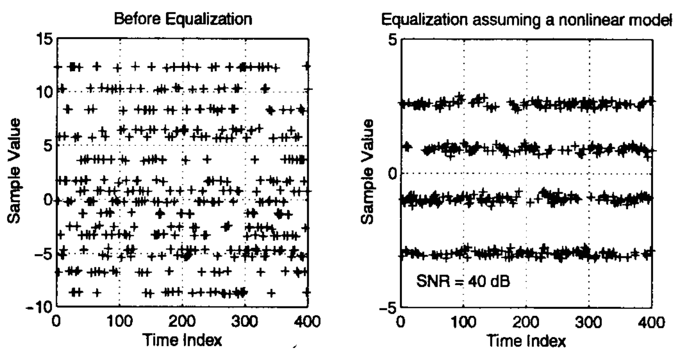

Fig. 7. Eye-patterns for Example 2.

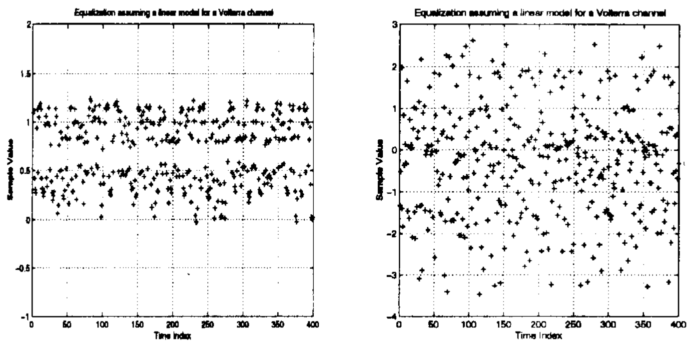

Fig. 8. Eye-patterns—linear approximations.

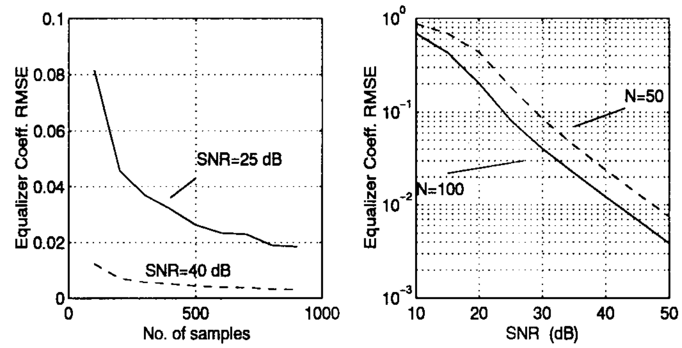

Fig. 9. RMSE curves for Example 3.

and using one SVD, we computed the 3×1 vector equalizer of order $K = 2$ by solving (22) with $d = L_1 + K$. According to Theorem 1, note that Assumption (a2.2) is not necessary for this example. Fig. 4(a) depicts root mean-square error (RMSE) between the true and estimated equalizer coefficients for lengths $N = 100, \cdots, 900$ at SNR $= 20$ dB and 40 dB; RMSE versus SNR is shown in Fig. 4(b) for $N = 50, 100$ (averages were computed based on 100 Monte Carlo runs). Interestingly, with as little as $N = 200$ symbols, it is possible to equalize linear-quadratic channels with RMSE $= O(10^{-2})$ at SNR $= 20$ dB. A typical eye-diagram of one channel's output is plotted in Fig. 5(a) along with its equalized version in Fig. 5(b).

Example 2. Blind Equalization of a Real Channel with $\dim\{\mathcal{N}[\mathcal{X}^{(0,\, L+K)}]\} = 2$*:* A similar simulation was carried with four-level PAM data [$s(n) = \pm 3, \pm 1$] and $M = 4$ channel outputs were generated according to the model ($m = 1, 2, 3, 4$)

$$x^{(m)}(n) = \sum_{l=0}^{1} h_1^{(m)}(l) s(n-l) + \sum_{l=0}^{1} h_{2,0}^{(m)}(l) s^2(n-l) + h_{2,1}^{(m)}(0) s(n-1) s(n) + v^{(m)}(n)$$

with the kernels $\mathbf{h}_1(0) = [1, 0.5, 2, 0.1]'$, $\mathbf{h}_1(1) = [-2.5, 3, 0, -1.1]'$, $\mathbf{h}_{2,0}(0) = [0.01, 0.5, 0.2, 0.003]'$, $\mathbf{h}_{2,0}(1) = [0.2, 0.3, -0.7, -0.001]'$, $\mathbf{h}_{2,1}(0) = [0.007, 0.001, 0.3, -0.15]'$. Figs. 6 and 7 show that about an order of magnitude more data are required to achieve performance similar to that in Figs. 4 and 5, a consequence of the fact that two SVD's are required for this model (note that here $L_1 = L_{2,0} = 1$, $L_{2,1} = 0$, $K = 1$). To illustrate the importance of incorporating nonlinearities over adopting linear approximations, we supposed that the data come from a linear channel of order $L = 3$, and using $M = 2$ outputs we designed an order $K = 2$ linear equalizer by inverting the channel estimate of [33]. The equalized eye-patterns for the two- and four-level PAM data are shown in Fig. 8. The importance of adopting the correct model is evident if one compares Figs. 5 and 7 with Fig. 8(a) and, respectively, Fig. 8(b). In both simulations, the SNR was 40 dB. We also tested the procedure for order selection. By choosing $\overline{K}_1 = 12$, $\overline{K}_2 = 14$, and an appropriate threshold for selecting the dominant singular values, we found $\rho_{\overline{X}_1} = 41$, $\rho_{\overline{X}_2} = 47$ from where we correctly estimated $L_1 = L_2 = 1$.

Example 3. Blind Equalization of a Baseband Complex Channel with $\dim\{\mathcal{N}[\mathcal{X}^{(0,\, L+K)}]\} = 1$*:* In this case, the simulation was carried with a quadrature PSK (QPSK) input signal and with $M = 3$ channels. The channels included only odd-order kernels. This model belongs to the class of narrowband Volterra channels that has only odd order kernels [2, p. 58]. The mth channel is given by

$$x^{(m)}(n) = \sum_{l=0}^{3} h_1^{(m)}(l) x(n-l) + v^{(m)}(n) + \sum_{l=0}^{1} \cdot h_{3,12}^{(m)} x(n-l) x(n-l-1) x^{\star}(n-l-2)$$

with $\mathbf{h}_1(0) = [1+i, 0.5+0.4i, -1+i]'$, $\mathbf{h}_1(1) = [-2.5+2i, 3+2i, 1-2i]'$, $\mathbf{h}_1(2) = [1+i, -1+i, 2+1.3i]'$, $\mathbf{h}_1(3) = [4+0.3i, 5+i, -3+1.3i]'$, $\mathbf{h}_{3,1:2}(0) = [2, 0.3+0.2i, -0.7+0.7i]'$, $\mathbf{h}_{3,1:2}(1) = [0.7-0.8i, 1.2+i, 3+0.1i]'$, and $\star$ representing the complex conjugation operation. Again only one SVD is sufficient in order to determine all the equalizers

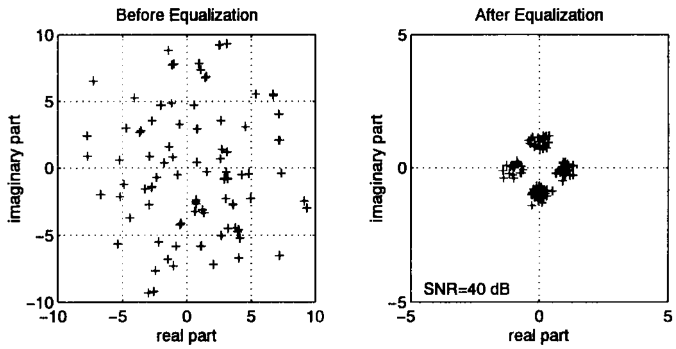

Fig. 10. Eye patterns for Example 3.

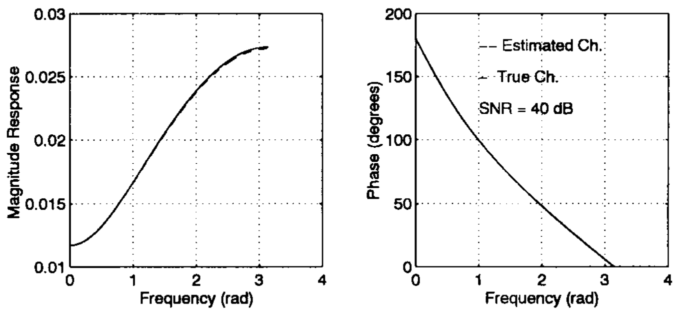

Fig. 11. Linear kernels for Example 4.

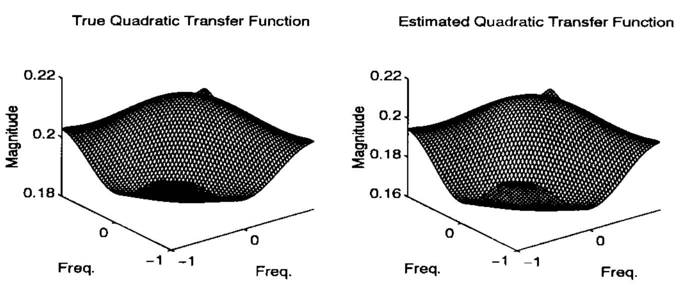

Fig. 12. Quadratic kernels for Example 4.

since $L_1 = 3 > L_{3,1:2} = 1$. In Figs. 9 and 10, the performance of the $\mathbf{g}_1^{(0)}$ equalizer with order $K = 3$ is shown. As in the above examples, the rmse values were computed using 100 Monte Carlo simulations.

Example 4. Blind Identification of a Real Channel with $\dim\{\mathcal{N}[\mathcal{X}^{(0,\,L+K)}]\} = 2$*:* The same channel as in Example 2 was used in this experiment, the only difference being that instead of a 2-PAM signal we used a pseudorandom sequence with normal distribution $N(0,\,1)$. We wanted to study the capability of the present approach to identify the channel. We performed the study for different SNR's. For relatively high SNR's (e.g., 40 dB) the estimated frequency responses for the linear/quadratic kernels were almost identical with the true ones (see Figs. 11 and 12). When we decreased the SNR below 30 dB the performance diminished significantly. We used 1000 samples and one Monte Carlo simulation.

Example 5. Blind Equalization of a Real Channel with $\dim\{\mathcal{N}[\mathcal{X}^{(0,\,L+K)}]\} = 3$*:* In this case we considered $M = 4$ channels with $\dim[\mathcal{N}(\mathcal{X}_{0,\,3})] = 3$, and $L_1 = L_{2,\,0} = L_{3,\,0:0} = L = 1$. The mth channel is described by

$$x^{(m)}(n) = \sum_{l=0}^{1} h_1^{(m)}(l)s(n-l) + \sum_{l=0}^{1} h_{2,\,0}^{(m)}(l)s^2(n-l) + \sum_{l=0}^{1} h_{3,\,0:0}^{(m)}(l)s^3(n-l) + v^{(m)}(n)$$

where $\mathbf{h}_1(0) = [1, 0.5, 2, -0.3]'$, $\mathbf{h}_1(1) = [-2.5, 3, 0., 0.9]'$, $\mathbf{h}_{2,0}(0) = [0.01, 0.5, 0.2, 1]'$, $\mathbf{h}_{2,0}(1) = [0.2, 0.3, -0.7, 0.02]'$, $\mathbf{h}_{3,0:0}(0) = [-0.6, 0.2, 0.9, 0.76]'$, $\mathbf{h}_{3,0:0}(1) = [-0.9, -0.7, 0.1, 0.02]'$. We first considered a pseudorandom sequence with normal distribution $N(0,\,1.4)$ as an input sequence. Only 500 samples were used to estimate a $K = 2$-order equalizer. A comparison between the true input $s(n)$ and the equalizer output waveforms for SNR = 30 dB, and respectively, SNR = 40 dB is shown in Fig. 13. In Fig. 14, the eye-patterns, when the input signal is a nine-level PAM signal and there is no noise present, are shown. The experiment was repeated with an eight-level PAM signal, but the equalizer failed to work properly. This is justifiable, since the input signal must take at least $P^2 = 9$ nonzero distinct values in order to satisfy the conditions of Theorems 2 and 3. Note that the received signal eye diagram looks completely different from the eye diagram of a nine-level PAM signal, even in the absence of noise, while the eye diagram of the equalized output shows nine distinct levels.

VI. Concluding Remarks

We proposed a linear multichannel equalizer for a nonlinear FIR Volterra channel. The approach required only that the input sequence satisfies a p.e. condition and the channel transfer matrix has full row rank. The equalization of nonlinear channels with linear FIR equalizers is appealing and can be justified intuitively if one views the vector equalizer as a beamformer which, thanks to its diversity, is capable of nulling the nonlinearities and equalizing the linear part.

A number of open questions arise: analytic performance evaluation, comparisons with Cramer–Rao bounds, selection of the optimum equalizer delay, explicit inclusion of the noise along the linear prediction formulation (see for the linear case [30]), and thorough study of determining the structure and order of the Volterra model from output data only. We envision a solution based on the minimization of an Akaike-type criterion by varying the order P in a certain interval, and by estimating L from a relation similar to (18) ($L_1 = \cdots = L_P = L$). The linear equations involved in the derivation of the equalizers suggest an adaptive version of the proposed algorithm, at least for the case when $\dim\{\mathcal{N}[\mathcal{X}^{(0,\,L+K)}]\} = 1$. The computation of a basis for $\mathcal{R}(\mathbf{A}_p) \cap \mathcal{R}(\mathbf{B})$ may be very sensitive to noise perturbations. Especially, for low SNR, a new optimal and general solution that avoids such intersections and takes into account noise statistics is desirable.

The proposed deterministic approach works well in the case of high SNR and requires a reduced number of samples/computational effort in comparison with a higher order statistics based approach [26] and [27]. For communication channels with varying nonlinearities (e.g., in mobile radio

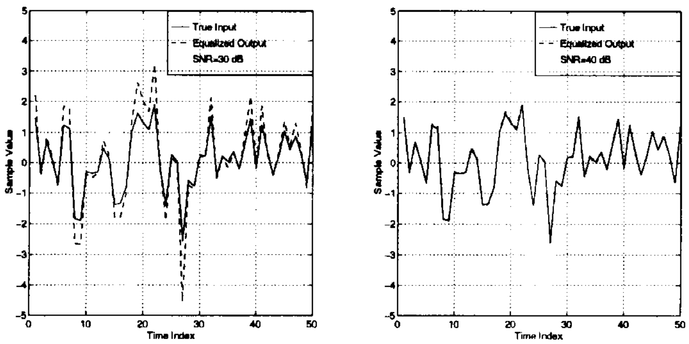

Fig. 13. True and equalized inputs.

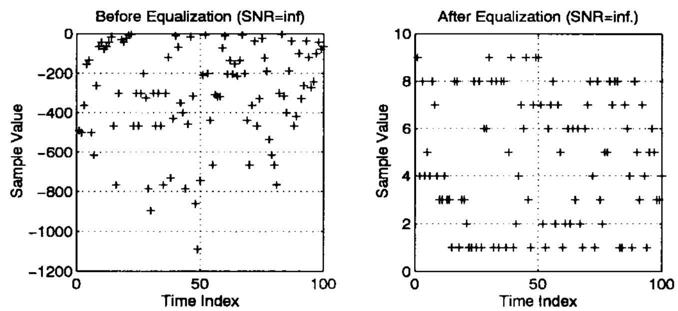

Fig. 14. Eye-patterns for Example 5.

communications), the present approach may exhibit good tracking properties. The complexity of the algorithm is relatively high, which makes difficult the implementation of an efficient on-line version. The computational complexity increases significantly with the order and the memory of the nonlinear channel; for example, for a LTI–ZMNL–LTI channel of order 3 and memory L, the number of antennas M and the equalizer order K have a quadratic and, respectively, cubic dependence on memory L. As in linear case, the issue of how often a real channel satisfies the coprimeness condition is not known.

Appendix A
Characterization of Assumption (a1.1)

An alternative characterization of Assumption (a1.1) is possible. It can be easily shown that an equivalence exists between the full row rank of the block Toeplitz matrix $\mathbf{H}$ and the maximum column rank of a generalized Sylvester resultant $\mathbf{R}_{K+1}$ of two polynomial matrices (see [1] and [4]). The resultant $\mathbf{R}_{K+1}$ is obtained by permuting the columns of $\mathbf{H}'$ and adding a certain number of columns with only zero entries to $\mathbf{H}'$. From [4, Th. 1] a characterization in terms of the dual dynamic indices of a pair of polynomial matrices can be deduced. In general, this characterization does not bring too much, except in the particular case when the channels (associated to the linear MIMO interpretation, i.e., $\mathbf{h}_{p,\,i_1:i_{p-1}}$) have the same length. In this case, no zero column has to be appended to $\mathbf{H}'$ for obtaining the generalized Sylvester resultant $\mathbf{R}_{K+1}$, and (a1.1) reduces to the coprimeness condition of a pair of polynomial matrices.

Appendix B
Characterization of Assumption (a2.1)

We provide a short justification for the result: if the input signal constellation has at least $P+1$ distinct points and N is large enough, then Assumption (a2.1) holds with probability 1. Indeed, if $\mathcal{S}_{rc}^{(0,\,L+K)}$ is the matrix obtained from $\mathcal{S}^{(0,\,L+K)}$ by eliminating C redundant columns (i.e., $\mathcal{S}_{rc}^{(0,\,L+K)}$ does not have any pair of identical columns), then (a2.1) is equivalent to $\mathcal{S}_{rc}^{(0,\,L+K)}$ having full column rank. Suppose $\mathcal{S}_{rc}^{(0,\,L+K)}$ does not have full column rank, then there exists a vector $\mathbf{t}$, $\mathbf{t} \neq \mathbf{0}$, such that

$$\mathcal{S}_{rc}^{(0,\,L+K)}\,\mathbf{t} = \mathbf{0}. \tag{45}$$

Note that the entries of any row of $\mathcal{S}_{rc}^{(0,\,L+K)}$ are distinct and have the form $s_{p,\,i_1:i_{p-1}}(n)$, with $0 \le i_1 \le \cdots \le i_{p-1} \le L_p$, $1 \le p \le P$. Equation $\mathcal{S}_{rc}^{(0,\,L+K)}(N-n,\,:)\mathbf{t} = \mathbf{0}$ can be rewritten as a Pth-order recursive equation

$$w_1\, s^P(n) + w_2\, s^{P-1}(n) + \cdots + w_P\, s(n) + w_{P+1} = 0 \tag{46}$$

where $w_1, \cdots, w_{P+1}$ are dependent on $s(n-m)$, $m \ge 1$, and $\mathbf{t}$. Equation (46) admits at most P roots. The probability that for any n, $s(n)$ takes a value from the set of $P+1$ points, which is a root of (46) is less than or equal to $P/(P+1)$, assuming equal probability distribution for the input values. So considering (46) for all rows of $\mathcal{S}_{rc}^{(0,\,L+K)}$, it follows that the probability for $\mathcal{S}_{rc}^{(0,\,L+K)}$ not to be full column rank is less than or equal to $[P/(P+1)]^{(N-2K-L)}$, and it converges to zero as $N \to \infty$.

Appendix C
Proof of Theorem 3

Using (39) and (42)–(44), we obtain

$$\mathbf{X}(N-n,\,:)\hat{\mathbf{g}}_{P-1,\,0:0}^{(0)} = f_{11}s^{P-1}(n) + f_{21}s^{P}(n) \tag{47}$$

$$\mathbf{X}(N-n,\,:)\hat{\hat{\mathbf{g}}}_{P-1,\,0:0}^{(0)} = f_{12}s^{P-1}(n) + f_{22}s^{P}(n). \tag{48}$$

We want to show that $\mathbf{g}_{P-1,\,0:0}^{(0)}$ can be found as a linear combination of $\hat{\mathbf{g}}_{P-1,\,0:0}^{(0)}$ and $\hat{\hat{\mathbf{g}}}_{P-1,\,0:0}^{(0)}$. Let α_1 and α_2 be two constants chosen such that $\alpha_1 f_{11} + \alpha_2 f_{12} = 1$ and $\alpha_1 f_{21} + \alpha_2 f_{22} = 0$. It follows from (47) and (48), that $\mathbf{g}_{P-1,\,0:0}^{(0)} := \alpha_1\hat{\mathbf{g}}_{P-1,\,0:0}^{(0)} + \alpha_2\hat{\hat{\mathbf{g}}}_{P-1,\,0:0}^{(0)}$ satisfies $\mathbf{X}\mathbf{g}_{P-1,\,0:0}^{(0)} = \tilde{\mathbf{s}}_{P-1}$. So the problem of finding $\mathbf{g}_{P-1,\,0:0}^{(0)}$ reduces to finding constants α_1 and α_2. Because we know the output of equalizer $\mathbf{g}_{P,\,0:0}^{(0)}$, the identity $[s^{P-1}(n)]^P = [s^P(n)]^{P-1}$ will be used to recover α_1 and α_2

$$\begin{aligned}
&[\mathbf{X}(N-n,\,:)\mathbf{g}_{P-1,\,0:0}^{(0)}]^P = [\mathbf{X}(N-n,\,:)\mathbf{g}_{P,\,0:0}^{(0)}]^{P-1}\\
&\{\alpha_1[\mathbf{X}(N-n,\,:)\hat{\mathbf{g}}_{P-1,\,0:0}^{(0)}] + \alpha_2[\mathbf{X}(N-n,\,:)\hat{\hat{\mathbf{g}}}_{P-1,\,0:0}^{(0)}]\}^P\\
&\quad = s^{P(P-1)}(n).
\end{aligned} \tag{49}$$

So we obtain from (47)–(49)

$$\begin{aligned}
&\{\alpha_1[f_{11}s^{P-1}(n) + f_{21}s^{P}(n)]\\
&\quad + \alpha_2[f_{12}s^{P-1}(n) + f_{22}s^{P}(n)]\}^P = s^{P(P-1)}(n).
\end{aligned} \tag{50}$$

Defining $a_n := f_{11}s^{P-1}(n) + f_{21}s^P(n)$ and $b_n := f_{12}s^{P-1}(n) + f_{22}s^P(n)$, and using the binomial expansion, (50) reduces to

$$\binom{P}{0}\alpha_1^P a_n^P + \binom{P}{1}\alpha_1^{P-1}a_n^{P-1}\alpha_2 b_n + \cdots + \binom{P}{P}\alpha_2^P b_n^P = s^{P(P-1)}(n)$$

and for $n = N-1, \cdots, K$, in the matrix form relation $\overline{\mathbf{A}}\alpha = \tilde{\mathbf{s}}_{P(P-1)}$,

$$\begin{bmatrix} a_{N-1}^P & a_{N-1}^{P-1}b_{N-1} & \cdots & b_{N-1}^P \\ a_{N-2}^P & a_{N-2}^{P-1}b_{N-2} & \cdots & b_{N-2}^P \\ \vdots & \vdots & \vdots & \vdots \\ a_K^P & a_K^{P-1}b_K & \cdots & b_K^P \end{bmatrix} \begin{bmatrix} \binom{P}{0}\alpha_1^P \\ \binom{P}{1}\alpha_1^{P-1}\alpha_2 \\ \vdots \\ \binom{P}{P}\alpha_2^P \end{bmatrix} = \begin{bmatrix} s^{P(P-1)}(N-1) \\ s^{P(P-1)}(N-2) \\ \vdots \\ s^{P(P-1)}(K) \end{bmatrix}. \tag{51}$$

We wish to show that matrix $\overline{\mathbf{A}}$ is full column rank if the input $s(n)$, for $n = K, \cdots, N-1$, takes at least $P+1$ distinct values. Indeed, considering the determinant of submatrix $\overline{\mathbf{A}}(1 : P+1, :)$, we have

$$\det[\overline{\mathbf{A}}(1:P+1, :)] = b_{N-1}^P \cdots b_{N-P-1}^P \times \det\left\{\begin{bmatrix} \left(\frac{a_{N-1}}{b_{N-1}}\right)^P & \cdots & \left(\frac{a_{N-1}}{b_{N-1}}\right) & 1 \\ \left(\frac{a_{N-2}}{b_{N-2}}\right)^P & \cdots & \left(\frac{a_{N-2}}{b_{N-2}}\right) & 1 \\ \vdots & \vdots & \vdots & \vdots \\ \left(\frac{a_{N-P-1}}{b_{N-P-1}}\right)^P & \cdots & \left(\frac{a_{N-P-1}}{b_{N-P-1}}\right) & 1 \end{bmatrix}\right\}$$

or

$$\det[\overline{\mathbf{A}}(1:P+1, :)] = b_{N-1}^P \cdots b_{N-P-1}^P \prod_{i>j}\left(\frac{a_i}{b_i} - \frac{a_j}{b_j}\right).$$

From the definitions of a_i and b_i, we obtain

$$\frac{a_i}{b_i} = \frac{f_{11} + f_{21}s(i)}{f_{12} + f_{22}s(i)}.$$

Because $\mathbf{F}$ is nonsingular, it turns out easily that if $s(i) \neq s(j)$ then $a_i/b_i \neq a_j/b_j$. So if $s(i)$, $i = N-1, \cdots, N-P-1$ are distinct, then $\det[\overline{\mathbf{A}}(1 : P+1, :)] \neq 0$, and so $\overline{\mathbf{A}}$ is full column rank. Having established that $\overline{\mathbf{A}}$ is nonsingular, from (51) we can recover uniquely α. Indeed, taking the ratio of its first two entries, it follows that α_1/α_2 can be uniquely determined. Hence, the equalizer $\mathbf{g}_{P-1,\,0:0}^{(0)}$ can be uniquely determined (up to a constant).

Acknowledgment

The authors wish to thank the reviewers, and especially Reviewer 2, for their constructive comments. The first author wishes to thank Prof. V. Z. Marmarelis of the University of Southern California for sparking his interest in the theory of nonlinear systems, and for his support and friendship.

References

[1] B. D. O. Anderson and E. I. Jury, "Generalized Bezoutian and Sylvester matrices in multivariable linear control," *IEEE Trans. Automat. Contr.*, pp. 551–556, Aug. 1976.

[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] R. R. Bitmead, S. Y. Kung, B. D. O. Anderson, and T. Kailath, "Greatest common divisors via generalized Sylvester and Bezout matrices," *IEEE Trans. Automat. Contr.*, pp. 1043–1047, Dec. 1978.

[5] A. Bjorck and G. H. Golub, "Numerical methods for computing angles between linear subspaces," *Math. Comput.*, vol. 27, no. 123, pp. 579–594, July 1973.

[6] Z. Ding, "Blind channel identification and equalization using spectral correlation measurements, Part I: Frequency-domain analysis," in *Cyclostationarity in Communications and Signal Processing*, W. A. Gardner, Ed. New York: IEEE, 1994, p. 424.

[7] G. D. Forney, "Minimal bases of rational vector spaces, with applications to multivariable linear systems," *SIAM J. Contr.*, pp. 493–520, May 1975.

[8] G. B. Giannakis and S. Halford, "Blind fractionally-spaced equalization of noisy FIR channels: Adaptive and optimal solutions," in *Proc. ICASSP*, Detroit, MI, May 1995, vol. 3, pp. 1972–1975.

[9] G. H. Golub and C. F. Van Loan, *Matrix Computations.* Baltimore, MD: John Hopkins Univ. Press, 1983.

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

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

[12] T. Kailath, *Linear Systems.* Englewood Cliffs, NJ: Prentice-Hall, 1980, ch. 6.

[13] K. I. Kim and E. J. Powers, "A digital method of modeling quadratically nonlinear systems with a general random input," *IEEE Trans. Acoust., Speech, Signal Processing*, pp. 1758–1769, Nov. 1988.

[14] T. Koh and E. J. Powers, "Second-order Volterra filtering and its application to nonlinear system identification," *IEEE Trans. Acoust., Speech, Signal Processing*, pp. 1445–1455, Dec. 1985.

[15] M. J. Korenberg and L. D. Paarman, "Orthogonal approaches to time series analysis and system identification," *IEEE Signal Processing Mag.*, pp. 29–43, July 1991.

[16] P. Koukoulas and N. Kalouptsidis, "Nonlinear system identification using Gaussian inputs," *IEEE Trans. Signal Processing*, vol. 43, pp. 1831–1841, Aug. 1995.

[17] G. Lazzarin, S. Pupolin, and A. Sarti, "Nonlinearity compensation in digital radio systems," *IEEE Trans. Commun.*, pp. 988–998, Feb. 1994.

[18] J. Lee and V. J. Mathews, "A fast recursive least squares adaptive second-order Volterra filter and its performance analysis," *IEEE Trans. Signal Processing*, vol. 41, pp. 1087–1101, Mar. 1993.

[19] Y. Li and Z. Ding, "Blind channel identification based on second order cyclostationary statistics," in *Proc. ICASSP*, vol. 4, 1993, pp. 81–84.

[20] L. Lovasz, *Combinatorial Problems and Exercises.* Amsterdam, The Netherlands: North-Holland, 1979.

[21] P. Z. Marmarelis and V. Z. Marmarelis, *Analysis of Physiological Systems.* New York: Plenum, 1978.

[22] V. J. Mathews, "Adaptive polynomial filters," in *IEEE Signal Processing Mag.*, pp. 10–26, July 1991.

[23] E. Moulines, P. Duhamel, J. 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.

[24] R. Nowak and B. Van Veen, "Random and pseudorandom inputs for volterra filter identification," *IEEE Trans. Signal Processing*, vol. 42, pp. 2124–2135, Aug. 1994.

[25] ——, "Invertibility of higher order moment matrices," *IEEE Trans. Signal Processing*, vol. 43, pp. 705–708, Mar. 1995.

[26] S. Prakriya and D. Hatzinakos, "Blind identification of nonlinear channels with higher-order cyclic spectral analysis," in *Proc. Int. Conf. HOS*, Spain, 1995, p. 366.

[27] S. Prakriya and D. Hatzinakos, “Blind identification of LTI-ZMNL-LTI nonlinear channel models,” *IEEE Trans. Signal Processing,* vol. 43, pp. 3007–3013, Dec. 1995.
[28] N. Rozario and A. Papoulis, “The identification of certain nonlinear systems by only observing the output,” in *Proc. HOSA Workshop,* Vail, CO, 1989, pp. 78–81.
[29] M. Schetzen, *The Volterra and Wiener Theories of Nonlinear Systems.* New York: Wiley, 1980.
[30] D. Slock, “Blind fractionally-spaced equalization, perfect-reconstruction filter banks and multichannel linear prediction,” in *Proc. ICASSP,* vol. 4, 1994, pp. 585–588.
[31] L. Tong, G. Xu, and T. Kailath, “Blind identification and equalization based on second-order statistics: A time domain approach,” *IEEE Trans. Inform. Theory,* pp. 340–349, 1994.
[32] P. A. Wedin, “On angles between subspaces of a finite dimensional inner product space,” in *Matrix Pencils,* B. Kagstrom and A. Ruhe, Eds. New York: Springer-Verlag, 1983.
[33] 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, 1995.

Georgios B. Giannakis (S’84–M’86–SM’91) received the Dipl. degree in electrical engineering from the National Technical University of Athens, Greece, 1981. From September 1982 to July 1986, he was with the University of Southern California (USC), Los Angeles, where he received the M.Sc. degree in electrical engineering in 1983, the M.Sc. degree in mathematics in 1986, and the Ph.D. degree in electrical engineering in 1986.

After lecturing for one year at USC, he joined the Department of Electrical Engineering at the University of Virginia, Charlottesville, in September 1987, where he has been an Associate Professor since 1992. His general interests lie in the areas of signal processing, estimation and detection theory, and system identification. Specific research areas of current interest include time, lag, space, and code-diversity techniques for channel estimation and multiuser communications, nonstationary and cyclostationary signal analysis, wavelets in statistical signal processing, and non-Gaussian signal processing using higher order statistics with applications to sonar, array, and image processing.

Dr. Giannakis received the IEEE Signal Processing Society’s 1992 Paper Award in the Statistical Signal and Array Processing (SSAP) area. He coorganized the 1993 IEEE Signal Processing Workshop on Higher Order Statistics and the 1996 SSAP Workshop. He guest co-edited two special issues on higher order statistics (*International Journal of Adaptive Control and Signal Processing*, *EURASIP Journal of Signal Processing*), and a special issue on signal processing advances in communications of the IEEE TRANSACTIONS ON SIGNAL PROCESSING. He has served as an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING and the IEEE SIGNAL PROCESSING LETTERS, as Secretary of the Signal Processing Conference Board, and a member of the SSAP Technical Committee. He is also a member of the IMS and the European Association for Signal Processing.

Erchin Serpedin (S’95) received (with highest distinction) the Dipl. Electrical Engineer degree from the Department of Automatic Control and Computer Science, Politechnic Institute of Bucharest, Bucharest, Romania, in 1991. He received the Specialization degree in signal processing and transmission of information from Supélec, Lorraine, Metz, France, and the M.S.E.E. degree from Georgia Institute of Technology (Georgia Tech), Atlanta, both in 1992. Since August 1995, he has been working toward the Ph.D. degree in electrical engineering at University of Virginia, Charlottesville.

From 1993 to 1995, he was an instructor at the Department of Automatic Control, Politechnic Institute of Bucharest. His research interests lie in the areas of statistical signal and array processing, system identification, digital signal processing, communication theory, and mathematics.