Direct Blind Equalizers of Multiple FIR Channels: A Deterministic Approach

Georgios B. Giannakis, *Fellow, IEEE*, and Cihan Tepedelenlioğlu

***Abstract*—Blind equalization of single-input multi-output channels has practical value for inverse problems encountered in communications, sonar, and seismic data processing. Relying on diversity (sufficient number of multiple outputs), we bypass the channel estimation step and derive direct blind FIR equalizers of co-prime FIR channels. There are no constraints on the inaccessible input, apart from a minimum persistence of excitation condition; the input can be deterministic or random with unknown color or distribution. At moderate SNR (>20 dB), the resulting algorithms remain operational even with very short data records (<100 samples), which makes them valuable for equalization of rapidly fading multipath channels. Complexity, persistence of excitation order, and mean-square error performance tradeoffs are delineated for equalizers of single-shift (semi-blind), pair, or, multiple shifts estimated separately or simultaneously. Optimum and suboptimum combinations of the equalizers' outputs are also studied. Simulations illustrate the proposed algorithms and compare them with dual deterministic channel identification algorithms.**

***Index Terms*— Communications, equalization, multichannel system identification.**

I. Introduction

CONSIDER the $M \times 1$ data vector $\mathbf{x}(n) := [x_1(n) \cdots x_M(n)]'$ modeled as the noisy output of an $M \times 1$ vector FIR channel with impulse response coefficients $\mathbf{h}(l) := [h_1(l) \cdots h_M(l)]'$, $l = 0, 1, \ldots, L$ and scalar input $s(n)$ (see also Fig. 1)

$$\mathbf{x}(n) = \sum_{l=0}^{L} \mathbf{h}(l)s(n-l) + \mathbf{v}(n), \quad n = 0, 1, \ldots, N-1. \quad (1)$$

Input $s(n)$ is inaccessible (blind scenario) and can be either deterministic or a realization of a random (white or colored) process. Additive noise $v(n)$ is assumed zero-mean, white, and independent of $s(n)$ if the latter is random.

The multichannel model of (1) arises when $M > 1$ antennas receive a single information bearing source (spatial diversity) and/or when continuous-time communication signals are sampled at a rate faster than the symbol rate (temporal diversity). Tong *et al.* [17], [18] exploited the diversity offered by the multiple outputs in order to establish that under mild persistence of excitation conditions on $s(n)$, the channels $\{h_m(l)\}_{m=1}^{M}$ can be identified from the outputs $\{x_m(n)\}_{m=1}^{M}$ only, provided that the channel transfer functions $\{H_m(z)\}_{m=1}^{M}$ are co-prime. In addition to equalization of (e.g., multipath induced) communication channels, the multichannel model (1) appears in a variety of problems including sonar dereverberation and seismic deconvolution [9].

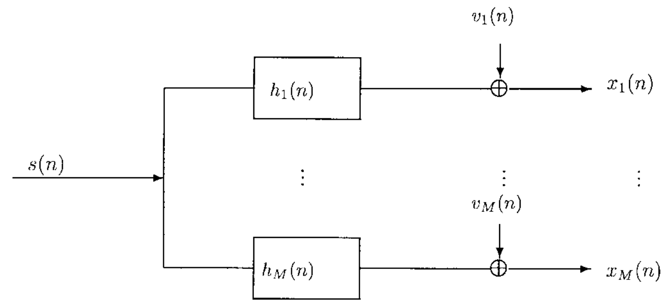

Fig. 1. Single-input, M-output channel model.

The original work of [17] sparked interest toward blind identification of possibly nonminimum phase channels without using high-order output statistics (e.g., see [5], [12], [14], [20], and references therein). Most of the existing algorithms identify $\mathbf{h}(l)$ in (1), and according to complexity tradeoffs, they either adopt the (nonlinear) Viterbi algorithm or take a second step to construct a linear FIR (zero-forcing or mean-square error) equalizer for recovering $s(n)$. However, when low-complexity online demodulation is desired, it may be advantageous to bypass the channel identification step and estimate the equalizer directly. Especially with slowly time-varying channels in wireless environments, low-complexity direct adaptive equalizers are highly desirable.

The objective of this paper is to derive Kth-order FIR equalizers $\{\mathbf{g}_i(k)\}_{k=0}^{K}$ given $\{\mathbf{x}(n)\}_{n=0}^{N-1}$, which in the noise-free case satisfy the so-called zero-forcing condition[1] (see also Fig. 2)

$$\sum_{k=0}^{K} \mathbf{x}'(n-k)\mathbf{g}_i(k) = s(n-i), \quad i = 0, 1, \ldots, L+K. \quad (2)$$

Index i denotes a delay that is nonidentifiable from output data only.

Direct blind equalizers corresponding to zero delay $(i = 0)$ were also developed in [5] using an equivalent cyclostationary formulation but assuming that $s(n)$ is white. This paper's approach not only allows for deterministic inputs but also

[1] Prime denotes Hermitian transpose, * conjugate, ^{T} transpose, and † pseudo-inverse.

Manuscript received March 21, 1997; revised July 14, 1998. This work was supported by the Office of Naval Research under Grant N00014-93-1-0485. The associate editor coordinating the review of this paper and approving it for publication was Prof. Barry D. Van Veen.

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(99)00136-1.

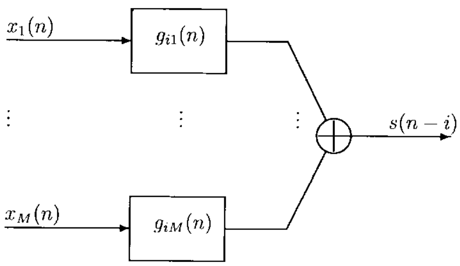

Fig. 2. FIR M-channel equalizer corresponding to shift i.

estimates $L+K+1$ equalizers corresponding to all possible shifts. Blind estimation of equalizers corresponding to all possible shifts was also suggested (without uniqueness proofs) in [15], using an interesting linear prediction framework. An algorithm for the direct estimation of all possible equalizers was also derived independently in [2] and [3] by estimating the equalizer taps via second-order statistics of the output.[2] In contrast, this paper's framework invokes no statistical assumptions on the input and, at moderate SNR's (>20 dB), yields operational equalizers using as few as 50 to 100 samples. This is because the equalizing solutions herein are algebraic and exact in the absence of noise. In addition, estimation of all equalizers can be done pairwise, which requires less data compared with methods that estimate equalizers simultaneously. The novelties of this paper's approach include the following.

1) derivation of the minimum $(i = 0)$ and maximum $(i = L+K)$ delay equalizers *separately*, when all data are available or when guard bits separate data records (a semi-blind approach);
2) joint blind estimation of $\mathbf{g}_0(k)$ and $\mathbf{g}_{L+K}(k)$ and, subsequently, $\mathbf{g}_i(k)$ for $i = 1, \ldots, L+K-1$, allowing $s(n)$ to be either deterministic, or (non-)white random;
3) identifiability proofs revealing the role of persistence-of-excitation;
4) methods to optimally combine the outputs of all equalizers corresponding to different delays $i \epsilon [0, L+K]$.

[2] Reference [2] assumes white inputs, whereas [3], which appeared after the conference version of this paper [6], dispenses with the whiteness assumption.

A conference version of the results here can be found in [6], and extensions to time-varying channels appeared in [4].

II. Direct Blind Equalizers

We will assume carrier synchronization throughout this work. Like most other blind channel identification and equalization algorithms, the proposed methods will require frequency offset estimation (see [7] and references therein for a detailed study of this topic).

Consider (2) with $n = N-1, N-2, \ldots, K$ and rewrite it in a matrix form as $\mathbf{X}\mathbf{g}_i = \mathbf{s}_i$, i.e.,

$$\underbrace{\begin{bmatrix} \mathbf{x}'(N-1) & \mathbf{x}'(N-2) & \cdots & \mathbf{x}'(N-1-K) \\ \mathbf{x}'(N-2) & \mathbf{x}'(N-3) & \cdots & \mathbf{x}'(N-2-K) \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{x}'(K) & \mathbf{x}'(K-1) & \cdots & \mathbf{x}'(0) \end{bmatrix}}_{\mathbf{X}:\,(N-K)\times M(K+1)} \times \underbrace{\begin{bmatrix} \mathbf{g}_i(0) \\ \mathbf{g}_i(1) \\ \vdots \\ \mathbf{g}_i(K) \end{bmatrix}}_{\mathbf{g}_i:\,M(K+1)\times 1} = \underbrace{\begin{bmatrix} s(N-1-i) \\ s(N-2-i) \\ \vdots \\ s(K-i) \end{bmatrix}}_{\mathbf{s}_i:\,(N-K)\times 1}. \tag{3}$$

Even if $\mathbf{s}_i$ were available, uniqueness of $\mathbf{g}_i$ would require $\mathbf{X}$ to be full column rank. To study the rank of $\mathbf{X}$, we temporarily assume noise-free data in (1) and decompose it as $\mathbf{X} = \mathbf{SH}$, where we have (4) and (5), shown at the bottom of the page. We adopt the following assumptions.

a1) The data length N and channel (equalizer) order $L(K)$ satisfy

$$N - K \geq L + K + 1 \tag{6}$$

which is easily met in practice by collecting sufficient data.

a2) The input $s(n)$ is persistently exciting (p.e.) of order at least $L+K+1$, and in view of (4) and (6)

$$\rho_s := \text{rank}(\mathbf{S}) = L + K + 1. \tag{7}$$

$$\mathbf{S} := \begin{bmatrix} s(N-1) & s(N-2) & \cdots & s(N-1-K-L) \\ s(N-2) & s(N-3) & \cdots & s(N-2-K-L) \\ \vdots & \vdots & \vdots & \vdots \\ s(K) & s(K-1) & \cdots & s(-L) \end{bmatrix}_{(N-K)\times(L+K+1)} \tag{4}$$

$$\mathbf{H} := \begin{bmatrix} \mathbf{h}'(0) & \mathbf{0}' & \cdots & \mathbf{0}' \\ \vdots & \mathbf{h}'(0) & \ddots & \\ \mathbf{h}'(L) & \vdots & \cdots & \mathbf{h}'(L-K) \\ \vdots & \mathbf{h}'(L) & \ddots & \vdots \\ \mathbf{0}' & \vdots & & \mathbf{h}'(L) \end{bmatrix}_{(L+K+1)\times M(K+1)}. \tag{5}$$

Assumption a2) amounts to requiring at least $L+K+1$ nonzero frequencies in the spectrum of $s(n)$; white noise is p.e. of any order. For the estimation of double or multiple delay equalizers, we will need $\rho_s = 2(L+K+1)-1$ because both of these methods require concatenation of two portions of a block Hankel matrix $\mathbf{X}$, which translates in a doubling of the p.e. requirement on the input $s(n)$.

a3) The triplet $(M, K, L) =$ (number of channels, equalizer order, channel order) obeys

$$M(K+1) \geq L+K+1 \tag{8}$$

which for a given M and L is satisfied by choosing the equalizer order

$$K \geq K_{\min} := \left\lceil \frac{L}{M-1} - 1 \right\rceil \tag{9}$$

where $\lceil \alpha \rceil$ denotes the smallest integer $\geq \alpha$. With a single channel ($M = 1$), it follows that (8) is satisfied only by an infinite-order equalizer ($K_{\min} = \infty$). However, with $M = 2$, an equalizer of order $K_{\min} = L-1$ satisfies (8). Therefore, as long as $K > K_{\min}$ in (9), the methods proposed in this paper will work, but unlike the single input single output case, as we also illustrate in the simulations, in the presence of noise, increasing the equalizer order does not improve performance.

a4) The block Toeplitz channel matrix $\mathbf{H}$ has full row rank, and in view of (8)

$$\text{rank}(\mathbf{H}) = L+K+1. \tag{10}$$

Equation (10) holds true if and only if the transfer functions $\{H_m(z)\}_{m=1}^{M}$ are co-prime (see e.g., [18]).

Assumptions a1)–a4) appear in all blind identification algorithms of multiple FIR channels [12], [14], [17], [18], [20]. We underscore that apart from (7), no extra assumptions are imposed on $s(n)$; in fact, similar p.e. conditions are required even for input–output identification algorithms, where the input is also available (see, e.g., [16, pp. 117–125]). Equation (8) expresses the need for diversity and implies that $\mathbf{H}$ in (5) is fat or square. On the other hand, (10) reinforces diversity by requiring channels to be co-prime (or "sufficiently different").

In this section, we will focus on the noise-free case. We first establish the following (also proved in [20] with matrices of different dimensions):

Lemma 1: Under a1)–a4), the $(N-K) \times M(K+1)$ matrix $\mathbf{X}$ in (3) has rank$(\mathbf{X}) = L+K+1$.

Proof: It follows from a1) and a2) that rank$(\mathbf{S}) = L+K+1$, and from a3) and a4) that rank$(\mathbf{H}) = L+K+1$; hence, rank$(\mathbf{X}) =$ rank$(\mathbf{SH}) = L+K+1$. □

As was done in [20], by knowing an upper bound $\bar{L}$ on the channel order L, we can estimate L and the equalizer order K. Let $\bar{L}$ be a known upper bound on L and with M given, choose $\bar{K} = \lceil \bar{L}/(M-1) - 1 \rceil$. Following the proof of Lemma 1, we have rank$(\mathbf{X}) = L + \bar{K} + 1$. In [20], this observation was used to find the order L as rank$(\mathbf{X}) - \bar{K} - 1$; here, we will assume that L is known (or has been estimated by the outlined procedure), and together with the known $M \geq 2$, we will choose K to satisfy the inequality in (8).

A. Single-Shift: Off-Line Solutions

Because $s(n)$ is not available, to solve (3) for $\mathbf{g}_i$, we need to eliminate $\mathbf{s}_i$. Our solution in this section requires that all $s(n)$ data have been processed. Such an assumption holds, e.g., with packet radio transmission where the beginning and end of each packet are specified by control bits (guard interval). The method relies on the simple observation that if N is the length of $\{\mathbf{x}(n)\}_{n=0}^{N-1}$, the nonzero support of $s(n)$ is $[0, N-1-L]$, i.e., $s(n) = 0$ for $n < 0$ and $n > N-1-L$.

Let us first consider (2) with $i = 0$ and $n = N-1+K, N-2+K, \ldots, N-1-L$; we obtain the matrix equation in (11), shown at the bottom of the page, where, similar to $\mathbf{X}$, matrix $\bar{\mathbf{X}}$ can be decomposed as $\bar{\mathbf{X}} = \bar{\mathbf{S}}\mathbf{H}$ with $\mathbf{H}$ as in (5) and (12), shown at the bottom of the page. Alternately, we consider (2) with the maximum delay $i = L+K$ and

$$\underbrace{\begin{bmatrix} \mathbf{0}' & \cdots & \mathbf{0}' & \mathbf{x}'(N-1) \\ \mathbf{0}' & \cdots & \mathbf{x}'(N-1) & \mathbf{x}'(N-2) \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{x}'(N-1-L) & \cdots & \mathbf{x}'(N-L-K) & \mathbf{x}'(N-1-L-K) \end{bmatrix}}_{\bar{\mathbf{X}}:\,(L+K+1)\times M(K+1)} \mathbf{g}_0 = \underbrace{\begin{bmatrix} 0 \\ 0 \\ \vdots \\ s(N-1-L) \end{bmatrix}}_{(L+K+1)\times 1} \tag{11}$$

$$\bar{\mathbf{S}} := \begin{bmatrix} 0 & 0 & \cdots & s(N-1-L) \\ 0 & 0 & \cdots & s(N-2-L) \\ \vdots & \vdots & \vdots & \vdots \\ s(N-1-L) & s(N-2-L) & \cdots & s(N-1-2L-K) \end{bmatrix}_{(L+K+1)\times(L+K+1)}. \tag{12}$$

$n = L+K, L+K-1, \ldots, 0$ to obtain

$$\underbrace{\begin{bmatrix} \mathbf{x}'(L+K) & \mathbf{x}'(L+K-1) & \cdots & \mathbf{x}'(L) \\ \mathbf{x}'(L+K-1) & \mathbf{x}'(L+K-2) & \cdots & \mathbf{x}'(L-1) \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{x}'(0) & \mathbf{0}' & \cdots & \mathbf{0}' \end{bmatrix}}_{\tilde{\mathbf{X}}:(L+K+1)\times M(K+1)} \mathbf{g}_{L+K} = \underbrace{\begin{bmatrix} s(0) \\ 0 \\ \vdots \\ 0 \end{bmatrix}}_{(L+K+1)\times 1} \tag{13}$$

where, again, matrix $\tilde{\mathbf{X}} = \tilde{\mathbf{S}}\mathbf{H}$ with (14), shown at the bottom of the page.

Let $\mathbf{e}_i$ denote an $(L+K+1) \times 1$ vector having unity as its $(i+1)$st entry and zero elsewhere. From (11) and (12), we know that $\bar{\mathbf{X}}\mathbf{g}_0 = \bar{\mathbf{S}}\mathbf{H}\mathbf{g}_0 = \bar{\mathbf{S}}\mathbf{e}_0$, where the last equality follows because $\bar{\mathbf{S}}\mathbf{e}_0$, which is the first column of $\bar{\mathbf{S}}$, is nothing but the right-hand side of (11). The columns of $\bar{\mathbf{S}}$ are linearly independent, and since $\mathbf{H}$ is full rank, $\mathbf{e}_0$ is in the range of $\mathbf{H}$; hence, $\mathbf{H}\mathbf{g}_0 = \mathbf{e}_0$. Thus, the $\mathbf{g}_0$ that solves (11) is the first column of the right inverse of $\mathbf{H}$ in (5) and satisfies $\mathbf{X}\mathbf{g}_0 = \mathbf{S}\mathbf{H}\mathbf{g}_0 = \mathbf{s}_0$, i.e., the solution of (11) is the desired 0-delay equalizer. The argument for $\mathbf{g}_{L+K}$ being the last column of the right inverse of $\mathbf{H}$ is very similar. These results lead to the following theorem.

Theorem 1: If a1)–a4) hold, the minimum- and maximum-delay equalizers $\mathbf{g}_0$ and $\mathbf{g}_{L+K}$ are identifiable as solutions of (11) and (13) within a scalar ambiguity inherently nonidentifiable from output data only.

Since we do not know $s(N-1-L)$ and $s(0)$, a scalar ambiguity is introduced in determining the equalizers. To fix the scale ambiguity, we divide (11) and (13) by $s(N-1-L)$ and $s(0)$, respectively, and identify the equalizers in closed form as

$$\begin{aligned} \mathbf{g}_0 &= \frac{1}{s(N-1-L)} \bar{\mathbf{X}}'(\bar{\mathbf{X}}\bar{\mathbf{X}}')^{-1}\mathbf{e}_{L+K} \\ \mathbf{g}_{L+K} &= \frac{1}{s(0)} \tilde{\mathbf{X}}'(\tilde{\mathbf{X}}\tilde{\mathbf{X}}')^{-1}\mathbf{e}_0. \end{aligned} \tag{15}$$

If $s(N-1-L)$ is zero, we need to rewrite (11), replacing $s(N-1-L)$ with $s(n_0)$, where n_0 is the greatest index for which $s(n)$ is nonzero, and decrease each index in (11) by $N-1-L-n_0$. A similar manipulation of (13) can be made if $s(0) = 0$.

Because such equalizers are forcing zeros outside the $[0, N-1-L]$ support of $s(n)$, they are known as zero-forcing (ZF) equalizers. Interestingly, since with $M = 1$ a3) is not satisfied, exact FIR solutions like (11) and (13) are not possible in the single-channel case.

We also note that equalizers with delays $i \epsilon [1, L+K-1]$ cannot be obtained directly, e.g., equalizer $\mathbf{g}_1$ would require not only $s(N-1-L)$, as $\mathbf{g}_0$ does in (11), but also the input sample $s(N-2-L)$. Hence, the ambiguity would be greater than that of a scale. However, once $\mathbf{g}_0$ (or $\mathbf{g}_{L+K}$) is found, it will be shown in the next section how (3) can be used to find a relationship between $\mathbf{g}_0$ and $\mathbf{g}_i$ for $0 < i < L+K$, thus enabling us to estimate equalizers corresponding to $0 < i < L+K$ using the single-shift equalizer estimate $\mathbf{g}_0$. These "mid-delay" equalizers may perform better than the minimum and maximum delay equalizers because practical channel impulse responses usually taper off at both ends (see also [3]). Indeed, if the components of $\mathbf{h}(l)$ are small for small and large values of l, then the equalizing solutions of $\mathbf{H}\mathbf{g}_0 = \mathbf{e}_0$ and $\mathbf{H}\mathbf{g}_{L+K} = \mathbf{e}_{L+K}$ must have large norms, which results in noise amplification in the equalization process.

Notice that the minimum number of consecutive zeros needed for the single-shift method is $L+K$, as is clear from (11) and (13). Considering equation (9), we have $L+K_{\min} \leq 2L$. Hence, the amount of time for the guard interval is less than $2LT$, where T is the symbol duration. In addition, recall that in both (11) and (13), $L+K+1$ different vectors $\mathbf{x}(n)$ are used. Equalizer estimates cannot use more data since it would again yield an ambiguity greater than that of a scale on the right-hand side of (11) or (13). Because the equalizing solutions are algebraic (i.e., deterministic), their performance does not improve as more data are acquired (there is no explicit incorporation of noise averaging). On the positive side, however, the method can work with very small data lengths $(N < 100)$ and exploits small guard times to acquire quick equalizer estimates with partial knowledge of the periodic bursts of zeros in the input sequence, which is a feature present in TDMA systems like GSM. The periodic presence of guard times in the input sequence enables block synchronization, which is necessary for this method (see [13] and [19] for detailed discussion on blind block synchronization). That the single-shift method can work with very small data lengths is also advantageous in scenarios where the time invariance assumption on the channel is violated during the time it takes to collect sufficient samples for estimating the received data statistics required by statistical methods. Exploitation of the known guard times leads us to term this method as *semi-blind*.

Next, we will describe *blind* deterministic algorithms capable of identifying equalizers pairwise and corresponding to all shifts, simultaneously. More importantly, they will be capable of operating with any segment of the data, thereby allowing for more flexible operation (recall that zeroing the unavailable

$$\tilde{\mathbf{S}} := \begin{bmatrix} s(L+K) & s(L+K-1) & \cdots & s(0) \\ s(L+K-1) & s(L+K-2) & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ s(0) & 0 & \cdots & 0 \end{bmatrix}_{(L+K+1)\times(L+K+1)}. \tag{14}$$

data makes sense only when the entire record is available or when guard bits are inserted).

B. Multiple Shifts: Simultaneous Solutions

Our approach of eliminating $s(n)$ in this section relies on pairwise combination of equalizers corresponding to two different shifts. From this viewpoint, our algorithms are "dual" to those in [20], where pairs of the multiple outputs were combined for channel identification. Here, we combine pairs of the equalizer outputs (corresponding to two or more shifts) in order to identify the equalizers.

Let us consider (2) with $i = 0$ to find

$$\sum_{k=0}^{K} \mathbf{x}'(n-k)\mathbf{g}_0(k) = s(n) \tag{16}$$

and substitute $n + i \leftarrow n$ in (2) to obtain

$$\sum_{k=0}^{K} \mathbf{x}'(n+i-k)\mathbf{g}_i(k) = s(n). \tag{17}$$

Eliminating $s(n)$ from (16) and (17) yields the desired cross relation, which makes direct blind equalization possible by solving a system of linear equations in

$$\sum_{k=0}^{K} \mathbf{x}'(n-k)\mathbf{g}_0(k) = \sum_{k=0}^{K} \mathbf{x}'(n+i-k)\mathbf{g}_i(k). \tag{18}$$

To describe our approach, we adopt the MATLAB notation $\mathbf{X}(i_1 : i_2, :)$ to denote a submatrix of $\mathbf{X}$ formed by the i_1 through i_2 rows and all columns of $\mathbf{X}$.

From the definition of $\mathbf{s}_i$ in (3), it follows that $\mathbf{s}_0(i+1 : N-K) = \mathbf{s}_i(1 : N-K-i)\ \forall i = 1, \ldots, L+K$. On defining the corresponding submatrices of $\mathbf{X}$ as

$$\mathbf{X}_{0,i} := \mathbf{X}(i+1 : N-K, :), \quad \mathbf{X}_i := \mathbf{X}(1 : N-K-i, :) \tag{19}$$

we infer that for i- and 0-delay equalizers, the corresponding left-hand sides of (3) are related via $\mathbf{X}_{0,i}\mathbf{g}_0 = \mathbf{X}_i\mathbf{g}_i$ or, equivalently

$$\underbrace{[\mathbf{X}_{0,i} \quad -\mathbf{X}_i]}_{\mathcal{X}_{0,i}:(N-K-i)\times 2M(K+1)} \underbrace{\begin{bmatrix}\mathbf{g}_0\\ \mathbf{g}_i\end{bmatrix}}_{\mathbf{g}_{0,i}:2M(K+1)\times 1} = \mathbf{0}, \quad i = 1, \ldots, L+K. \tag{20}$$

We will show that the pair of equalizers $(\mathbf{g}_0, \mathbf{g}_i)$ that solve (20) yield, up to a scale, perfect input estimates when multiplied with the data matrix $\mathbf{X}$ in the absence of noise if and only if $i = L + K$. The following theorem details why the solution $(\mathbf{g}_0, \mathbf{g}_{L+K})$, which is obtained from (20) with $i = L + K$, yields the first and last columns of the right inverse of $\mathbf{H}$ and, hence, yield the desired minimum $(i = 0)$ and maximum $(i = L + K)$ delay equalizers.

Theorem 2: Suppose $\mathbf{x}(n)$ obeys the noise-free $(\mathbf{v}(n) = \mathbf{0})$ model (1) with (N, M, L, K) satisfying (6) and (8). If input $s(n)$ is p.e. of order $\rho_s = 2(L+K+1)-1$, the channels satisfy (10), and the pair $(\mathbf{g}_0, \mathbf{g}_{L+K})$ satisfies (20) with $i = L + K$; then

$$\mathbf{X}_{0,L+K}\mathbf{g}_0 = \mathbf{X}_{L+K}\mathbf{g}_{L+K} = \alpha\mathbf{s}_{L+K} \tag{21}$$

for some complex constant α, i.e., the pair $(\mathbf{g}_0, \mathbf{g}_{L+K})$ that solves (20) is the desired equalizer pair that yields (within a scale) $\mathbf{s}_{L+K}$.

Proof: The first equality follows from (20) with $i = L+K$. Following (19) with $i = L+K$, define the submatrices $\mathbf{S}_{0,L+K}$ and $\mathbf{S}_{L+K}$ of matrix $\mathbf{S}$ in (4). Decompose $\mathcal{X}_{0,L+K}$ as

$$\mathcal{X}_{0,L+K} = \mathcal{S}_{0,L+K}\mathcal{H}_{0,L+K} = [\mathbf{S}_{0,L+K} \quad \mathbf{S}_{L+K}]\begin{bmatrix}\mathbf{H} & \mathbf{0}\\ \mathbf{0} & -\mathbf{H}\end{bmatrix} \tag{22}$$

where $\mathbf{H}$ is given by (5), and $\mathcal{S}_{0,i}$ is an $(N-K-i) \times 2(L+K+1)$ matrix given by (23), shown at the bottom of the page, and $i = L + K$. We observe from (23) that except for the first column of submatrix $\mathbf{S}_{0,L+K}$ (which is identical to the last column of submatrix $\mathbf{S}_{L+K}$), all other columns of $\mathcal{S}_{0,L+K}$ are distinct and, due to the p.e. condition, linearly independent. The first equality in (21) implies $\mathbf{S}_{0,L+K}\mathbf{H}\mathbf{g}_0 = \mathbf{S}_{L+K}\mathbf{H}\mathbf{g}_{L+K}$, which, due to the linear independence of the *distinct* columns of $\mathcal{S}_{0,L+K}$, implies that $\mathbf{H}\mathbf{g}_0 = \alpha\mathbf{e}_0$ and $\mathbf{H}\mathbf{g}_{L+K} = \alpha\mathbf{e}_{L+K}$ for some complex constant α (such a pair of equalizers exists since $\mathbf{H}$ is full rank). Hence, the products in (21) pick up the first and last column of $\mathcal{S}_{0,L+K}$, respectively, which establishes the second equality in (21). □

We have shown in Theorem 2 that if $i = L + K$ and (20) is solved, then the solution $(\mathbf{g}_0, \mathbf{g}_{L+K})$ will yield the input as in (21). It is interesting to note that if $i < L + K$ is used, the corresponding submatrices $\mathbf{S}_{0,i}$ and $\mathbf{S}_i$ would have $L+K+1-i$ identical columns from (23), which means that the solution to (20) would not necessarily yield the input $\alpha\mathbf{s}_i$ when multiplied by $\mathbf{X}$. Therefore, for (21) to hold, we must have $i = L + K$, and $\mathbf{H}$ in (5) must be square or fat. If $\mathbf{H}$ is square, the solution to (20) will be unique (up to a scale) since in that case, the right inverse of $\mathbf{H}$ is unique; if $\mathbf{H}$ is fat, which is the case when (8) is a strict inequality, then the first and last columns of the right inverse of $\mathbf{H}$ will not be unique, but they will still satisfy (21).

It is known that in the noisy case, performance of the equalizers depends on the delay i, which in turn, depends on the shape of the channel $\mathbf{h}(l)$ (see e.g., [5]). We are thus prompted to consider equalizers of shifts $i < L + K$

$$\mathcal{S}_{0,i} := \underbrace{\left[\begin{array}{ccc|ccc} s(N-1-i) & \cdots & s(N-1-L-K-i) & s(N-1) & \cdots & s(N-1-L-K)\\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots\\ s(K) & \cdots & s(-L) & s(K+i) & \cdots & s(-L+i)\end{array}\right]}_{\mathbf{S}_{0,i} \qquad\qquad \mathbf{S}_i} \tag{23}$$

simultaneously with the $(0, L+K)$ equalizers that guarantee identifiability under p.e. conditions. Concatenating equations like (20) with pairs $(0,1), (0,2), \ldots, (0, L+K)$, we get

$$\underbrace{\begin{bmatrix} \mathbf{X}_{0,1} & -\mathbf{X}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{X}_{0,2} & \mathbf{0} & -\mathbf{X}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{X}_{0,L+K} & \mathbf{0} & \mathbf{0} & \cdots & -\mathbf{X}_{L+K} \end{bmatrix}}_{\mathcal{X}:\ \sum_{i=1}^{L+K}(N-K-i)\times(L+K+1)M(K+1)} \underbrace{\begin{bmatrix} \mathbf{g}_0 \\ \mathbf{g}_1 \\ \vdots \\ \mathbf{g}_{L+K} \end{bmatrix}}_{\mathbf{g}} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix}. \tag{24}$$

Similar to $\mathcal{X}_{0,i}$ in (22), supermatrix $\mathcal{X}$ can be decomposed as

$$\mathcal{X} = \underbrace{\begin{bmatrix} \mathbf{S}_{0,1} & \mathbf{S}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{S}_{0,2} & \mathbf{0} & \mathbf{S}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{S}_{0,L+K} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{S}_{L+K} \end{bmatrix}}_{\mathcal{S}} \times \underbrace{\begin{bmatrix} \mathbf{H} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & -\mathbf{H} & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & -\mathbf{H} \end{bmatrix}}_{\mathcal{H}}. \tag{25}$$

The matrix entries $\mathbf{S}_{0,i}$ and $\mathbf{S}_i$ in (25) have dimensionality $(N-K-i)\times(L+K+1)$. Equations (24) and (25) together imply that $\mathbf{S}_{0,i}\mathbf{H}\mathbf{g}_0 = \mathbf{S}_i\mathbf{H}\mathbf{g}_i, i = 0, \ldots, L+K$. Then $\mathbf{S}_{0,L+K}\mathbf{H}\mathbf{g}_0 = \mathbf{S}_{L+K}\mathbf{H}\mathbf{g}_{L+K}$ holds. From the proof of Theorem 2, we have that $\mathbf{H}\mathbf{g}_0 = \alpha\mathbf{e}_0$ for some complex number α. This means that $\mathbf{S}_i\mathbf{H}\mathbf{g}_i = \mathbf{S}_{0,i}\mathbf{H}\mathbf{g}_0 = \mathbf{S}_{0,i}\alpha\mathbf{e}_0 = \alpha\mathbf{s}_i, i = 0, \ldots, L+K$. Hence, we have established the following theorem.

Theorem 3: If $\mathbf{x}(n)$ comes from (1) with $\mathbf{v}(n) \equiv \mathbf{0}$, and assumptions a1)–a4) hold true, and $\mathbf{g}_i, i = 0, \ldots, L+K$ satisfy (24), then

$$\mathbf{X}_{0,i}\mathbf{g}_0 = \mathbf{X}_i\mathbf{g}_i = \alpha\mathbf{s}_i, \quad i = 0, \ldots, L+K.$$

for some scale ambiguity α.

We emphasize that in comparison with the pairwise recovery of the minimum and maximum delay equalizers from (20), the simultaneous solution of (24) is useful because it gives us a choice of selecting the minimum mean-square error (MMSE) equalizer among the $L+K+1$ zero-forcing equalizers. This will be discussed in more detail in Section III. Another way of obtaining all possible equalizers is to use $\mathbf{X}_{0,i}\mathbf{g}_0 = \mathbf{X}_i\mathbf{g}_i$ for $i = 0, \ldots, L+K$; having estimated $\mathbf{g}_0$ from (20), any other equalizer can be obtained as

$$\mathbf{g}_i = \mathbf{X}_i^{\dagger}\mathbf{X}_{0,i}\mathbf{g}_0, \quad i = 1, \ldots, L+K. \tag{26}$$

Such an estimator may be preferred over (24) since there are fewer parameters at a time to be estimated, and the matrices involved have smaller size compared with $\mathcal{X}$ in (24). It should also be noted that (26) can also be used in the single-shift method of the previous subsection to find equalizers corresponding to $0 < i < L+K$. In any case, all methods proposed in this subsection, when compared with [12], [17], [18], and [20], find the equalizers directly and appear to be suitable for on-line linear equalization of communication signals. It should also be noted that the methods proposed here do not necessitate the estimation of all possible delay equalizers for the estimation of one; hence, they are computationally more attractive than [2] and [3]. Nevertheless, when it is believed that the minimum or the maximum delay equalizers will do poorly compared with the other delays, then having the estimates of the zeroth and $(L+K)$th delay equalizers will enable estimation of other delays using (26).

If the objective is blind channel identification, the equalizers obtained can be used as columns of a $M(K+1)\times(L+K+1)$ matrix $\mathbf{G}$, which must be inverted to recover $\mathbf{H}$ based on the zero-forcing condition, cf., (1) and (2)

$$\mathbf{HG} = \mathbf{I}. \tag{27}$$

We could be tempted to think that $\mathbf{g}_i$'s are linearly dependent, shifted versions of each other. However, (27) shows that $\mathbf{g}_i$'s are nothing but the columns of $\mathbf{H}^{-1}$, which by (10) are linearly independent. As usual, two options become available when solving (24) or (20). The first option is to fix the first entry of $\mathbf{g}$ to one, move the first column $\mathbf{x}_1$ of $\mathcal{X}$ to the right-hand side, and solve for the normalized $\bar{\mathbf{g}}$ using least-squares

$$\bar{\mathbf{g}}_{\text{LS}} = (\bar{\mathcal{X}}'\ \bar{\mathcal{X}})^{-1}\bar{\mathcal{X}}'\bar{\mathbf{x}} \tag{28}$$

where $\bar{\mathcal{X}}$ is the matrix left after removing the first column of $\mathcal{X}$, and $\bar{\mathbf{x}} := -\mathbf{x}_1$. The second approach is to set $||\mathbf{g}||^2 = \mathbf{g}'\mathbf{g} = 1$ and find $\mathbf{g}$ as the eigenvector corresponding to the minimum eigenvalue of $\mathcal{X}'\mathcal{X}$, which can be done by calculating the singular value decomposition of the matrix $\mathcal{X}$.

III. Input Recovery and Noisy Case

The equalizers derived so far recover the input perfectly in the absence of noise. In this section, we will discuss in what sense our methods are optimal in the presence of noise.

For the single shift case (Section II-A), the equalizers obtained are

$$\begin{aligned} \mathbf{g}_0 &= (\bar{\mathbf{X}} + \bar{\mathbf{V}})^{\dagger}\frac{1}{s(N-1-L)}\mathbf{e}_0 \\ \mathbf{g}_{L+K} &= (\tilde{\mathbf{X}} + \tilde{\mathbf{V}})^{\dagger}\frac{1}{s(0)}\mathbf{e}_{L+K} \end{aligned} \tag{29}$$

where $\bar{\mathbf{V}}(\tilde{\mathbf{V}})$ is a "noise matrix" with the same structure as $\bar{\mathbf{X}}(\tilde{\mathbf{X}})$. As a consequence of the zero-forcing criterion employed, the estimate in (29) does not take the noise into account and will be close to the true estimate in (15) whenever the SNR is relatively high (>15–20 dB).

Unlike the single-shift case, the estimators in (20) and (24) possess a certain optimality inherent in the least-squares approach pursued in their derivation. In the presence of noise,

$\mathcal{X}_{0,L+K}$ in (20) will be full rank, in which case, our algorithm finds $\mathbf{g}_{0,L+K}$ with the criterion [see also (18)]

$$\hat{\mathbf{g}}_{0,L+K} = \underset{\|\mathbf{g}_{0,L+K}\|=1}{\arg\min} \sum_{n=0}^{N-1} \sum_{k=0}^{K} |\mathbf{x}'(n-k)\mathbf{g}_0(k) - \mathbf{x}'(n+L+K-k)\mathbf{g}_{L+K}(k)|^2 \tag{30}$$

which is the squared norm of the left-hand side in (20). Notice that the inner product of the data $\mathbf{x}(n)$ with the corresponding equalizer in (30) is an estimate of the shifted (by $i = 0$ or $i = L+K$) input. Therefore, in the noisy case, we are estimating the equalizers to minimize the error between two input estimates corresponding to two equalizer outputs with shifts $i = 0$ and $i = L+K$. This approach can be seen as the "deterministic" counterpart of what is referred to in [2] as mutually referenced equalizers.

We can argue similarly for (24), where equalizers corresponding to all delays are involved. In the presence of noise, matrix $\mathcal{X}$ will be full rank. We obtain $\hat{\mathbf{g}}$ in (24) based on the criterion

$$\hat{\mathbf{g}} = \underset{\|\mathbf{g}\|=1}{\arg\min} \sum_{i=1}^{L+K} \sum_{n=0}^{N-1} \sum_{k=0}^{K} |\mathbf{x}'(n-k)\mathbf{g}_0(k) - \mathbf{x}'(n+i-k)\mathbf{g}_i(k)|^2. \tag{31}$$

In this case, the criterion is an average of the difference of input estimates for different delays.

We are capable of estimating $L+K+1$ different equalizers, each of which might perform differently in the presence of noise, depending on the shape of the channel [5]. To make use of this multitude of equalizers, we might either choose to select one that performs "best" in some sense, or we might choose to combine the outputs of these equalizers under some criterion. First, we will discuss the former option. For this purpose, we will assume that $s(n)$ is a zero-mean, stationary, uncorrelated sequence, and we wish to find, among the $L+K+1$ zero-forcing equalizers, the one that minimizes $E|\hat{s}(n) - s(n-i)|^2$, where i is the delay index. Using (2) and the orthogonality principle, it is not difficult to show that

$$\min_i E|\hat{s}(n) - s(n-i)|^2 = \min_i \mathbf{g}_i'\mathbf{R}_v\mathbf{g}_i \tag{32}$$

where $\mathbf{R}_v$ is the $(L+K+1)\times(L+K+1)$ covariance matrix of the fractionally sampled noise process (or the covariance matrix of the noise vector for the case of antennas). This result (which is also mentioned in [2] and [5] without elaboration) is intuitively appealing since the equalizer with the minimum norm will amplify the noise the least.

Availability of equalizers corresponding to all possible shifts also allows for averaging the deconvolved signals (see also [15]). Specifically, relying on (17), we can recover $s(n)$ by aligning the equalizer outputs and computing a *weighted* average over all possible shifts (see Fig. 3)

$$\bar{s}(n) = \frac{1}{L+K+1} \sum_{i=0}^{L+K} w_i \left[\sum_{k=0}^{K} \mathbf{x}'(n+i-k)\mathbf{g}_i(k) \right] \tag{33}$$

where w_i are the weights that sum up to $L+K+1$. Such a constraint on the weights guarantees that there is no scale change in the average estimate $\bar{s}(n)$. In order to find the $\hat{\mathbf{w}} = [\hat{w}_0 \cdots \hat{w}_{L+K}]^T$ that minimizes $E|\bar{s}(n) - s(n)|^2$, we use (33) and (2) to write

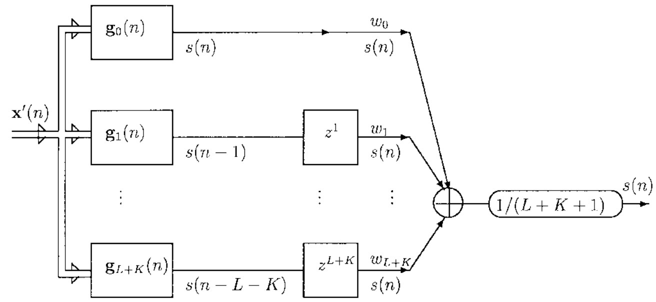

Fig. 3. Multishift equalizer and averaged input recovery.

$$E|\bar{s}(n) - s(n)|^2 = E\left| \frac{1}{L+K+1} \sum_{i=0}^{L+K} w_i \sum_{k=0}^{K} \mathbf{v}'(n+i-k)\mathbf{g}_i(k) \right|^2. \tag{34}$$

It is well known that the solution to this quadratic minimization problem $\hat{\mathbf{w}} = \arg\min_{\mathbf{w}} E|\bar{s}(n) - s(n)|^2$ under the linear constraint $\sum_{i=0}^{L+K} w_i = L+K+1$ has a solution $\hat{\mathbf{w}} = c\mathbf{A}^{-1}\mathbf{1}$, where c is a scalar so that $\hat{\mathbf{w}}$ satisfies the constraint $\mathbf{1} := [1 \cdots 1]^T$, and the (i,j) element of $\mathbf{A}$ is given by

$$[\mathbf{A}]_{i,j} = \sum_{k_1,k_2=0}^{K} [\mathbf{g}_i'(k_1)\mathbf{R}_v(j-i+k_1-k_2)\mathbf{g}_j(k_2) + \mathbf{g}_j'(k_1)\mathbf{R}_v(i-j+k_1-k_2)\mathbf{g}_i(k_2)] \tag{35}$$

where $\mathbf{R}_v(m) := E[\mathbf{v}(n)\mathbf{v}'(n+m)]$. This result can be obtained using Lagrange multipliers (see, e.g., [8]). To find the optimal weights in (33), we first obtain $\{\hat{\mathbf{g}}_i\}_{i=0}^{L+K}$ estimates by solving (24) or (26) and use (35) to compute the $\hat{\mathbf{A}}$ estimates. To satisfy the constraint $\mathbf{1}^T\hat{\mathbf{w}} = L+K+1$, we choose $c = (L+K+1)/(\mathbf{1}^T\hat{\mathbf{A}}^{-1}\mathbf{1})$ and select the weights as

$$\hat{\mathbf{w}} = \frac{L+K+1}{\mathbf{1}^T\hat{\mathbf{A}}^{-1}\mathbf{1}}\hat{\mathbf{A}}^{-1}\mathbf{1}. \tag{36}$$

Notice that matrix $\mathbf{A}$ and, therefore, $\hat{\mathbf{w}}$, turn out to be real, even though the search for $\mathbf{w}$ was made over the complex field. A possible value for the weight vector is $\hat{\mathbf{w}} = \mathbf{e}_i$, which amounts to choosing the ith equalizer. Therefore, using a general weighting vector enables the selection of the equalizer with the MMSE. Therefore, combining the outputs of all equalizers with the optimum $\hat{\mathbf{w}}$ in (36) will yield a lower (or equal) MSE than the single zero-forcing equalizer with the MMSE suggested in (32).

Assuming the noise is zero-mean Gaussian and that we have the true zero-forcing equalizers, we can also calculate the probability of symbol error. Because $v(n)$ is assumed Gaussian and the equalizers are linear, the error $e(n) = \bar{s}(n) - s(n)$ is also Gaussian. The distribution of $e(n)$ can be computed by the knowledge of the variance of $e(n)$ in (34), which enables the calculation of the probability of error. We assume here that we have the zero-forcing equalizers. Since the true zero-forcing equalizers cannot be obtained with noisy data, the above procedure will only approximate the probability of error.

In [5] and [11], a performance index is optimized in order to select the "optimum-shift" equalizer, and links of the present approach with [11] and the linear prediction framework of [15] are worth further investigation. Thorough performance analysis is beyond the scope of this paper, and results will be reported elsewhere.

IV. Adaptive Algorithms

Equation (20) can be recast in a least squares framework by setting the first coefficient of $\mathbf{g}_{0,L+K}$ to 1 and can be rewritten as

$$\bar{\mathcal{X}}\bar{\mathbf{g}} = \bar{\mathbf{x}} = -\mathbf{x}_1 \tag{37}$$

where

- $\bar{\mathcal{X}}$ $\mathcal{X}_{0,L+K}$ without its first column;
- $\mathbf{x}_1$ vector containing the elements of that column;
- $\bar{\mathbf{g}}$ $\mathbf{g}_{0,L+K}$ without its first element.

It is known that RLS is a recursive way of computing $\bar{\mathbf{g}}_{\mathrm{LS}} = (\bar{\mathcal{X}}'\bar{\mathcal{X}})^{-1}\bar{\mathcal{X}}'\bar{\mathbf{x}}$, which is the solution to the least squares problem in (37) (see, e.g., [8, p. 314]). We will use this algorithm to update the vector of equalizer coefficients online.

Let $\boldsymbol{\eta}_t^T$ be the tth row of $\bar{\mathcal{X}}$, $\hat{\mathbf{g}}_t$ the estimate of the vector of equalizer coefficients at iteration t, $\mathcal{X}_t := [\boldsymbol{\eta}_1 \cdots \boldsymbol{\eta}_t]^T$, $\mathbf{P}_t := (\bar{\mathcal{X}}_t' \ \bar{\mathcal{X}}_t)^{-1}$, and scalar x_{1t} the tth location in $\mathbf{x}_1$. Using the matrix inversion lemma, we can relate the inverse of the correlation matrix estimate at time t, which is $\mathbf{P}_t$, to its previous values and the input of the equalizer $\boldsymbol{\eta}_t$ [8]. This enables the update of $\hat{\mathbf{g}}_t$ without having to invert the correlation matrix estimate.

The recursion used for RLS is (see e.g., [8])

$$\begin{aligned} \mathbf{k}_t &= \frac{\lambda^{-1}\mathbf{P}_{t-1}\boldsymbol{\eta}_t}{1+\lambda^{-1}\boldsymbol{\eta}_t'\mathbf{P}_{t-1}\boldsymbol{\eta}_t} \\ e_t &= x_{1t} - \hat{\mathbf{g}}_{t-1}'\boldsymbol{\eta}_t \\ \hat{\mathbf{g}}_t &= \hat{\mathbf{g}}_{t-1} + \mathbf{k}_t e_t' \\ \mathbf{P}_t &= \lambda^{-1}\mathbf{P}_{t-1} - \lambda^{-1}\mathbf{k}_t\boldsymbol{\eta}_t'\mathbf{P}_{t-1}. \end{aligned} \tag{38}$$

Initialization of the correlation matrix and the estimate of the equalizer coefficients can be made, as usual, from the batch estimates discussed in Section II. Alternatively, $\mathbf{P}_0 = \delta^{-1}\mathbf{I}$ and $\hat{\mathbf{g}}_0 = \mathbf{0}$ could be used for initialization where δ is a small number. The forgetting factor λ should be chosen as unity whenever the channel is known to be time invariant. This would yield the least squares solution online. Choosing $\lambda < 1$ is necessary whenever the algorithm needs to track "slow" variations.

We could also be interested in using the computationally less intensive LMS algorithm at the expense of accuracy and slow convergence. In the absence of a training sequence (desired input), we consider the elements of $\mathbf{x}_1 = -\bar{\mathbf{x}}$ in (28) as our desired sequence that we would like $\hat{\mathbf{g}}_t'\boldsymbol{\eta}_t$ to estimate. At each iteration, the vector of equalizer coefficients is updated by using the recursions

$$\begin{aligned} \hat{\mathbf{g}}_{t+1} &= \hat{\mathbf{g}}_t + \mu e_{t+1}^*\boldsymbol{\eta}_t \\ e_t &= \hat{\mathbf{g}}_t'\boldsymbol{\eta}_t - x_{1t} \end{aligned} \tag{39}$$

TABLE I
Channel Coefficients

	Channel I		Channel II	
i	1	2	1	2
$h_i(0)$	1.0	-1.023-0.501i	0.1662 - 0.0372i	0.8404 - 0.0862i
$h_i(1)$	-1.280-0.301i	0.106+1.164i	1.0156 - 0.0036i	0.3931 + 0.1373i
$h_i(2)$	1.617+2.385i	1.477+1.850i	-0.1114 + 0.1899i	-0.0816 + 0.1385i
$h_i(3)$	0.178+0.263i	-0.482-0.523i	0.0572 + 0.0474i	0.0552 - 0.0125i
$h_i(4)$	0	0	-0.0069 - 0.0155i	-0.0367 + 0.0061i
$h_i(5)$	0	0	-0.0086 − 0.0167i	0.0085 + 0.0126i
$h_i(6)$	0	0	-0.0464 − 0.0074i	-0.0528 + 0.0010i
$h_i(7)$	0	0	-0.0267 − 0.0098i	-0.0869 + 0.0185i

where the step size μ is a parameter that should be adjusted according to the tradeoff between speed of convergence and steady-state performance.

V. Simulations

We used two different channels to test our algorithms. The first one has order $L = 3$ and the coefficients were chosen randomly, whereas the second channel is a length-16 version of an empirically measured $T/2$ spaced digital microwave radio channel $(M = 2)$ with 230 taps, which we truncated to obtain a channel with $L = 7$. The shortened version is derived by linear decimation of the FFT of the "full-length" $T/2$-spaced impulse response and taking the IFFT of the decimated version (see [1] for more details on this channel). The channel coefficients for both sets of channels are listed in Table I. Both channels were excited by a QPSK sequence. All plots, with the exception of the eye diagrams and probability-of-error curves, were averaged over 100 Monte Carlo runs.

Test Case 1—Single Shift Equalizers (Batch Method): Fig. 4 illustrates a comparison of the method in [20] where the equalizer is obtained by estimating the channel matrix in (5) and using the column of its inverse with the minimum norm. The method in [20] will be abbreviated as XLTK hereafter. On the left, we see that the probability of symbol error for both methods we proposed [single shift (SS) and multiple shift (MS) with equalizers with the minimum norm being used] is better especially in the critical range of SNR < 25 dB. We see a similar outcome for the MSE plot on the right. The RMSE between a vector $\mathbf{a}$ and its estimate $\hat{\mathbf{a}}$ was computed as

$$\mathrm{RMSE} = \sqrt{\frac{1}{R}\sum_{r=1}^{R}\frac{\|\hat{\mathbf{a}}_r - \mathbf{a}_r\|^2}{\|\mathbf{a}_r\|^2}} \tag{40}$$

where r stands for realization, and R is the number of realizations. $N = 100$ samples were used in Fig. 4. Note, however, that our equalizer estimators only use $L + K + 1 = 3 + 2 + 1 = 6$ vector $\mathbf{x}(n)$ samples. Recall that SS is a semi-blind method that exploits existing guard times, whereas MS and XLTK make no use of such knowledge.

Fig. 5 shows how, at SNR = 20 dB, the eye diagrams of the equalized symbols look for different equalizer delays with $N = 80$ data points. The zero-delay equalizer is estimated with the single-shift method, and the equalizer corresponding to $i = 4$ is estimated using the estimated $\mathbf{g}_0$ and (26). Observe

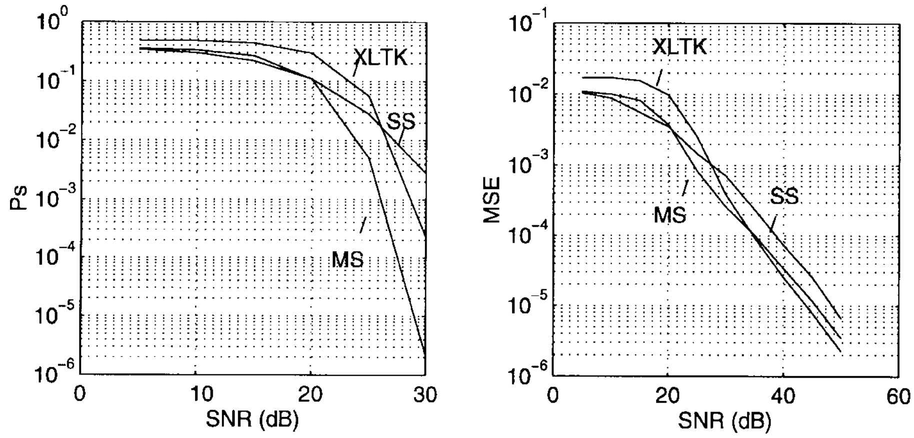

Fig. 4. Single-shift compared with XLTK and multiple shift.

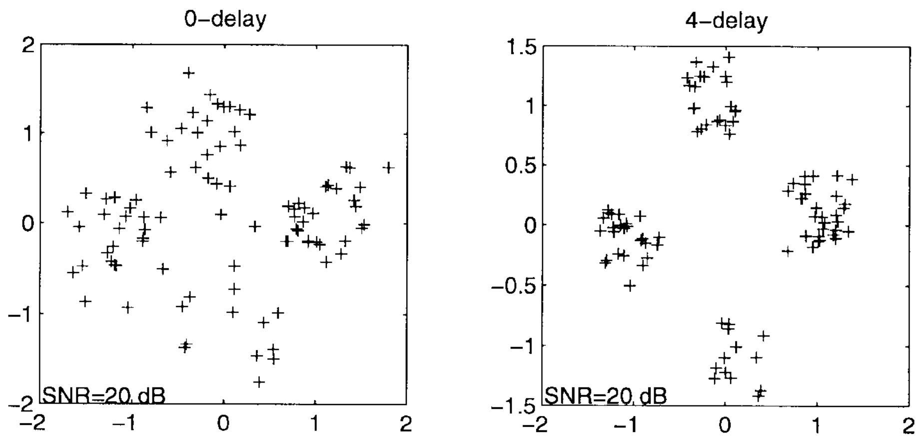

Fig. 5. Eye diagrams for different delay equalizers.

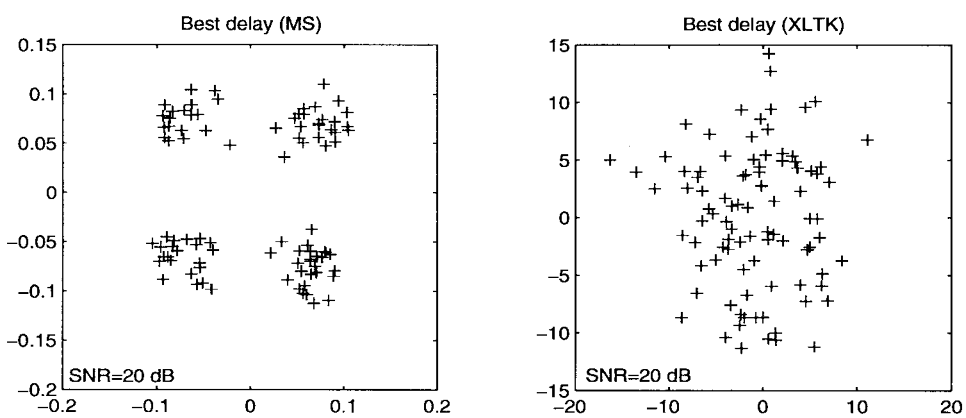

Fig. 6. Eye diagrams for MS and XLTK.

how important it is to have the minimum or maximum delay equalizer available for the estimation of other delays.

Fig. 6 illustrates an eye-diagram comparison of the MS method with the method in [20] at an SNR of 20 dB and $N = 100$ data points. Together with Fig. 4, Fig. 6 illustrates that the MS method does better than [20] in this SNR region.

In Fig. 7, robustness of the single-shift estimation algorithm with the equalizer order estimate is illustrated. Here, $N = 100$ data points were used, and the SNR was 30 dB. Note that although $K_{\min} = 2$, MSE performance does not deteriorate significantly, even with equalizer orders as high as $K = 10$.

Test Case 2—Batch and RLS Direct Equalizers: Fig. 8 shows the RMSE of the estimated input by straight averaging of the outputs of all equalizers calculated via (24) for Channel I. Fig. 9 depicts eye diagrams for one realization of $N = 500$ samples, illustrating how well the RLS algorithm equalizes the output of the channels when SNR = 25 dB. RLS recursions in (39) were initialized using the batch estimate in (20) with $N =$

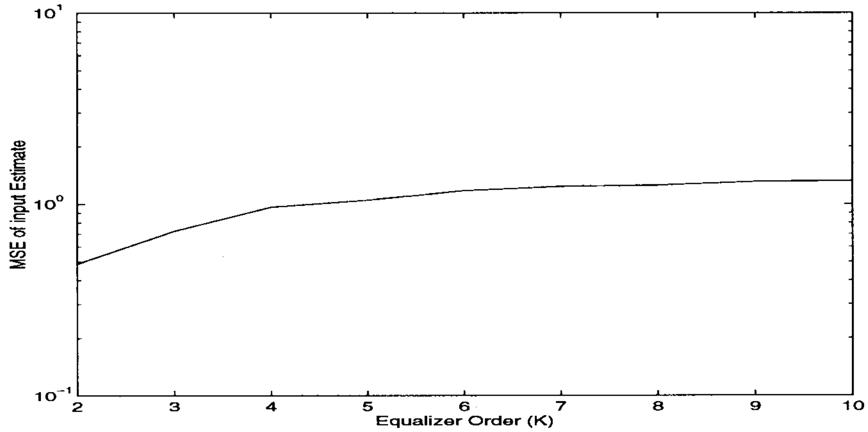

Fig. 7. Effects of the equalizer order.

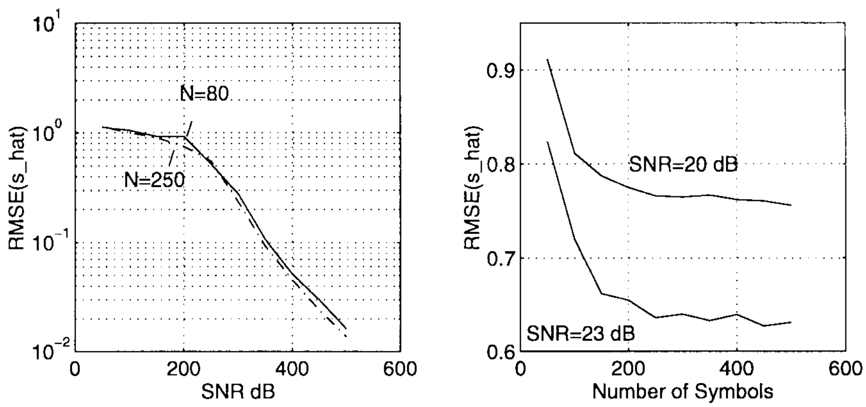

Fig. 8. Performance of the batch method.

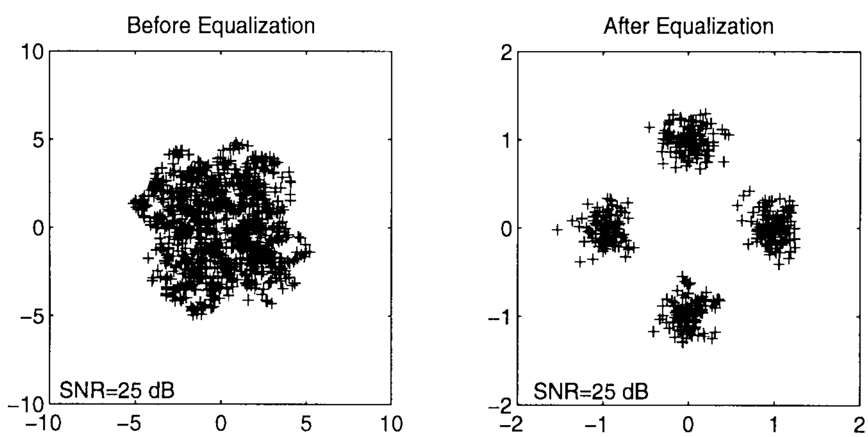

Fig. 9. Eye diagrams for the RLS algorithm initialized with the batch estimate.

$2M(K+1) = 16$, which is the minimum number of samples required for the batch method to work. How well the RLS improves on the batch estimate with the number of iterations is illustrated in Fig. 10. RMSE$(\hat{\mathbf{g}})$ at iteration t is defined as $||\mathbf{g}_t - \hat{\mathbf{g}}_t||/||\mathbf{g}_t||$ averaged over 100 realizations, where $\mathbf{g}_t$ and $\hat{\mathbf{g}}_t$ are the true and estimated equalizer coefficients at iteration t, respectively; see (40). Here, the channel was time-invariant, so we used $\lambda = 1$.

Test Case 3—Adaptive LMS Equalizer: In Fig. 11, the LMS learning curve (initialized with the batch estimate with $N = 16$) illustrates the tracking of the true equalizer after 3 is subtracted from the real part of the last coefficient of h_1 of channel I at the 200th iteration. In Fig. 12, both plots are averages over 100 realizations, and SNR $= 22$ dB. Stepsize $\mu = 0.009$ was used in (39).

Test Case 4—Comparison with the Deterministic Method in [20]: To illustrate the suitability of the linear equalizers proposed here for adaptive equalization, we compared our method with XLTK. Among the many other blind methods we could pick, we chose XLTK because it is also a "deterministic"

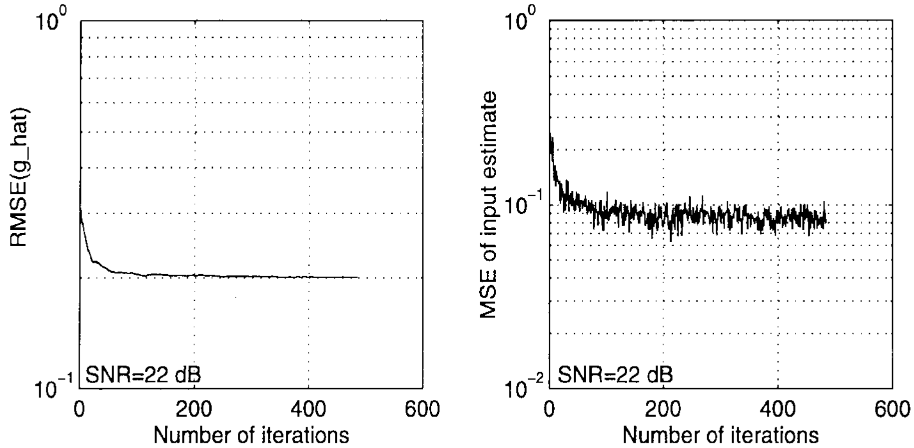

Fig. 10. Performance of the RLS with the number of iterations.

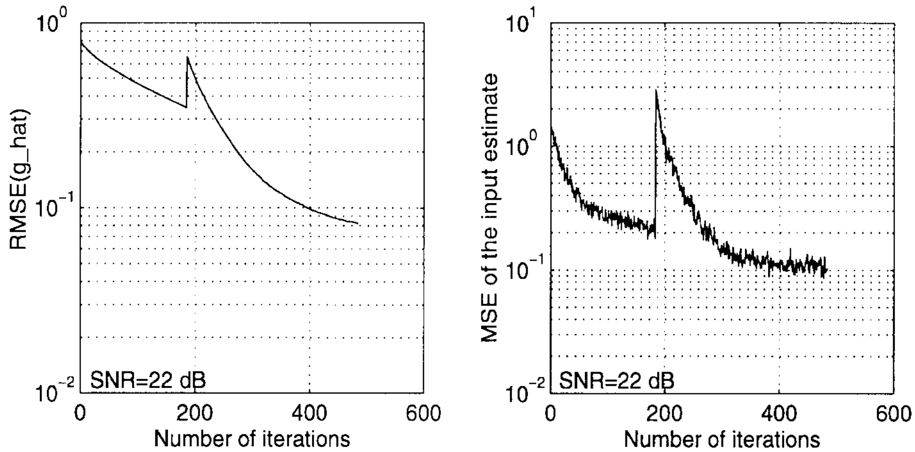

Fig. 11. LMS with one coefficient abruptly changing at the 200th iteration.

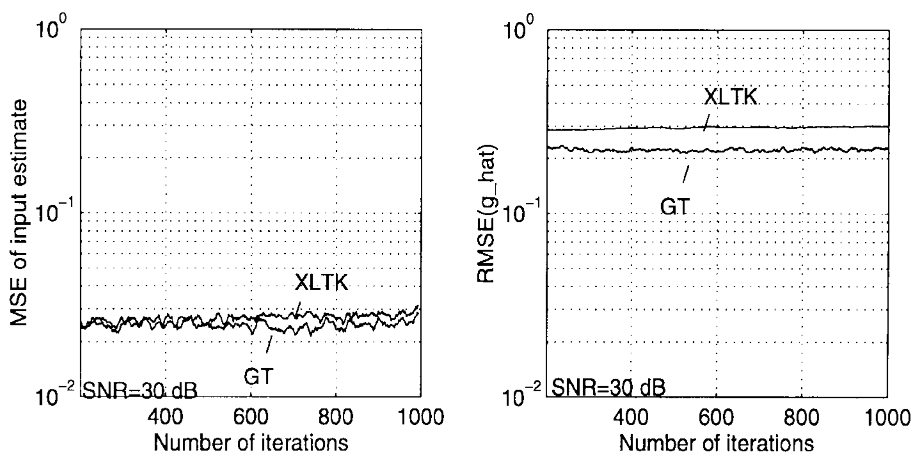

Fig. 12. Comparison with XLTK on a slowly varying channel.

method of comparable complexity, unlike [10], which is another deterministic algorithm that requires an SVD every time a portion of the input sequence is to be estimated; hence, it is computationally very demanding. XLTK is a channel identification method. The equalizers we use for the XLTK are obtained by substituting the estimates of the channel in (5) and then taking the columns of the right inverse of this matrix. We used the RLS algorithm in (39) to adaptively update the minimum and the maximum delay equalizers and observed how their performance compares with updating the batch estimate of [20] every 200 symbols. We allowed the channel coefficients $h_1(0)$ and $\text{Re}(h_2(1))$ in Channel I to vary as a Gaussian AR process with mean as listed in Table I (AR coefficients were 1 and -0.01, and the innovation variance

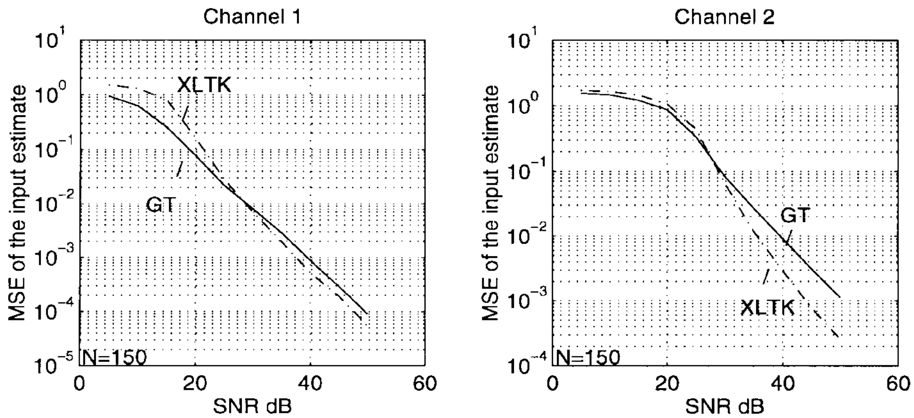

Fig. 13. Comparison with XLTK of weighted average of all equalizers on the two channels.

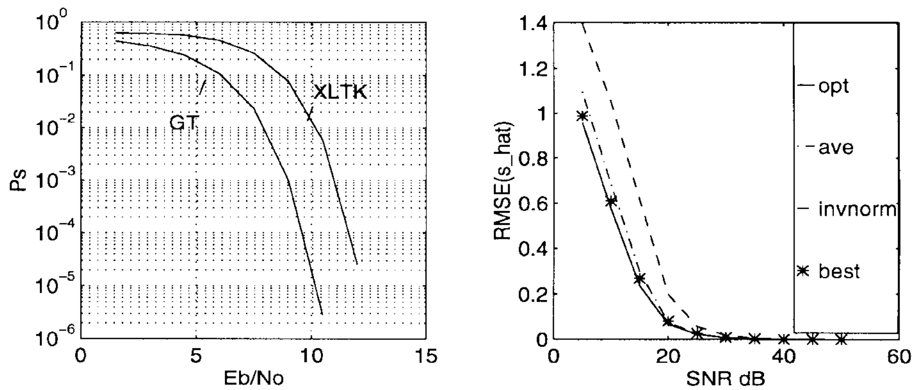

Fig. 14. (left) Ps comparison with XLTK and (right) different choice of weights for GT.

was 1). For such a setup, the MSE of the input estimate and the error in the norm of the equalizer estimates are plotted in Fig. 12 against the number of iterations at SNR $= 30$ dB and $\lambda = 0.95$. To make the difference between the two overlapping MSE curves visible, we averaged them with a rectangular window of length 10. This figure illustrates that in addition to the savings in computational complexity, the adaptive approach, by updating the equalizers at every iteration, can perform better than XLTK.

The simulations have shown that minimum- and maximum-delay equalizer estimates with the proposed method do slightly worse, in general, than that of [20]. However, when the estimates of all the equalizers are combined, at SNR's < 30 dB, our method performs slightly better. Fig. 13 illustrates this effect with the two different sets of channels. In this case, the equalizers' estimates are combined with weights given by (36) for both the XLTK and our method.

The left plot in Fig. 14 illustrates the superiority of our method in terms of probability of symbol error for Channel I. A data length of $N = 250$ was used over 10 000 realizations. On the right, the effect of different choices for $\mathbf{w}$ is illustrated. These are in increasing performance: $\mathbf{w} = \mathbf{1}$ (ave for average); $\mathbf{w}$ such that the ith component $w_i = 1/\|\mathbf{g}_i\|$ (invnorm for inverse of the norm); $\mathbf{w} = \mathbf{e}_i$, where $\mathbf{g}_i$ is the equalizer with minimum norm (best equalizer in the sense of MMSE); and $\mathbf{w}$ corresponding to (36) (opt). Especially as the SNR increases, $\mathbf{w}$ in (36) yields the smallest RMSE. It is worth noting that since all the equalizer estimates use the same data, they are strongly correlated, and hence, (unweighted) averaging does not necessarily improve performance over the single equalizer with the minimum norm.

VI. Conclusion

Diversity available in the form of multiple outputs makes blind equalization of single-input coprime channels possible under minimal persistence of excitation (p.e.) conditions. The required data length for identification is smaller than most methods in the literature and even smaller than most deterministic methods due to the pairwise (or single) equalizer estimation scheme. When the entire record of data is available, simple blind solutions result for the equalizers corresponding to the minimum and maximum delays. Simultaneous solution of the latter requires stricter p.e. conditions when estimates rely on a segment of the output data. Complexity increases when equalizers corresponding to all possible shifts are estimated, but all the information about the channel is recovered. In addition, extra averaging improves RMSE of the coefficient estimates and the equalized output in the presence of additive white noise. The linear equation form of the direct blind equalizer estimates lends itself naturally to adaptive solutions. Optimum combinations of the equalizer outputs improve performance over existing algorithms in terms of mean-square error and probability of error in the equalized data.

REFERENCES

[1] T. J. Endres, S. D. Halford, C. R. Johnson Jr., and G. B. Giannakis, "Simulated comparisons of blind equalization algorithms for cold start-up applications," *Int. J. Adaptive Contr. Signal Process.*, vol. 12, no. 3, pp. 283–301, May 1998.

[2] D. Gesbert, P. Duhamel, and S. Mayrargue, "Subspace-based adaptive algorithms for the blind equalization of multichannel fir filters," in *Proc. EUSIPCO*, 1994.

[3] D. Gesbert, P. Duhamel, and S. Mayrargue, "On-line blind multichannel equalization based on mutually referenced filters," *IEEE Trans. Signal Processing*, vol. 45, pp. 2307–2317, Sept. 1997.[3]

[4] G. B. Giannakis, "Blind equalization of time varying channels: A deterministic multichannel approach," in *Proc. 8th Signal Process. Workshop Stat. Signal Array Process.*, Corfu, Greece, June 24–26, 1996, pp. 180–183.

[5] G. B. Giannakis and S. Halford, "Blind fractionally-spaced equalization of noisy FIR channels: Direct and adaptive solutions," *IEEE Trans. Signal Processing*, vol. 45, pp. 2277–2292, Sept. 1997.

[6] G. B. Giannakis and C. Tepedelenlioğlu, "Batch and adaptive direct blind equalizers of multiple FIR channels: A deterministic approach," in *Proc. 30th Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, Nov. 3–6, 1996, vol. 1, pp. 290–294.

[7] F. Gini and G. B. Giannakis, "Frequency offset and symbol timing recovery in flat-fading channels: A cyclostationary approach," *IEEE Trans. Commun.*, vol. 46, pp. 400–412, Mar. 1998.

[8] S. Haykin, *Adaptive Filter Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1986.

[9] ——, *Blind Deconvolution*. Englewood Cliffs, NJ: Prentice-Hall, 1994.

[10] H. Liu and G. Xu, "A deterministic approach to blind symbol estimation," *IEEE Signal Processing Lett.*, vol. 1, pp. 205–207, Dec. 1994.

[11] K. Abed Meraim *et al.*, "Prediction error methods for time-domain blind identification of multichannel FIR filters," in *Proc. ICASSP*, 1995, vol. 3, pp. 1968–1971.

[12] 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.

[13] A. Scaglione, G. B. Giannakis, and S. Barbarossa, "Self-recovering multirate equalizers using redundant filterbank precoders," in *Proc. Int. Conf. Acoust., Speech, Signal Process., (ICASSP)*, Seattle, WA, May 12–15, 1998, vol. VI, pp. 3501–3504.

[14] D. T. M. Slock, "Blind fractionally-spaced equalization, perfect-reconstruction filter banks and multichannel linear prediction," in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, Adelaide, Australia 1994, vol. IV, pp. 585–588.

[15] D. T. M. Slock and C. B. Papadias, "Further results on blind identification and equalization of multiple FIR channels," in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, Detroit, MI, May 8–12, 1995, vol. 3, pp. 1972–1975.

[16] T. Söderström and P. Stoica, *System Identification*. London, U.K.: Prentice-Hall, 1989.

[17] L. Tong, G. Xu, and T. Kailath, "Blind identification and equalization based on second-order statistics: A time domain approach," *IEEE Trans. Inform. Theory*, vol. 40, pp. 340–349, 1994.

[18] L. Tong, G. Xu, B. Hassibi, and T. Kailath, "Blind channel identification based on second-order statistics: A frequency-domain approach," *IEEE Trans. Inform. Theory*, vol. 41, pp. 329–334, Jan. 1995.

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

[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.

[3] Appeared after the original submission of the present manuscript in Mar. 1997 and its conference version [6] in Nov. 1996.

Georgios B. Giannakis (F'97) received the Diploma in electrical engineering from the National Technical University of Athens, Athens, Greece in 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 University of Virginia, Charlottesville, in September 1987, where he is now a Professor with the Department of Electrical Engineering. His general interests lie in the areas of signal processing, communications estimation and detection theory, and system identification. Specific research areas of current interest include diversity techniques for channel equalization and multi-user communications, nonstationary and cyclostationary signal analysis, wavelets in statistical signal processing, and non-Gaussian signal processing with applications to SAR, 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 co-organized the 1993 IEEE Signal Processing Workshop on Higher Order Statistics, the 1996 IEEE Workshop on Statistical Signal and Array Processing, and the first IEEE Signal Processing Workshop on Wireless Communications in 1997. He guest (co)-edited two special issues on high-order statistics (*International Journal of Adaptive Control and Signal Processing* and the EURASIP journal *Signal Processing*) and the January 1997 Special Issue on Signal Processing for Advanced 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, a secretary of the Signal Processing Conference Board, a member of the SP Publications board, and a Member and Vice-Chair of the SSAP Technical Committee. He now chairs the Signal Processing for Communications Technical Committee. He is also a member of the IMS and the European Association for Signal Processing.

Cihan Tepedelenlioğlu was born in Ankara, Turkey, in 1973. He received the B.S. degree with highest honors from the Florida Institute of Technology, Melbourne, in 1995 and the M.S. degree from the University of Virginia, Charlottesville, in 1998, both in electrical engineering.

Since Fall 1995, he has been working as a Teaching and Research Assistant at the University of Virginia and is currently a Ph.D. student in the CCSP laboratory. His research interests include statistical signal processing, system identification, time-varying systems, and equalization of fading channels in digital communications.