

Blind Identification With Periodic Modulation: A Time-Domain Approach

Ching An Lin and Jwo Yuh Wu

Abstract—We propose a method for blind identification of finite impulse response (FIR) channels with periodic modulation. The time-domain formulation in terms of block signals is simple compared with existing frequency-domain formulations. It is shown that the linear equations relating the products of channel coefficients and the autocorrelation matrix of the received signal can be further arranged into decoupled groups. The arrangement reduces computations and improves accuracy of the solution; it also leads to very simple identifiability conditions and a very natural formulation of the optimal modulating sequence selection problem. The proposed optimal selection minimizes the effects of channel noise and error in autocorrelation matrix estimation; it results in a consistent channel estimate when the channel noise is white. Simulation results show that the method yields good performance: It compares favorably with an existing subspace modulation-induced-cyclostationarity method, and it is robust with respect to channel order overestimation. The effect of modulation period and threshold of the modulating sequence are also discussed.

Index Terms—Blind identification, periodic modulation precoder, transmitter induced cyclostationarity.

I. INTRODUCTION

BLIND identification and equalization of finite-impulse-response (FIR) channels that exploit cyclostationarity of second-order statistics of the received data was first proposed by Tong *et al.* [16]. Various schemes have since been proposed [2]–[5], [8], [11]–[15]. See [5] for detailed references. Cyclostationarity can be induced at the receiver or at the transmitter. While receiver-induced cyclostationarity has always resulted from oversampling [11], many different schemes have been proposed to induce cyclostationarity at the transmitter. They include periodic modulation [2], [12], repetition coding [15], and combinations of repetition and modulation [4] and filterbank precoding [13]. A general analysis of the precoding frameworks is in [13]. A related performance analysis of these so-called transmitter-induced-cyclostationarity precoders can be found in [3].

In this paper, we study the problem of blind channel identification with periodic modulation of source symbols. In contrast to the frequency-domain approaches in [2] and [12], we formu-

late the problem in the time domain and in terms of block signals. The method exploits the linear relation between the products of channel coefficients and the autocorrelation matrix of the received signal and computes the products first by solving a set of linear equations. The channel coefficients are then obtained (to within a scalar ambiguity) by computing the dominant eigenvector of an associated Hermitian matrix. We note that similar “bilinear” approach is also used in [8]. We show that the set of linear equations relating the products of coefficients and the autocorrelation matrix can be further arranged into decoupled groups. The arrangement reduces computations and improves accuracy of the solution; it also leads to very simple identifiability conditions, which depend on the modulating sequence alone, and a very natural formulation of the optimal modulating sequence selection problem. We note that identifiability conditions based on the modulating sequence alone are also reported in [2]. The proposed optimal selection minimizes the effects of channel noise and error in autocorrelation matrix estimation. Moreover, the resultant channel estimate is consistent when the channel noise is white.

The paper is organized as follows. Section II is the problem statement and preliminary. Section III establishes the identifiability conditions, proposes an identification algorithm, and discusses numerical aspects associated with it. In Section IV, the problem of selecting the modulating sequence is formulated and solved. In Section V, simulation examples are given to illustrate the performance of the proposed method. Section VI contains conclusions.

II. PROBLEM STATEMENT AND PRELIMINARY

A. Problem Statement

We consider the baseband transmission system. The source symbol sequence $s(n)$ is modulated by a (real) periodic sequence $p(n)$ with period N to obtain the modulated sequence

$$w(n) = p(n)s(n) \quad (2.1)$$

which is then pulse shaped and transmitted through the “composite” channel including the transmit filter, the channel, and the receiving filter. The received continuous-time signal is sampled at the symbol rate. A commonly used discrete-time model of the channel characteristic is an FIR filter [1]. The input–output relation is described by

$$z(n) = \sum_{l=0}^L h(l)w(n-l) \quad (2.2)$$

Manuscript received December 20, 2001; revised June 24, 2002. This work was supported by the National Science Council under Grant NSC-89-2213-E009-218. The associate editor coordinating the review of this paper and approving it for publication was Prof. Nicholas D. Sidiropoulos.

The authors are with the Department of Electrical and Control Engineering, National Chiao-Tung University, Hsinchu, Taiwan, R.O.C. (e-mail: calin@cc.nctu.edu.tw; u8712803@cc.nctu.edu.tw).

Digital Object Identifier 10.1109/TSP.2002.804145.

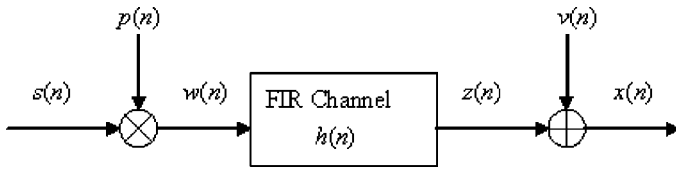


Fig. 1. Discrete-time baseband model.

where the sequence $h(n)$ is the impulse response of the channel, and L is the channel order. The received signal sequence $x(n)$ is the sum of filtered signal $z(n)$ and an additive noise, that is

$$x(n) = z(n) + v(n) \quad (2.3)$$

where $v(n)$ is the channel noise. A schematic description of (2.2) and (2.3) is shown in Fig. 1.

In this paper, we propose a method for identifying $h(n)$ using second-order statistics of the received data $x(n)$ and discuss an optimal design of the modulating sequence $p(n)$. The following assumptions are made in the sequel. We note that essentially the same assumptions are made in [2] and [12].

- The source sequence $s(n)$ is zero mean, uncorrelated, and $E s(k)s(l)^* = \delta(k-l)$, where $E y$ denotes the expectation of random variable y , and $\delta(\cdot)$ is the Kronecker delta function.
- The channel noise $v(n)$ is stationary with zero mean and is uncorrelated with source sequence $s(n)$.
- An upper bound \hat{L} on channel order L is known, and the period is $N > \hat{L} + 1$.
- The receiver is synchronized with the transmitter.

B. Preliminary

Define the block signal

$$\bar{x}(n) := [x(Nn)x(Nn+1) \cdots x(Nn+N-1)]^T \in C^N \quad (2.4)$$

and let the block signals $\bar{w}(n)$, $\bar{s}(n)$, $\bar{z}(n)$, and $\bar{v}(n)$ be similarly defined. From (2.1), we have

$$\bar{w}(n) = G\bar{s}(n) \quad (2.5)$$

where G is a diagonal matrix

$$G := \text{diag}\{p(0) \cdots p(N-1)\} \in \mathfrak{R}^{N \times N}. \quad (2.6)$$

In terms of the block signals, we write the discrete-time model (2.2) and (2.3) as

$$\bar{x}(n) = \bar{z}(n) + \bar{v}(n) \quad (2.7)$$

where

$$\bar{z}(n) = H_0 G \bar{s}(n) + H_1 G \bar{s}(n-1), \quad n \geq 1 \quad (2.8)$$

in which $H_0 \in C^{N \times N}$ is a lower triangular Toeplitz matrix with

$$[h(0) \cdots h(L)0 \cdots 0]^T \quad (2.9a)$$

as its first column, and $H_1 \in C^{N \times N}$ is an upper triangular

Toeplitz matrix with

$$[0 \cdots 0h(L) \cdots h(1)] \quad (2.9b)$$

as its first row. We note that i) the input–output relation between $x(n)$ and $s(n)$ are periodically time-varying, whereas (2.7), in terms of block signals, is time-invariant; ii) the splitting of H_0 and H_1 as lower triangular and upper triangular matrices, respectively, is possible by the assumption $N > \hat{L} + 1$.

III. CHANNEL IDENTIFICATION

A. Identification Equations: Noise-Free Case

We will first consider the noise-free case, i.e., $x(n) = z(n)$ for all n . We assume for the moment that the channel order L is known and consider the general case later. Since $\bar{x}(n) = \bar{z}(n)$ and by assumption a), the autocorrelation matrix of $\bar{x}(n)$ is computed from (2.8) as

$$R_{\bar{x}}(0) := E\bar{x}(n)\bar{x}(n)^* = H_0 G^2 H_0^* + H_1 G^2 H_1^* \in C^{N \times N} \quad (3.1)$$

where the matrices H_0 and H_1 , defined in (2.9), contain the unknown channel impulse response to be identified. Given the matrix $R_{\bar{x}}(0)$, (3.1) defines a set of N^2 nonlinear equations in the unknowns $h(0), \dots, h(L)$. However, if we consider the products of channel coefficients $h(k)h(l)^*$ as unknowns, (3.1) becomes a set of N^2 linear equations. Hence, instead of solving for $h(0), \dots, h(L)$ directly, we propose to first compute the products $h(k)h(l)^*$, $0 \leq k, l \leq L$. Since $R_{\bar{x}}(0)$ is Hermitian, there are only $N(N+1)/2$ independent equations in (3.1). Thus, we will consider the upper triangular part of $R_{\bar{x}}(0)$. The Toeplitz structure of matrices H_0 and H_1 allows us to simplify the equations even further. We will see that these $N(N+1)/2$ equations can be divided into $L+1$ decoupled groups of equations of smaller dimensions. The reduction in dimension reduces the amount of computations, especially when N is large, and improves numerical robustness with respect to noise and error in the estimation of $R_{\bar{x}}(0)$.

Define the matrix

$$J := \begin{bmatrix} 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix} \in \mathfrak{R}^{N \times N}. \quad (3.2)$$

Thus, for any $i \geq 1$, all entries of J^i are zero, except that those on the i th lower diagonal are equal to one. From (2.9), the matrices H_0 and H_1 can be, respectively, expressed as

$$H_0 = \sum_{k=0}^L h(k)J^k \quad \text{and} \quad H_1 = \sum_{m=1}^L h(m)(J^T)^{N-m}. \quad (3.3)$$

From (3.3), it follows that

$$H_0 G^2 H_0^* = \sum_{k=0}^L \sum_{l=0}^L h(k)h(l)^* J^k G^2 (J^T)^l \quad (3.4a)$$

$$H_1 G^2 H_1^* = \sum_{m=1}^L \sum_{n=1}^L h(m)h(n)^* (J^T)^{N-m} G^2 J^{N-n}. \quad (3.4b)$$

From (3.4), the matrix $R_{\bar{x}}(0)$ can be written as the weighted sum of the *known* matrices $J^k G^2 (J^T)^l$ and $(J^T)^{N-k} G^2 J^{N-l}$, with the *unknown* channel coefficients of the form $h(k)h(l)^*$ as weighting factors. In the following proposition, we will see that these two matrices are upper triangular for $l \geq k$ and lower triangular with zero diagonal entries for $l < k$. Thus, we only have to consider those terms in (3.4) with $l \geq k$. Moreover, all entries of both matrices are zero, except those on one of the upper diagonals. This allows us to rearrange the independent equations in the upper triangular part of (3.1) into decoupled sets of equations. The proof of the proposition is given in Appendix A.

For any $N \times N$ complex matrix $M = [m_{kl}]_{0 \leq k, l \leq N-1}$, define, for $0 \leq j \leq N-1$, $\Gamma_j[M] := [m_{0j} m_{1(j+1)} \cdots m_{(N-1-j)(N-1)}]^T \in C^{N-j}$. Namely, $\Gamma_j[M]$ is the vector containing the entries on the j th upper diagonal of M . Here, we identify the diagonal with $j = 0$.

Proposition 3.1: Let G and J be the $N \times N$ matrices, respectively, defined in (2.6) and (3.2). Let $1 \leq k, l \leq L$ be two non-negative integers. For each k , we have the following results.

- 1) If $l = k + j$, where $0 \leq j \leq L - k$, both $J^k G^2 (J^T)^l$ and $(J^T)^{N-k} G^2 J^{N-l}$ are upper triangular, with only the respective j th upper diagonals nonzero, and

$$\begin{aligned} \Gamma_j[J^k G^2 (J^T)^l] \\ = [0 \quad \cdots \quad 0 p(0)^2 \quad \cdots \quad p(N-1-k-j)^2]^T \in \Re^{N-j}. \end{aligned} \quad (3.5)$$

and

$$\begin{aligned} \Gamma_j[(J^T)^{N-k} G^2 J^{N-l}] \\ = [p(N-k)^2 \quad \cdots \quad p(N-1)^2 0 \quad \cdots \quad 0]^T \in \Re^{N-j}. \end{aligned} \quad (3.6)$$

- 2) If $l < k$, both $J^k G^2 (J^T)^l$ and $(J^T)^{N-k} G^2 J^{N-l}$ are lower triangular with zero diagonal entries.

□

Comments:

- 1) For $k = 0$, the right-hand side (RHS) of (3.5) becomes

$$[p(0)^2 \quad p(1)^2 \quad \cdots \quad p(N-2-j)^2 \quad p(N-1-j)^2]^T \quad (3.7)$$

whereas that of (3.6) is a zero vector.

- 2) We note that the vector on the RHS of (3.5) has the first k entries zero, whereas that in (3.6) has only the first k entries nonzero. Thus, there is no overlap between the locations of nonzero entries of the vectors given in the RHS of (3.5) and (3.6). In particular, we have (3.8), shown at the bottom of the page. □

Proposition 3.1 enables us to divide the upper triangular part of (3.1) into $L+1$ decoupled groups of equations based on each upper diagonal of $R_{\bar{x}}(0)$. More precisely, from (3.1), (3.4), and

Proposition 3.1, we have, for *each* $0 \leq j \leq N-1$

$$\begin{aligned} \Gamma_j[R_{\bar{x}}(0)] &= \sum_{k=0}^L \sum_{l=0}^L h(k)h(l)^* \Gamma_j[J^k G^2 (J^T)^l] \\ &\quad + \sum_{m=1}^L \sum_{n=1}^L h(m)h(n)^* \Gamma_j[(J^T)^{N-m} G^2 J^{N-n}] \\ &= \sum_{k=0}^{L-j} h(k)h(k+j)^* \Gamma_j[J^k G^2 (J^T)^{k+j}] \\ &\quad + \sum_{m=1}^{L-j} h(m)h(m+j)^* \Gamma_j[(J^T)^{N-m} G^2 J^{N-m-j}]. \end{aligned} \quad (3.9)$$

From (3.9), the $N-j$ equations defined by the j th upper diagonal of $R_{\bar{x}}(0)$ depend only on the $L-j+1$ unknowns of the form $h(k)h(k+j)^*$. Thus, if $j > L$, that is, if j is greater than the channel order, then $\Gamma_j[R_{\bar{x}}(0)]$ is a zero vector since $h(k)h(k+j)^* = 0$ for all $k \geq 0$ and $j > L$. Hence, we only have to consider $0 \leq j \leq L$ in (3.9). Combine the two summations on the RHS of (3.9) to obtain

$$\begin{aligned} \Gamma_j[R_{\bar{x}}(0)] &= h(0)h(j)^* \Gamma_j[G^2 (J^T)^j] \\ &\quad + \sum_{k=1}^{L-j} h(k)h(k+j)^* (\Gamma_j[J^k G^2 (J^T)^{k+j}] \\ &\quad + \Gamma_j[(J^T)^{N-k} G^2 J^{N-k-j}]). \end{aligned} \quad (3.10)$$

From (3.10), it follows that the vector $\Gamma_j[R_{\bar{x}}(0)]$ can be written as a linear combination of $L-j+1$ columns of the forms (3.7) and (3.8). To write (3.10) in matrix form, we define, for each $0 \leq j \leq L$

$$f_j := [h(0)h(j)^* \quad h(1)h(j+1)^* \quad \cdots \quad h(L-j)h(L)^*]^T \in C^{L-j+1}. \quad (3.11)$$

Thus, f_j contains all the unknown product coefficients of the form $h(k)h(k+j)^*$, $0 \leq k \leq L-j$. With (3.7) and (3.8), (3.10) can be expressed as

$$\Gamma_j[R_{\bar{x}}(0)] = M_j f_j \quad (3.12)$$

where, for $0 \leq j \leq L$, we have (3.13), shown at the bottom of the next page.

We note that the matrix M_0 is an $N \times (L+1)$ circulant matrix with first column equal to

$$q := [p(0)^2 \quad p(1)^2 \quad \cdots \quad p(N-2)^2 \quad p(N-1)^2]^T \quad (3.14)$$

and, for $1 \leq j \leq L$, M_j is obtained from M_0 by deleting its last j rows and last j columns.

Hence, the upper triangular part of matrix (3.1) is rearranged as $L+1$ decoupled groups of linear equations given in the matrix form (3.12), where the j th group of equations depends only on

$$\Gamma_j[J^k G^2 (J^T)^l] + \Gamma_j[(J^T)^{N-k} G^2 J^{N-l}] = [p(N-k)^2 \quad \cdots \quad p(N-1)^2 \quad p(0)^2 \quad \cdots \quad p(N-1-k-j)^2]^T. \quad (3.8)$$

the unknowns of the form $h(k)h(k+j)^*$. The proposed channel identification method is based on solving (3.12) and is discussed next.

Remark: Note that if the source sequence $s(n)$ is correlated, the equations defined by the j th upper diagonal of $R_{\bar{x}}(0)$ will involve additional unknowns $h(k)h^*(l)$ for $l \neq k+j$. In this situation, we can no longer arrange the $L+1$ equations defined by the upper triangular part of the matrix $R_{\bar{x}}(0)$ into $L+1$ decoupled groups, as shown in (3.12). The computation of the product channel coefficients $h(k)h^*(l)$ instead requires solving a single least squares problem of relatively large dimension. \square

B. Identifiability Condition

Consider the $L+1$ decoupled groups of equations in (3.12). Since $N > L+1$, each group of equations is overdetermined and consistent. Assume that for each $0 \leq j \leq L$, the matrix M_j is of full column rank. Then, the vector f_j , which contains unknowns of the form $h(k)h(l)^*$ with $l = k+j$, can be uniquely determined as

$$f_j = (M_j^T M_j)^{-1} M_j^T \Gamma_j [R_{\bar{x}}(0)]. \quad (3.15)$$

Successively using (3.15) for $j = 0, \dots, L$, we can obtain $h(k)h(l)^*$, $0 \leq k \leq l \leq L$. To identify the channel, that is, to determine the unknowns $h(0), \dots, h(L)$ up to a scalar ambiguity, form the $(L+1) \times (L+1)$ Hermitian matrix

$$Q = [q_{kl}]_{0 \leq k, l \leq L}, \quad \text{where } q_{kl} = h(k)h(l)^*. \quad (3.16)$$

We note that the j th upper diagonal vector of the matrix Q is simply f_j , that is, $\Gamma_j[Q] = f_j$ for $0 \leq j \leq L$. In the ideal case, the matrix Q is of rank one and can be factorized as $Q = hh^*$, where

$$h := [h(0) \quad \dots \quad h(L)]^T \in C^{L+1} \quad (3.17)$$

is the vector containing the channel impulse response. Thus, the channel is identified, up to a scalar ambiguity, by computing the unit-norm eigenvector associated with the maximal eigenvalue of the matrix Q . Hence, a sufficient condition for channel identifiability is that each M_j defined in (3.13) is of full column rank. We note that the matrices M_j are completely determined by the modulating sequence $p(n)$. By appropriately selecting $p(n)$, we can make M_j full rank and well conditioned.

C. On Channel Order Overestimation

The previous analysis is based on the assumption that the channel order L is known. In case that only an upper bound

$\hat{L} \geq L$ is available, with $N > \hat{L} + 1$, (3.15) will give the product coefficients $h(k)h(l)^*$ for $0 \leq k \leq l \leq \hat{L}$. The associated $(\hat{L} + 1) \times (\hat{L} + 1)$ matrix Q constructed as in (3.16) will theoretically have only one nonzero eigenvalue with the corresponding eigenvector given by

$$\alpha [h(0)h(1) \quad \dots \quad h(L)0 \quad \dots \quad 0]^T \in C^{\hat{L}+1} \quad (3.18)$$

where α is a scalar. Thus, it follows from (3.18) that the actual channel order and impulse response can be determined by the most significant entries in the unit-norm eigenvector associated with the maximal eigenvalue of Q . Hence, the proposed method is applicable, as long as an upper bound $\hat{L} \geq L$ is known; channel order overestimation, however, does increase the amount of computations involved.

D. Identification Algorithm and Computational Aspects

We summarize the proposed channel identification method as an algorithm.

Channel Identification Algorithm:

- 1) Select a modulating sequence $p(n)$ such that each matrix M_j defined in (3.13) is of full column rank.
- 2) Estimate the autocorrelation matrix $R_{\bar{x}}(0)$ via the time average

$$\hat{R}_{\bar{x}}(0) = \frac{1}{K} \sum_{i=1}^K \bar{x}(i)\bar{x}(i)^* \quad (3.19)$$

where K is the number of data blocks.

- 3) Compute the product channel coefficients $h(k)h(l)^*$, $0 \leq k \leq l \leq L$ using (3.15).
- 4) Form the matrix Q as in (3.16), and compute the channel impulse response vector h as the unit-norm eigenvector associated with the maximal eigenvalue of Q .

Some of computational aspects are discussed in the following.

1) *Least Squares Solution (3.15):* Since each group of equations in (3.12) is overdetermined and consistent, (3.15) will give the exact solution as long as the autocorrelation matrix $R_{\bar{x}}(0)$ is obtained *exactly* and that there is no noise. In practice, when

$$M_j = \begin{bmatrix} p(0)^2 & p(N-1)^2 & p(N-2)^2 & \dots & p(N-L-j)^2 \\ p(1)^2 & p(0)^2 & p(N-1)^2 & \dots & p(N-L-j+1)^2 \\ p(2)^2 & p(1)^2 & p(0)^2 & \dots & p(N-L-j+2)^2 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ p(N-3-j)^2 & p(N-4-j)^2 & p(N-5-j)^2 & \dots & p(N-3-L)^2 \\ p(N-2-j)^2 & p(N-3-j)^2 & p(N-4-j)^2 & \dots & p(N-2-L)^2 \\ p(N-1-j)^2 & p(N-2-j)^2 & p(N-3-j)^2 & \dots & p(N-1-L)^2 \end{bmatrix} \in \Re^{(N-j) \times (L-j+1)}. \quad (3.13)$$

only a time average $\hat{R}_{\bar{x}}(0)$ is available, the f_j in (3.15) is the least squares solution of (3.12).

2) *On Computation of f_0 Using (3.15)*: We note that from (3.11), the product coefficient vector f_0 contains unknowns of the form $|h(k)|^2$, for $0 \leq k \leq L$. However, the least squares solution (3.15) for $j = 0$ does not necessarily yield an f_0 with non-negative entries. If some entries of f_0 computed by (3.15) are negative, it is an indication that either the signal-to-noise (SNR) ratio is fairly small or that the quality of the estimated autocorrelation data is poor. A direct remedy for the latter case is to improve the quality of estimation by using more data samples. If, however, the cause is due to low SNR or a longer data record for improving the estimation quality is unavailable, an alternative is to compute f_0 using the non-negative least squares method [9] to remove the inconsistency. In such a situation, however, our simulation results show that the non-negative least squares method does not actually improve the overall estimation accuracy. A plausible reason is that the available data is ‘‘inherently bad.’’ Hence, although this alternative avoids negative values of the elements in the computed coefficient vector f_0 , it does not seem to be a good choice in practice since it does not seem to improve accuracy and increases the algorithm complexity.

3) *On Selection of Modulation Period N* : The number of equations in each group of (3.12) increases with the period N ; the j th group has $N - j$ equations. If the equation errors in (3.12) resulting from noise and imperfect estimation of $\hat{R}_{\bar{x}}(0)$ can be modeled as a zero mean white noise with fixed variance, then the error covariance can be made arbitrary small if the number of equations is sufficiently large [10, p. 178]. Thus, it seems desirable to choose large modulation period since more equations tend to improve the accuracy of the least squares solutions (3.15). However, for a fixed number of data samples S , the number of available data blocks for estimating $R_{\bar{x}}(0)$ is approximately S/N . Therefore, if N is large, we will have a small number of data blocks and, hence, usually a less accurate $\hat{R}_{\bar{x}}(0)$. Simulations show that different choices of N ($\geq \hat{L} + 2$) yield similar performance. It thus seems reasonable to choose $N = \hat{L} + 2$ to reduce computation.

4) *On Condition of the Matrix*: The accuracy of solution (3.15) is determined by the numerical condition of the matrix M_j . Since the construction of M_j entirely depends on $p(n)$, the condition of M_j is thus closely related to the selection of $p(n)$. We discuss how to choose $p(n)$ to improve numerical robustness in the next section.

5) *Computational Complexity*: Compared with the structured subspace method [2], [14] and the one-cycle subspace method [12], the proposed method requires fewer computations than the former and more computations than the latter. Detailed flop counts for these methods are given in Appendix E.

IV. OPTIMAL SELECTION OF THE MODULATING SEQUENCE $p(n)$

We consider the general case, that is, the channel noise is present, and discuss the problem of selecting the modulating sequence $p(n)$. We first propose an optimality criterion to select $p(n)$ to reduce the effect of noise. We will find a class of solu-

tions that are optimal for noise attenuation. Among this class of solutions, we then choose the ‘‘best’’ $p(n)$ with which the channel coefficients can be most reliably computed. Finally, we show that with the optimal $p(n)$, the resultant channel estimate is consistent. Effects of the modulation period and the threshold of the modulating sequence are also discussed.

A. Optimality Criteria

Assume that the additive channel noise is white. Then, from (2.7), (2.8), and assumption b), the autocorrelation matrix is

$$R_{\bar{x}}(0) = R_z(0) + \sigma^2 I_N \quad (4.1a)$$

where

$$R_z(0) = H_0 G^2 H_0^* + H_1 G^2 H_1^*. \quad (4.1b)$$

From (4.1), noise has contribution to only the diagonal entries of $R_{\bar{x}}(0)$, and thus, we have

$$\Gamma_0[R_{\bar{x}}(0)] = \Gamma_0[R_z(0)] + \sigma^2 b \quad (4.2)$$

where

$$b := [1 \quad \dots \quad 1]^T \in \Re^N. \quad (4.3)$$

With (4.1b) and (4.2), the $L+1$ decoupled groups of equations in (3.12) remains unchanged, except that the $j = 0$ group becomes

$$\Gamma_0[R_{\bar{x}}(0)] = M_0 f_0 + \sigma^2 b. \quad (4.4)$$

Since σ^2 is unknown, the actual product coefficient vector f_0 cannot be determined using (4.4). Instead, given $\Gamma_0[R_{\bar{x}}(0)]$, we solve the least squares solution \hat{f}_0 of the inconsistent equation $\Gamma_0[R_{\bar{x}}(0)] \approx M_0 f$. From (4.4), f_0 can be written as

$$\hat{f}_0 = (M_0^T M_0)^{-1} M_0^T \Gamma_0[R_{\bar{x}}(0)] = f_0 + \sigma^2 (M_0^T M_0)^{-1} M_0^T b. \quad (4.5)$$

Thus, the least squares solution \hat{f}_0 consists of the actual f_0 plus an additional perturbation term due to noise. From (4.5), the noise contribution is eliminated, that is, $\hat{f}_0 = f_0$, if and only if $M_0^T b = 0$, i.e., the vector b is orthogonal to $R(M_0)$, which is the range space of M_0 . Hence, if the modulating sequence $p(n)$ can be selected to achieve the above orthogonality condition, the effect of noise is completely eliminated. This turns out to be impossible since entries of M_0 are all non-negative. The best we can hope is to choose $p(n)$ as to make b and $R(M_0)$ as close to being orthogonal as possible. This suggests the following optimization scheme to select $p(n)$.

Define the quantity

$$\gamma := \frac{q^T b}{\|q\|_2 \cdot \|b\|_2} \quad (4.6)$$

where $\|\cdot\|_2$ denotes the 2-norm, vector b is defined in (4.3), and q defined in (3.14) is the first column of the matrix M_0 . Then, from (3.14), $\gamma > 0$ is a function of the modulating sequence $p(n)$. Let \tilde{q} be an arbitrary column of M_0 rather than the first. Since, from (3.13), M_0 is circulant, we have $\|\tilde{q}\|_2 = \|q\|_2$ and

¹Without lose of generality, we assume that the channel order L is known.

$\hat{q}^T b = q^T b$. Thus, the quantity γ defined in (4.6) is the cosine of the angle between the vector b and any column of M_0 and can be viewed as a measure of orthogonality between b and $R(M_0)$.

If γ is small for some choice of $p(n)$, then with this choice, the corresponding $R(M_0)$ is close to being orthogonal with b and, thus, will result in a small noise contribution to \hat{f}_0 in (4.5). Small γ also means that the projection of b onto $R(M_0)$ is small. If we think of $R(M_0)$ as the signal space, this would imply that the contribution of noise in the received data is small from the identification point of view. Therefore, we should keep γ as small as possible. Specifically, we propose to select the modulating sequence $p(n)$ to minimize γ with respect to all $p(n)$ satisfying

$$\frac{1}{N} \sum_{n=0}^{N-1} |p(n)|^2 = 1 \quad (4.7a)$$

and

$$|p(n)|^2 \geq \delta > 0, \quad \forall 0 \leq n \leq N-1. \quad (4.7b)$$

Constraint (4.7a) normalizes the average transmission power to unity since the sequence $w(n) = p(n)s(n)$ is actually transmitted, and the source sequence $s(n)$ is with unit variance. The constraint (4.7b) is necessary for equalization since, at the receiver end, it is impossible to recover a source symbol that is modulated with a zero value since the symbols are uncorrelated. Constraints (4.7) require that $\delta < 1$. We note that these two constraints are also used in [2], in which different optimality criteria are used.

By definitions of b and q , we have

$$q^T b = \sum_{n=0}^{N-1} |p(n)|^2, \quad \text{and} \quad \|b\|_2 = \sqrt{N}. \quad (4.8)$$

With (4.7a) and (4.8), the quantity γ defined in (4.6) can be expressed as

$$\gamma = \frac{\sqrt{N}}{\|q\|_2} = \sqrt{\frac{N}{\sum_{n=0}^{N-1} |p(n)|^4}}. \quad (4.9)$$

We note that the constraint (4.7a) is equivalent to

$$\|q\|_1 = \sum_{n=0}^{N-1} |p(n)|^2 = N \quad (4.10a)$$

where $\|\cdot\|_1$ denotes the 1-norm. In addition, if we let $q_n := |p(n)|^2$ be the n th component of q , $0 \leq n \leq N-1$, the constraint (4.7b) can be rewritten as

$$q_n \geq \delta > 0, \quad \forall 0 \leq n \leq N-1. \quad (4.10b)$$

Since N is fixed, the proposed optimality criterion for selecting $p(n)$ is equivalent to the following quadratic optimization problem.

(P): Maximize $\|q\|_2$ subject to the constraints (4.10a) and (4.10b).

The optimal solution to **(P)** is given in the following proposition.

Proposition 4.1: Let the integer $0 \leq m \leq N-1$ be fixed but arbitrary. Then, the vector q with entries

$$q_m = N(1-\delta) + \delta, \quad \text{and} \\ q_n = \delta, \quad 0 \leq n \leq N-1, \quad n \neq m$$

is an optimal solution to problem **(P)**. \square

[Proof]: Define $y := q - \Delta$, where $\Delta := [\delta \ \dots \ \delta]^T$, and let y_n be the n th component of y . Then, the problem is equivalent to maximizing $\|y + \Delta\|_2$ subject to

$$\|y\|_1 = N(1-\delta) \quad \text{and} \quad y_n \geq 0. \quad (4.11)$$

Since $y_n \geq 0$, we have

$$\|y + \Delta\|_2^2 = \|y\|_2^2 + 2N\delta - N\delta^2.$$

Since N and δ are fixed, our problem is the same as maximizing $\|y\|_2$ subject to (4.11). Since $\|y\|_2 \leq \|y\|_1$ for any y , the maximum $\|y\|_2$ is achieved if we can find some y that satisfies (4.11) and with which the equality holds. This is the case if we choose, for any fixed m , $0 \leq m \leq N-1$, $y_m = N(1-\delta)$ and $y_n = 0$ for $n \neq m$. The result thus follows. \square

From Proposition 4.1, the optimal modulating sequence is immediately given by, for any fixed $0 \leq m \leq N-1$

$$|p(m)| = \sqrt{N(1-\delta) + \delta}, \quad \text{and} \\ |p(n)| = \sqrt{\delta} \quad \text{for } n \neq m \quad (4.12)$$

where the magnitudes of the sequence assumes two values with one and only one peak. We should note that the optimal modulating sequence (4.12), which is obtained by minimizing the effect of (white) channel noise, is the same as the one reported in [2, Prop. 3]. In [2], this solution is obtained by maximizing the so-called degree-of-cyclostationarity $|F_2(k)|$ [2, p. 1577], subject to the same constraints. Since $|F_2(k)|$ can be regarded as a measure of SNR from a frequency domain point of view (large value of this quantity results in relatively large value of the cyclic correlations coefficients with respect to noise level), it is not unreasonable to expect that the two solutions should be the same; the choices in (4.12) tend to keep the noise effect on the signal component as small as possible.

With (4.12), the minimal value of the orthogonality measure γ is computed as

$$\bar{\gamma} := 1/\sqrt{N(1-\delta)^2 + \delta(2-\delta)}. \quad (4.13)$$

We note that $\bar{\gamma}$ is indeed independent of the index m at which the peak occurs. We should note that different index m , however, will result in different matrices M_j in (3.13). Hence, the choice of m is crucial to the properties of M_j . The issue of selecting m to obtain M_j with good numerical property is addressed in the next subsection.

Note that if we modulate the source $s(n)$ with the optimal sequence (4.12), there will be periodic peak value in the transmitting power. In case that the peak power due to modulation is greater than the maximal allowable power provided by the transmitter, we can reduce this peak value by imposing an upper

bound on the magnitude of the modulation sequence. This suggests to us that we should modify constraint (4.7b) as follows. For any $0 \leq n \leq N - 1$

$$\delta \leq |p(n)|^2 \leq \rho, \quad \text{for some } \rho \geq 1. \quad (4.14)$$

The resultant optimal modulating sequence can be obtained by solving problem **(P)** with constraint (4.10b) replaced by (4.14) (recall that $q_n = |p(n)|^2$). We should note that when no upper bound on $|p(n)|^2$ is imposed, (4.10a) and the condition $\delta \leq |p(n)|^2$ together imply $|p(n)|^2 \leq N - (N - 1)\delta, \forall 0 \leq n \leq N - 1$. As a result, for a given δ , if the upper bound ρ in (4.14) is chosen such that $\rho > N - (N - 1)\delta$, then it is inactive. The optimization problem, in this case, reduces to the original one, and the solution is given as in Proposition 4.1. Assume $\rho \leq N - (N - 1)\delta$. For the case of $N = 2$, it is easy to check that the optimal solution for $\{p(0), p(1)\}$ is $\{\sqrt{\rho}, \sqrt{2 - \rho}\}$ or $\{\sqrt{2 - \rho}, \sqrt{\rho}\}$. It appears that one component of the optimal solution attains the upper bound $\sqrt{\rho}$. For the general case, we conjecture that the optimal $p(n)$ is given by

$$\begin{aligned} p(m) &= \sqrt{\rho}, \quad \text{and} \\ p(n) &= \sqrt{(N - \rho)/(N - 1)}, \quad 0 \leq n \leq N - 1, \quad n \neq m \end{aligned} \quad (4.15)$$

where $0 \leq m \leq N - 1$ is fixed but arbitrary. In fact, if we form the Lagrangian function associated with the constrained optimization problem, it can be shown (see Appendix B) that $p(n)$ in (4.15) satisfies the first-order necessary condition. Numerical experiments (see Simulation 8) also tend to indicate that the two-level form solution in (4.15) is indeed the optimal solution.

B. On Selection of m

Assume that $p(n)$ is chosen as in (4.12). For each $0 \leq m \leq N - 1$, consider the associated matrices M_j in (3.13) for $0 \leq j \leq L$. For each j , let κ_j be the condition number [7] of the matrix $M_j^T M_j$. If κ_j is large, the matrix $M_j^T M_j$ is ill-conditioned, and the corresponding least squares estimate (3.15) is sensitive to data errors. Let

$$\mu_m := \max_{0 \leq j \leq L} \kappa_j \quad (4.16)$$

be the largest condition number among all $M_j^T M_j$ associated with m . If μ_m is large, then with this m in (4.12), we tend to have an ill-conditioned $M_j^T M_j$ for some j and, hence, a less-accurate least squares estimate (3.15). This suggests that we should consider each $\mu_m, 0 \leq m \leq N - 1$, and among them, we select the ‘‘optimal’’ m as the one whose corresponding μ_m is minimal. With the special form of $p(n)$ in (4.12), the procedure for finding the optimal m can be further simplified. Specifically, we will see in the next lemma that there exist some choices of m such that at least one associated M_j will lose rank. Such choices of m should be excluded since they will prevent channel identifiability. Moreover, with the special form (4.12), it can be shown that for certain ‘‘feasible’’ m , the corresponding μ_m are the same. With the aid of these facts, the optimal choice of m can

be determined without computing μ_m for all $0 \leq m \leq N - 1$. The proof of lemma is given in Appendix C.

Lemma 4.2: Assume that the modulating sequence $p(n)$ is chosen as in (4.12). For $0 \leq j \leq L$, let $M_j \in \mathfrak{R}^{(N-j) \times (L-j+1)}$ defined in (3.13) be the matrices associated with the index m . In addition, let μ_m defined in (4.16) be the maximal condition number among the $M_j^T M_j$ matrices. Then, we have the following results.

- 1) If $0 \leq m \leq N - L$ or $m = N - 1$, all the resultant M_j are of full column rank.
- 2) In particular, for each $0 \leq m \leq N - L - 1$, the corresponding μ_m is completely determined by the matrix M_0 and is equal to $\mu_m = 1 + ((L + 1)\delta(2 - \delta))/(N(1 - \delta)^2)$.
- 3) If $N - L + 1 \leq m \leq N - 2$, at least one M_j is not of full column rank. \square

From Lemma 4.2, we should only consider those $p(n)$ in (4.12) with $0 \leq m \leq N - L$ and $m = N - 1$. Since μ_m is the same for $0 \leq m \leq N - L - 1$, we will simply identify μ_m with μ_0 for all such m . The optimal m_{opt} is thus given by

$$m_{\text{opt}} = \arg \min \{\mu_0, \mu_{N-L}, \mu_{N-1}\}. \quad (4.17)$$

Henceforth, we will restrict the modulating sequence $p(n)$ to be of the form (4.12) with the index m determined by (4.17).

Remark: If the noise is colored and has unknown nonzero correlations σ_j^2 for time lags $0 \leq j \leq j_1$, then the first $j_1 + 1$ groups of equations in (3.12) becomes

$$\begin{aligned} \Gamma_j[R_{\tilde{x}}(0)] &= \Gamma_j[R_{\tilde{z}}(0)] + \sigma_j^2 b \\ &= M_j f_j + \sigma_j^2 b, \quad 0 \leq j \leq j_1. \end{aligned} \quad (4.18)$$

In this case, our choice of $p(n)$, although not optimal for $j \geq 1$, is still a good candidate selection for the least squares estimate (3.15). This is because all the resultant $M_j^T M_j$ matrices still remain well conditioned. Thus, the corresponding estimate \hat{f}_j is expected to be relatively insensitive to data errors due to noise. Simulation results (see Simulation 3) seem to indicate that at reasonable levels of SNR, the proposed selection of $p(n)$ indeed works well against color noise. \square

C. Consistency of Channel Estimation: White Noise Case

In this subsection, we will show that if the channel noise is white and the optimal modulating sequence given in (4.12) (with m determined by (4.17)) is used, the resultant channel estimate is consistent.

Recall that when there is no noise, the diagonal vector of the rank-one matrix $Q(=hh^*)$ defined in (3.16) equals f_0 , that is, $\Gamma_0[Q] = f_0$. When the channel noise is white, it follows from (4.5) that the resultant ‘‘perturbed’’ Q matrix, say, Q_w , is equal to

$$Q_w = Q + D \quad (4.19)$$

where $D \in \mathfrak{R}^{N \times N}$ is a noise perturbation matrix that is diagonal such that

$$\Gamma_0[D] = \sigma^2 (M_0^T M_0)^{-1} M_0^T b. \quad (4.20)$$

It is shown in Appendix D that if $p(n)$ is an admissible² two-level modulating sequence such that, for any fixed $0 \leq k \leq N-1$

$$\begin{aligned} |p(k)| &= \sqrt{\alpha}, \quad \text{and} \\ |p(n)| &= \sqrt{\beta}, \quad \text{for } 0 \leq n \leq N-1, \quad n \neq k \end{aligned} \quad (4.21)$$

then the corresponding Q_w matrix is of the form

$$Q_w = Q + \zeta I_{L+1} = hh^* + \zeta I_{L+1} \quad (4.22)$$

for some $\zeta > 0$. That is, as long as $p(n)$ is chosen of the form (4.21), the resultant (white) noise perturbation matrix D reduces to a scalar multiple of the $(L+1) \times (L+1)$ identity matrix. Since such perturbations preserve eigenvectors, in particular the one associated with the largest eigenvalue, the channel vector h can still be obtained, up to a scalar ambiguity, by computing the eigenvector associated with the largest eigenvalue of the Q_w matrix. The optimal $p(n)$ in (4.12), which is a special case of (4.21), hence yields the channel estimate consistent and, at the same time, achieves the largest noise reduction.

Remarks:

- a) We should note that if $p(n)$ is not of the form (4.21), then the diagonal entries of D will assume *at least* two values, viz., the matrix D is no longer a scalar multiple of the identity matrix. Such a perturbation will not preserve the eigenvectors of the (noise-free) Q matrix and, hence, results in an inconsistent channel estimate.
- b) If the channel noise is colored, the channel estimate is inconsistent even if a modulating sequence of the form (4.21) is used. This is because the noise perturbation matrix D will contain nonzero off-diagonal entries [this follows from (4.18)] and does not preserve the eigenvectors of Q either. In this case, however, the optimal $p(n)$ in (4.12) still seems to be a good choice as far as noise effect on the estimated channel is concerned. \square

D. Effects of N and δ

In this subsection, with optimal modulating sequence in (4.12), we discuss the effects of modulation period N and the threshold δ on the minimal orthogonality measure $\bar{\gamma}$ defined in (4.13).

Suppose $0 < \delta < 1$ is fixed. It follows from (4.13) that $\bar{\gamma}$ decreases as the period N increases and $\bar{\gamma} \rightarrow 0$ as $N \rightarrow \infty$. Thus, identification performance is improved as modulation period N is increasing. However, with $p(n)$ in (4.12), large modulation period N leads to large transmission power peak. This is undesirable since, in practice, there is a constraint on maximal available power provided by the transmitter. Moreover, as discussed in Section III-D, when the number of data samples is fixed, large N will lead to poor estimation of $\hat{R}_{\bar{x}}(0)$. Thus, a large modulation period should be avoided. Based on simulations, if the channel order is L , the selection $N = L + 2$ suffices to yield a satisfactory performance.

To see the effect of δ , rewrite $\bar{\gamma}$ in (4.13) as $\bar{\gamma} = 1/\sqrt{N - (N-1)\delta(2-\delta)}$. For $0 < \delta < 1$, $\delta(2-\delta)$ is an increasing function in δ . Thus, for a fixed period N , the

²By this, we mean the constraints (4.7a) and (4.7b) are satisfied.

minimal value $\bar{\gamma}$ decreases as δ is decreased. We should also note that at the equalization stage, small values of δ may cause large symbol error rate even if the channel is perfectly equalized [2, p. 1578]. This imposes a tradeoff on the selection of δ .

V. SIMULATION RESULTS

To illustrate the performance of the proposed channel identification method, we consider the five-tap channel used in [2]

$$\begin{aligned} h &= [0.459 + 0.265j, -0.2078 - 0.12j, -0.4677 \\ &\quad - 0.277j, 0.0953 + 0.055j, -0.0312 - 0.018j]^T. \end{aligned}$$

The input source symbols are drawn from an i.i.d. QPSK constellation. The additive channel noise $v(n)$ is white with a Gaussian distribution. As channel identification performance measure, we consider the normalized root-mean-square error³ (NRMSE) defined by [2]

$$\text{NRMSE} := \frac{1}{\|h\|_2} \sqrt{\frac{1}{I} \sum_{i=1}^I \|\hat{h}^{(i)} - h\|_2^2} \quad (5.1)$$

where I is the number of Monte Carlo runs, and $\hat{h}^{(i)}$ is the estimate of the channel impulse response vector in i th trial with $\hat{h}_l^{(i)}$ as the l th component. Since $\hat{h}^{(i)}$ can only be estimated within an unknown scalar ambiguity, for the purpose of computing NRMSE, we compute the unknown scalar by performing a least square fit of the estimated channel to the actual one. This technique is used in [8] to remove the scalar ambiguity. The signal-to-noise-ratio (SNR) is defined as

$$\text{SNR} := \sqrt{\frac{\frac{1}{N} \sum_{n=0}^{N-1} E\{|z(n)|^2\}}{E\{|v(n)|^2\}}}. \quad (5.2)$$

For Simulations 1–6, 100 Monte Carlo runs are conducted, i.e., $I = 100$.

1) *Simulation 1—Optimal Selection of Periodic Sequence $p(n)$:* The effects of selecting $p(n)$ on the performance of the proposed channel identification method are demonstrated. In the following two experiments, we set $N = 6$, $\delta = 0.5978$, and SNR is fixed at 10 dB. In the first experiment, with $p(n)$ chosen as in (4.12), we illustrate the resultant performances obtained from different m . For each $0 \leq m \leq 5$, form the associated matrices M_j , $0 \leq j \leq 4$ according to (3.13). Then, from (4.16) and by computation, we have $\{\mu_0, \dots, \mu_5\} = \{5.0713, 5.0713, 109.9, \infty, \infty, 109.9\}$. Fig. 2 shows the respective channel NRMSE versus numbers of samples (S) for $m = 0, 1, 4, 5$. Note that for $m = 0$ and $m = 1$, the corresponding performance is almost identical and that the degradation in performance for $m = 4$ and 5 , owing to the ill-conditioned M_j . In the second experiment, we will see the effectiveness of the optimal noise-attenuation selection (4.12). We consider $p(n)$ in (4.12) with $m = 0$, which is $\{p(0), \dots, p(5)\} = \{1.75, 0.7665, 0.7665, 0.7665, 0.7665, 0.7665\}$ and another plausible

³Another commonly used performance measure is the average bias [2]. In our simulation, this quantity exhibits the same tendency as the NRMSE and, thus, is not displayed.

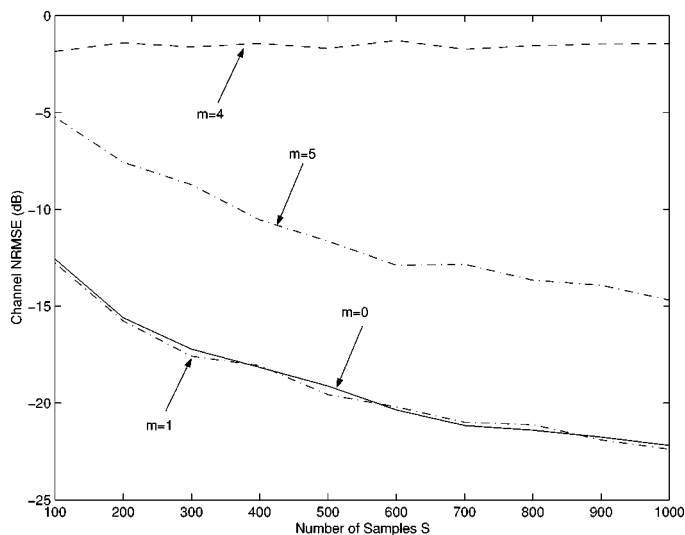


Fig. 2. Channel NRMSE. $p(n)$ in (4.12) with different choices of m .

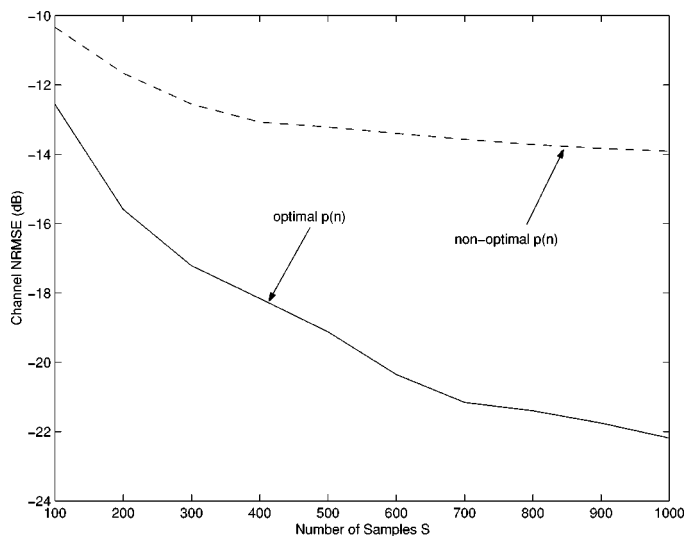


Fig. 3. Channel NRMSE. Different selections of $p(n)$.

two-level nonoptimal selection chosen as $\{p(0), \dots, p(5)\} = \{0.7665, 1.0404, 1.0404, 1.0404, 1.0404, 1.0404\}$, which also satisfies constraint (4.7) with the same $\delta = 0.5878$ and is selected in analogous way as in [2]. We note that both the optimal $p(n)$ and the comparative selection are of the form (4.21) and result in a consistent channel estimate. Fig. 3 shows the corresponding NRMSE of the estimated channels versus numbers of samples. As one can see, the optimal $p(\cdot)$ gives significantly better performance.

2) *Simulation 2—Comparations With Existing Subspace Approaches:* The performance of the proposed method is compared with those of the one cycle subspace method [12] and the structured subspace method [2], [14]. In all cases, we set $N = 6, \delta = 0.5878$, and choose $p(n)$ as in (4.12) with $m = 0$. We note that $\delta = 0.5878$ is the choice used for simulation in [2]. Fig. 4 shows the corresponding NRMSE versus numbers of samples for fixed SNR = 10 dB. Fig. 5 shows the corresponding NRMSE versus SNR, where the number of samples is fixed at

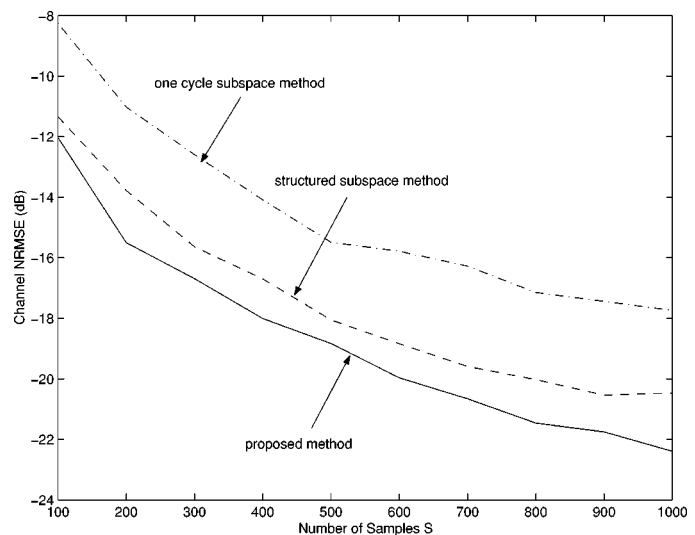


Fig. 4. Channel NRMSE. Proposed method versus subspace methods (white noise case).

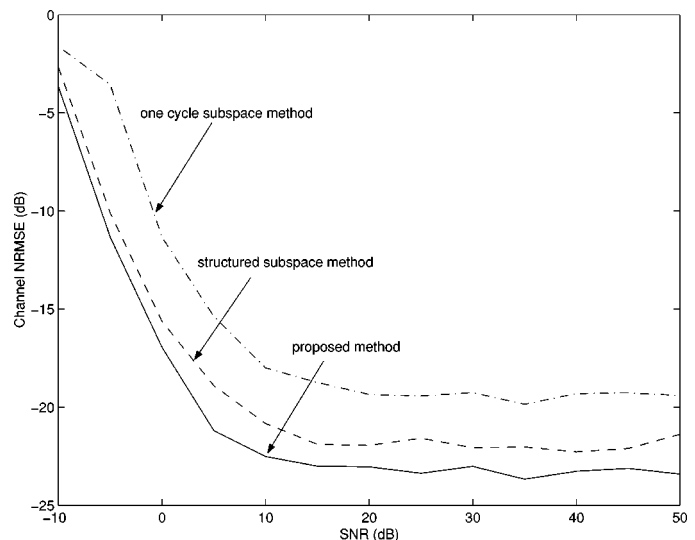


Fig. 5. Channel NRMSE versus SNR. Proposed method versus subspace methods (white noise case).

$S = 1000$. The results show that the proposed method gives better results when compared with the two subspace methods.

3) *Simulation 3—Robustness to Additive Color Noise:* We demonstrate the robustness of the proposed method when the channel noise is colored. In this simulation, the additive color noise $v(n)$ is generated by filtering a white noise sequence $v_w(n)$ using the second-order FIR filter $C(z) = 0.6 + 0.6z^{-1} + 0.5292z^{-2}$, that is, $v = [C(z)]v_w$. Fig. 6 shows the NRMSE versus SNR, computed, respectively, using the proposed method and the two subspace methods. The number of samples is fixed at $S = 1000$, and number of trials is $I = 100$. In all methods, we set $N = 6, \delta = 0.5878$ and choose $p(n)$ as in (4.12) with $m = 0$. The result shows that, for SNR < 5 dB, the subspace method in [2] achieves best performance. This is mainly because it provides a consistent channel estimate, irrespective of color noise, as opposed to our method. We note that although the method in [12] also preserves consistency in the colored noise case, it exploits,

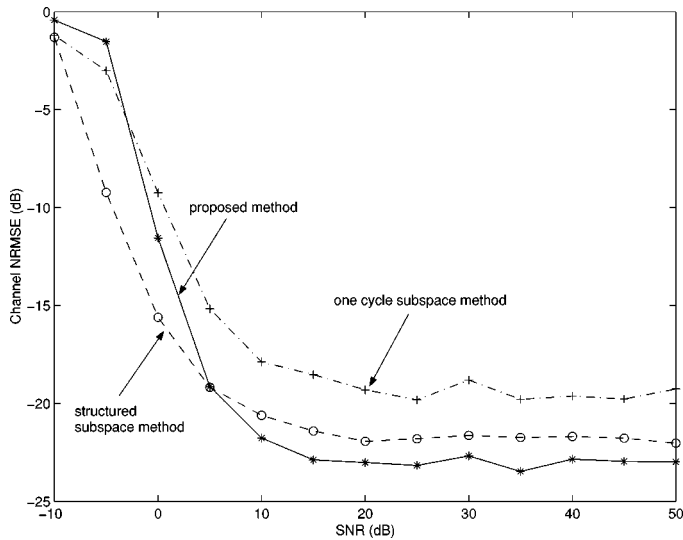


Fig. 6. Channel NRMSE versus SNR. Proposed method versus subspace methods (color noise case).

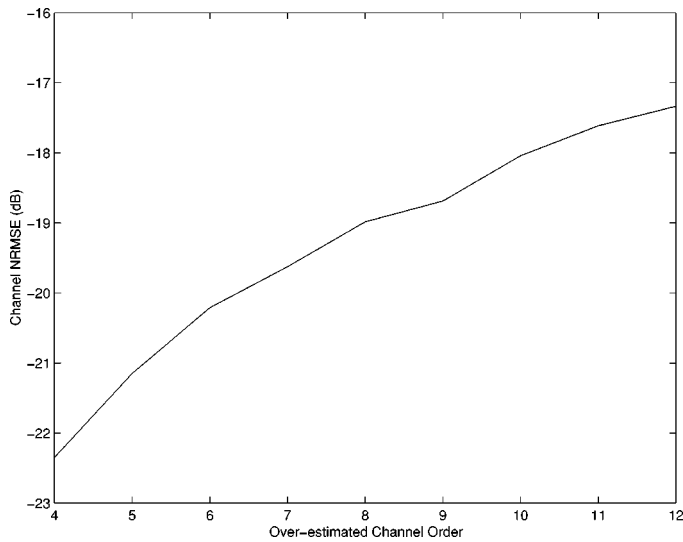


Fig. 7. Channel NRMSE versus overestimated channel order.

however, only the cyclic correlation of only one nonzero cycle for identification and, hence, leads to degeneration in performance. For $\text{SNR} > 5$ dB, the proposed method results in better performance. A reasonable explanation is that the well-conditioned M_j matrices tend to reduce the noise effect on the estimated channel coefficients.

4) *Simulation 4—Robustness to Channel Order Overestimation:* We test the proposed method when channel order is overestimated. For each upper bound \hat{L} , $4 \leq \hat{L} \leq 12$, we choose the modulation period $N = \hat{L} + 2$. The SNR is fixed at 10 dB, and number of samples at $S = 1000$. Fig. 7 shows that the proposed method is quite robust to channel order overestimation; the NRMSE increases less than 5 dB as \hat{L} increases from 4 to 12.

5) *Simulation 5—Effects of δ and N on Performance of the Proposed Method:* For $p(n)$ in (4.12) with $m = 0$, we illustrate the effects of δ and N on the performance. We fix $\text{SNR} = 10$ dB. For fixed $N = 6$, Fig. 8 shows the channel NRMSE versus numbers of samples corresponding to three different thresholds

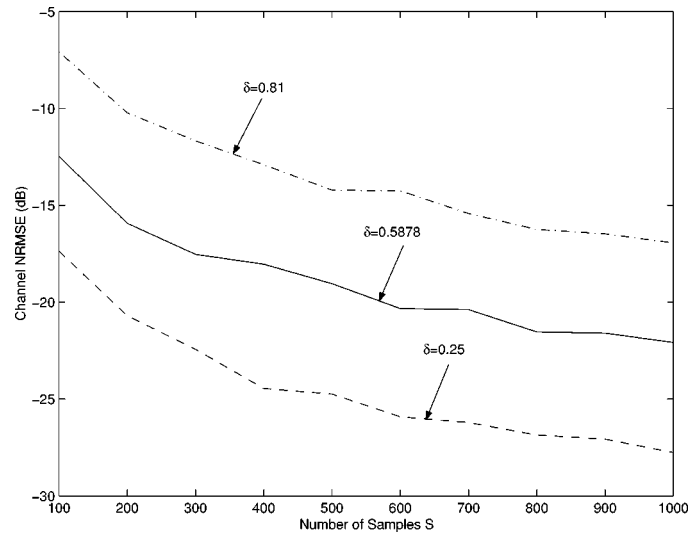


Fig. 8. Channel NRMSE. $N = 6$ with different choices of δ .

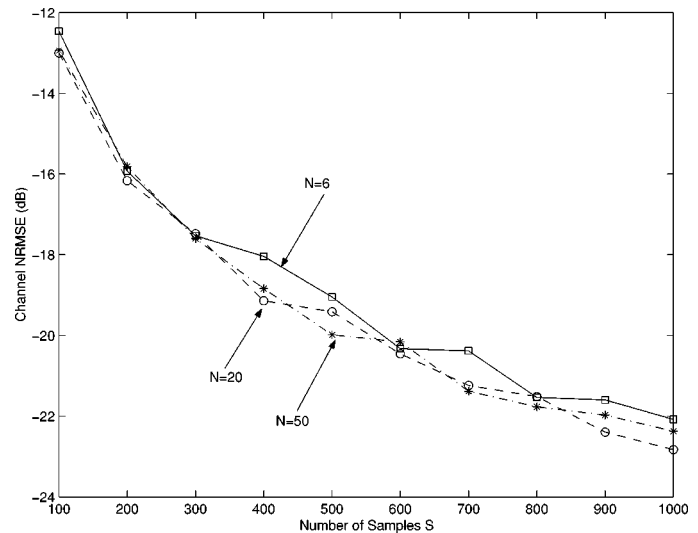


Fig. 9. Channel NRMSE. $\delta = 0.5878$ with different choices of N .

$\delta = 0.81$, $\delta = 0.5878$, and $\delta = 0.25$. For these three δ , the respective minimal orthogonality measures $\bar{\gamma}$ computed using (4.13) are 0.9204, 0.7353, and 0.5121. The result shows that small δ indeed leads to improved performance. For fixed $\delta = 0.5878$, Fig. 9 shows the channel NRMSE versus number of samples corresponding to modulation periods $N = 6$, $N = 20$, and $N = 50$, with which the respective minimal $\bar{\gamma}$ are 0.7353, 0.4863, and 0.3275. From the figure, it can be seen that the performance corresponding to the three different N are roughly the same. This is because the estimation quality of $\hat{R}_{\bar{x}}(0)$ is poor as N increases. This demerit may cancel out the benefit from large noise attenuation when a large N is used. Since increasing N will also increase the amount of computation and, at the same time, result in a large transmission power peak, the modulation period $N = \hat{L} + 2$ seems to be a desirable choice.

6) *Simulation 6—On Large Sample Performance:* In this simulation, we demonstrate the behavior of the channel NRMSE computed using the proposed method when the number of samples S is large. With $\delta = 0.5878$, Fig. 10 shows the NRMSE versus number of samples S (for $S = 1000$ – $50\,000$)

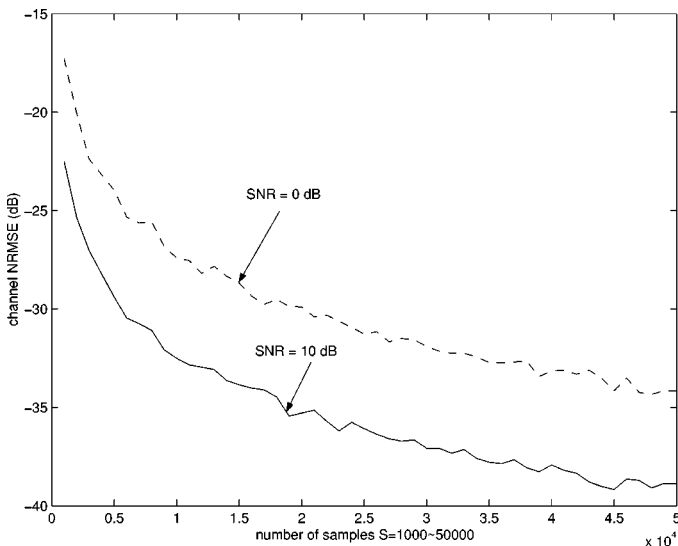


Fig. 10. Channel NRMSE for large numbers of data samples.

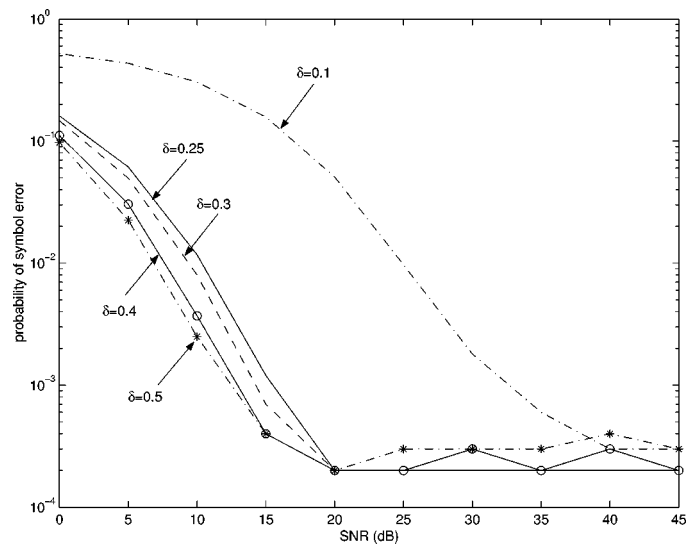


Fig. 12. Probability of symbol error with different values of δ (0.1–0.5).

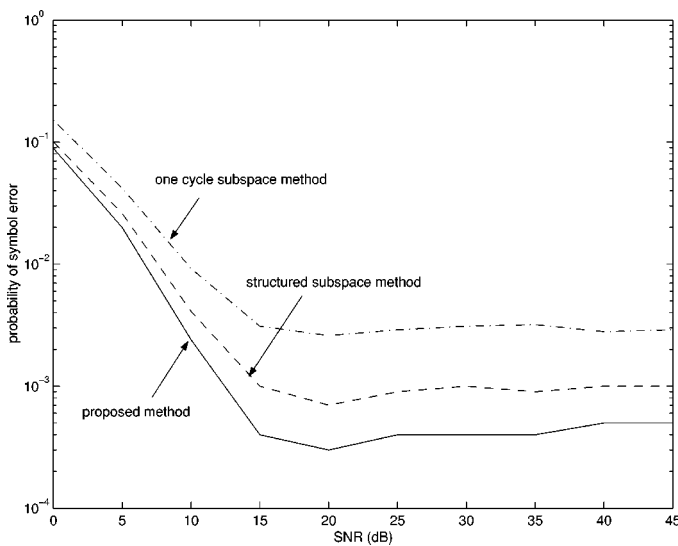


Fig. 11. Probability of symbol error. Proposed method versus subspace methods.

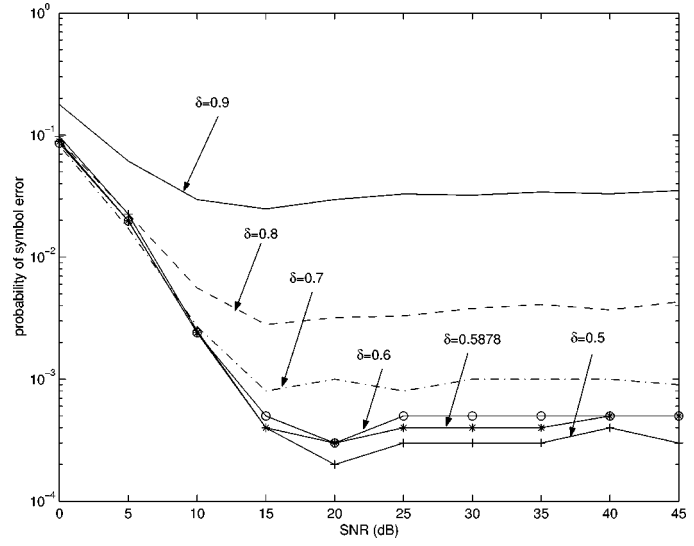


Fig. 13. Probability of symbol error with different values of δ (0.5–0.9).

with respect to two different SNR (0 and 10 dB). We note from the figure that both the resultant NRMSE versus S curves seem to decay at a rate $1/\sqrt{S}$. This can be seen as follows. Consider $S = 10\,000$ and $S = 50\,000$ on the SNR = 0 dB curve. The latter S is five times as large as the former, and the respective NRMSE are -34.1549 and -27.4118 dB. The difference of the two NRMSE ($= -6.7431$ dB) is approximately equal to $20\log_{10}(1/\sqrt{5})$ ($= -6.9897$ dB), viz., the latter NRMSE is roughly $1/\sqrt{5}$ times as small as the former. This phenomenon remains valid if, on each SNR curve, we consider two arbitrary numbers of samples and compute the difference in the respective NRMSE. Hence, although a precise asymptotic performance analysis is not established, this observation seems to indicate that the covariance of the estimated channel decays at a rate $1/S$. We note that the channel estimators in [2] and [14] indeed exhibit this property.

7) *Simulation 7—On Equalization Performance:* In this simulation, we compare the probability of symbol error (PSE)

when different channel estimation schemes are used. We also demonstrate the effect of the threshold δ on PSE. For both experiments, the equalizer is the same 17-tap causal cyclic Wiener filter as that in [2] and [12] (with 12-tap reconstruction delay). The number of samples is fixed at $S = 1000$. For $\delta = 0.5878$, Fig. 11 shows the PSE versus SNR for the proposed method and the two subspace-based methods in [2] and [12]. The PSE is averaged over 500 independent Monte Carlo trials per SNR point. The result shows that the proposed method leads to the lowest PSE. Figs. 12 and 13 show the PSE versus SNR, respectively, for $\delta = 0.1$ –0.5 and $\delta = 0.5$ –0.9. The channel is estimated using the proposed method and PSE is averaged over $I = 500$ trials. Note that large δ (≈ 1) and small δ (≈ 0.1) result in poor performance. The former case is a consequence of poor channel estimation accuracy when large δ is used. A plausible explanation for the latter case is that, although the channel is better estimated with small δ , symbols modulated with a small value ($\sqrt{\delta}$) would be more easily corrupted by noise and, hence, are more likely to cause decision errors. This imposes a

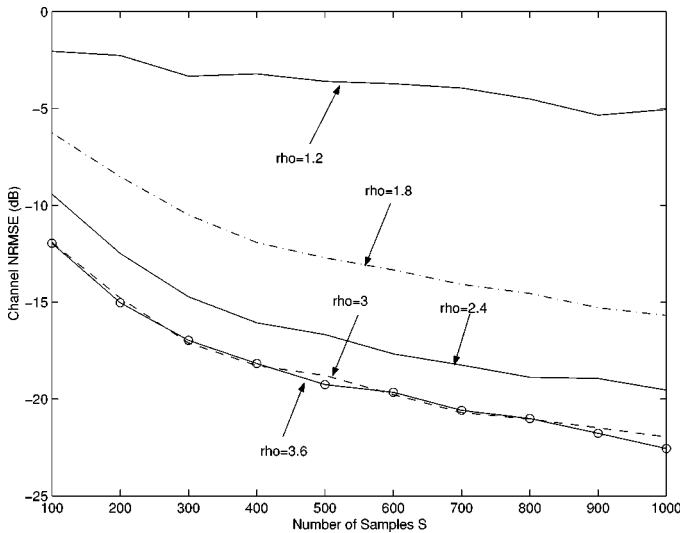


Fig. 14. Channel NRMSE with different choices of upper bound ρ , $\delta = 0.5878$.

tradeoff between PSE and accuracy of channel identification in selecting the threshold δ . We note that this phenomenon is also observed in [2, p. 1584]. Based on these observations, it seems that a good choice is $0.4 \leq \delta \leq 0.6$.

8) *Simulation 8—On Imposing Upper Bounds on the Magnitude of Modulating Sequence:* In this simulation, we provide numerical solutions to optimal modulating sequences when different choices of upper bound ρ in constraint (4.14) are used. The optimization problem is solved using *Matlab Optimization Toolbox*. With the computed solutions, the respective resultant identification performances are also illustrated. For $N = 6$ and $\delta = 0.5878$, we use the algorithm `qp` (see [6, p. 3–38]) to compute the solutions with respect to five choices of ρ : 1.2, 1.8, 2.4, 3, 3.6. Note that for the particular threshold $\delta = 0.5878$, the maximal feasible value of $|p(n)|^2$, when no upper bound on $|p(n)|^2$ is imposed, is $6 - 5 \times 0.5878 = 3.061$. Hence, the choice $\rho = 3.6$ is an inactive bound, and the corresponding solution is expected to be the same as the one computed using (4.12). By computations, it appears that the five solutions are all of the two-level form: $p(0) = \sqrt{\alpha}$ and $p(n) = \sqrt{\beta}$, $1 \leq n \leq 5$, where the respective (α, β) pairs are (1.2, 0.96), (1.8, 0.84), (2.4, 0.72), (3, 0.6), and (3.061, 0.5878). From the results, it can be checked that except for the choice $\rho = 3.6$, the computed optimal sequence for each ρ is the same as the one computed according to the conjectured formula (4.15) with $m = 0$. The solution corresponding to $\rho = 3.6$, as expected, is indeed the same as the one computed using (4.12) (with $m = 0$). Based on the computed optimal modulating sequences, Fig. 14 shows the respective channel NRMSE versus number of samples S . The SNR is fixed at 10 dB, and the number of trials is $I = 100$. The NRMSE curve computed using the sequence (4.12) is also shown. Fig. 14 also shows degeneration in performance as the upper bound ρ decreases toward unity.

VI. CONCLUSION

We propose a method for blind identification of FIR channels with periodic modulation of source symbols. The time-domain

formulation in terms of block signals is simple compared with existing frequency-domain approaches. The method exploits the linear relation between the products of channel coefficients and the autocorrelation matrix of the received signal as well as the decoupled structure of the resulting linear system of equations. The identifiability conditions so derived are particularly simple; they depend on the modulating sequence alone. Indeed, with the proposed method, *any* FIR channel is identifiable with an appropriate choice of the periodic modulating sequence, provided that the modulation period $N \geq L + 2$, where L is the channel order. In fact, *almost* all periodic modulating sequences yield channel identifiability. The optimal modulating sequence selection problem formulated as one of minimizing the effects of channel noise and error in estimating the autocorrelation matrix is straightforward and easy to solve. The proposed optimal solution also results in a consistent channel estimate when the channel noise is white. Simulation results show that the method yields good performance; it compares favorably with existing subspace modulation-induced-cyclostationarity methods, and it is robust with respect to channel order overestimation.

APPENDIX A

PROOF OF PROPOSITION 3.1

From (3.2), we note that for any $N \times N$ matrix M and integer $k \geq 1$, the multiplication $J^k M$ moves the rows of M to k places *below* and leaves the *first* k rows zero. In addition, for integer $l \geq 1$, the multiplication $M(J^T)^l$ moves the rows of M to l places *above* and leaves the *last* l rows zero. Equation (3.5) then follows since G is a diagonal matrix, and $l \geq k$. Equation (3.6) and 2) follow from similar arguments. \square

APPENDIX B

We set $m = 0$ without loss of generality. Let $q \in \mathbb{R}^N$, which is defined in (3.14), be the vector containing the squared magnitudes of the modulating sequence and $b \in \mathbb{R}^N$, which is defined in (4.3), be the vector with all entries equal to one. Then, the constrained optimization problem is

$$\min -q^T q, \text{ subject to } b^T q = N, \quad \text{and } \delta \leq q_n \leq \rho \\ 0 \leq n \leq N - 1. \quad (\text{A.1})$$

Write the inequality constraints in (A.1) directly in terms of vectors as

$$q - \rho b \leq 0_{N \times 1} \quad \text{and} \quad -q + \delta b \leq 0_{N \times 1}. \quad (\text{A.2})$$

The Lagrangian function associated with the problem is thus

$$L(q, \lambda, \mu, \zeta) = -q^T q + \lambda(b^T q - N) + \mu^T (q - \rho b) \\ + \zeta^T (-q + \delta b) \quad (\text{A.3})$$

where $\lambda \in \mathbb{R}$, $\mu = [\mu_0 \cdots \mu_{N-1}]^T \in \mathbb{R}^N$, and $\zeta = [\zeta_0 \cdots \zeta_{N-1}]^T \in \mathbb{R}^N$. Let

$$\bar{q} := [\rho \quad (N-\rho)/(N-1) \quad \cdots \quad (N-\rho)/(N-1)]^T. \quad (\text{A.4})$$

We will show that \bar{q} in (A.4) and, hence, $p(n)$ in (4.15) (with $m = 0$) satisfies the first-order necessary condition for opti-

mality, that is, there exist $\bar{\lambda} \in \mathfrak{R}$, $\bar{\mu} \in \mathfrak{R}^N$, $\bar{\zeta} \in \mathfrak{R}^N$ with $\bar{\mu} \geq 0_{N \times 1}$ and $\bar{\zeta} \geq 0_{N \times 1}$, such that

$$-\bar{q}^T + \bar{\lambda}b^T + \bar{\mu}^T - \bar{\zeta}^T = 0_{1 \times N} \quad (\text{A.5})$$

and

$$\bar{\mu}^T(\bar{q} - \rho b) + \bar{\zeta}^T(-\bar{q} + \delta b) = 0. \quad (\text{A.6})$$

Indeed, with \bar{q} given in (A.4), (A.5) and (A.6) hold with $\bar{\lambda} = (N - \rho)/(N - 1)$, $\bar{\zeta}^T = 0_{N \times 1}$, and $\bar{\mu} = [N(\rho - 1)/(N - 1) \ 0 \ \dots \ 0]^T$. \square

APPENDIX C

PROOF OF LEMMA 4.2

- 1) We first consider the cases $0 \leq m \leq N - L - 1$. Recall that for each $1 \leq j \leq L$, the matrix $M_j \in \mathfrak{R}^{(N-j) \times (L-j+1)}$ is obtained from M_0 by deleting its last j rows and last j columns. It can be checked that each such M_j contains an $(L - j + 1) \times (L - j + 1)$ square submatrix

$$R_j = (N(1 - \delta)) \cdot I_{L-j+1} + \delta \cdot [1 \ \dots \ 1]^T [1 \ \dots \ 1]. \quad (\text{A.7})$$

Since R_j is symmetric and positive definite, M_j is of full column rank. For $m = N - L$ ($m = N - 1$), it can be checked that each associated M_j has an $(L - j + 1) \times (L - j + 1)$ square submatrix with all first lower (upper, respectively) diagonal entries equal to $N(1 - \delta) + \delta$ and all the other entries equal to δ . It can be shown by induction that these square matrices are nonsingular if and only if $\delta \neq 0$ and $N(1 - \delta) \neq 0$. The conditions hold since $0 < \delta < 1$, and hence, M_j is of full column rank.

- 2) To compute μ_m for $0 \leq m \leq N - L - 1$, with (4.12) and (3.13), direct multiplication shows that for each j associated with such an m , we have (A.8), shown at the bottom of the page, where

$$\alpha := N^2(1 - \delta)^2, \quad \text{and} \quad \beta_j := \delta[2N - (N + j)\delta].$$

From (A.8), we know that the maximal and minimal eigenvalues of $M_j^T M_j$ are, respectively, $\alpha + (L - j + 1) \cdot \beta_j$ and α . By definition, the condition number of $M_j^T M_j$ is therefore $\kappa_j = 1 + (L - j + 1) \cdot \beta_j / \alpha$, which is a decreasing function of j . Thus, by definition of μ_m in (4.16), we have $\mu_m = \kappa_0 = 1 + (L + 1)\beta_0 / \alpha$. The result thus follows.

- 3) The result follows from the construction of M_j . Specifically, for $m = N - L + i$, where $1 \leq i \leq L - 2$, it can be checked that the matrix M_j for $j = i + 1$ is not of full column rank since it has two columns both equal to $[\delta \ \dots \ \delta]^T$. \square

APPENDIX D

PROOF OF (4.22)

Define $b_1 := [1 \ \dots \ 1] \in \mathfrak{R}^{L+1}$. It suffices to show that the vector

$$\Gamma_0[D] = \sigma^2 (M_0^T M_0)^{-1} M_0^T b = \zeta b_1. \quad (\text{A.9})$$

By definition of matrix M_0 [see (3.13)] and from constraint (4.7a), it follows immediately that

$$M_0^T b = N b_1. \quad (\text{A.10})$$

Again, by definition of M_0 , it can be verified that

$$M_0^T M_0 = (\alpha - \beta)^2 \cdot I_{L+1} + [2\alpha\beta + (N - 2)\beta^2] \cdot b_1^T b_1 \in \mathfrak{R}^{(L+1) \times (L+1)}. \quad (\text{A.11})$$

From (A.11) and by the matrix inversion lemma [10, p. 306], it can be shown that

$$\begin{aligned} (M_0^T M_0)^{-1} &= (\alpha - \beta)^{-2} I_{L+1} \\ &\quad - \frac{2\alpha\beta + (N - 2)\beta^2}{(\alpha - \beta)^4 + (\alpha - \beta)^2 [2\alpha\beta + (N - 2)\beta^2] (L + 1)} b_1 b_1^T. \end{aligned} \quad (\text{A.12})$$

From (A.10), (A.12), and by rearrangement, (A.9) follows with

$$\zeta = \frac{N\sigma^2}{(\alpha - \beta)^2 + [2\alpha\beta + (N - 2)\beta^2] (L + 1)}. \quad \square$$

APPENDIX E

FLOP COUNTS OF THREE METHODS

The proposed algorithm is compared with the two subspace methods in [2], [12], and [14] in terms of computational complexity. We define a ‘‘flop’’ to be a single complex multiply or addition [8]. Assume that the channel order L is known, the length of each data block is N , and that the number of available data blocks is K .

- a) *Proposed Method*: Estimate the (Hermitian) autocorrelation matrix $\hat{R}_{\bar{x}}(0)$ using (3.19); this requires $(2K - 1)N(N + 1)/2 + 1$ flops. Solving $L + 1$ linear least squares problems, using the QR factorization method [7, p. 226] requires $L^3(2N - L)/3 + L^2(57N - 32L)/6 + L(143N - 67L)/6 + (15N - 7L)/6 + 5$ flops. Compute an eigenvector associated with the largest eigenvalue of the $(L + 1) \times (L + 1)$ matrix Q . Since Q is Hermitian and is of rank-one, the computation of an eigenvector associated with the largest eigenvalue can be done by first computing a tridiagonal form of Q using the algorithm in [7, p. 420] followed by an inverse iteration; see ([7, p. 383]). This requires $4(L + 1)^3/3 + 21(L + 1)^2/2 - 59(L + 1)/6 - 6$ flops.

$$M_j^T M_j = \alpha \cdot I_{L-j+1} + \beta_j \cdot [1 \ \dots \ 1]^T [1 \ \dots \ 1] \in \mathfrak{R}^{(L-j+1) \times (L-j+1)}, \quad 0 \leq j \leq L \quad (\text{A.8})$$

- b) *Structured Subspace Method [2] and [14]*: For $1 \leq k \leq N-1$ and $-L \leq \tau \leq L$, estimate the $(N-1)(2L+1)$ cyclic correlation coefficients $\hat{C}_{xx}(k; \tau)$ using [14, Eq. (20)], and this requires $(N-1)(2L+1)(2NK+1)$ flops. For $1 \leq k \leq N-1$, compute the $N-1$ coefficients $F_2(k)$ using [2, Eq. (5)], and this requires $(2N-1)(N-1)+1$ flops. Based on [14, Eqs. (15) and (17)], form the two matrices $T(F(\mathbf{c}))$ and P , which are, respectively, of dimensions $(3L+1)(N-1)(N-2)/2 \times (L+1)(N-1)$ and $(L+1)(N-1) \times (L+1)$. Perform a matrix multiplication to obtain $T(F(\mathbf{c}))P$, and the required flop counts are $(L+1)(3L+1)(2N^3-9N^2+13N-6)/2$. Compute an SVD of the matrix $T(F(\mathbf{c}))P$, which is of dimension $(3L+1)(N-1)(N-2)/2 \times (L+1)$. With the algorithm in [7, p. 434], the approximate flop counts for computing an SVD of an $m \times n$ matrix, when only the singular values and an orthonormal basis of \mathfrak{R}^n are required, are $(4mn^2+8n^3)$ [7, p. 239], that is, $[2(3L+1)(N-1)(N-2)+8(L+1)](L+1)^2$ flops, as far as the matrix $T(F(\mathbf{c}))P$ is concerned.
- c) *One Cycle Subspace Method [12]*: For some $1 \leq k_0 \leq N-1$, estimate $\hat{C}_{xx}(k_0; \tau)$ for $-L \leq \tau \leq L$, and the flop counts are $(2NK+1)$. Compute $F_2(k_0)$, and the flop counts are $2N+1$. Compute the matrix $S_{k_0}(L)$ with dimension $(3L+1) \times (L+1)$, using [12, Eq. (23)], and this requires $(3L+1)(2L+1)+(4L+2)$ flops. Compute an SVD of $S_{k_0}(L)$, and the approximate required flops are $(20L+12)(L+1)^2$.

We note the following.

- i) In the subspace method [12], the flops cost required for estimating the statistics of received data is relatively small. This is because the method in [12] uses only the cyclic correlation coefficients of only one nonzero cycle (i.e., only one $\hat{C}_{xx}(k_0; \tau)$ for some $k_0 \neq 0$ is required).
- ii) For large channel order L , the computational cost of the method in [14] is relatively high. This is because it requires an SVD of the matrix $T(F(\mathbf{c}))P$ (which is of dimension $(3L+1)(N-1)(N-2)/2 \times (L+1)$).

ACKNOWLEDGMENT

The authors thank Reviewer #1, whose very detailed comments improve the paper.

REFERENCES

- [1] K. Abed-Meraim, W. Qiu, and Y. Hua, "Blind system identification," in *Proc. IEEE*, vol. 85, Aug. 1997, pp. 1310–1322.
- [2] A. Chevreuil, E. Serpedin, P. Loubaton, and G. B. Giannakis, "Blind channel identification and equalization using periodic modulation precoders: Performance analysis," *IEEE Trans. Signal Processing*, vol. 48, pp. 1570–1586, June 2000.
- [3] A. Chevreuil, P. Loubaton, and L. Vandendorpe, "Performance of general transmitter induced cyclostationarity precoder: Analysis based on MMSE-DF receiver," *IEEE Trans. Signal Processing*, vol. 48, pp. 3072–3086, Nov. 2000.

- [4] P. Ciblat, A. Chevreuil, and P. Loubaton, " α -repetition/modulation and blind second order identification," *IEEE Trans. Signal Processing*, vol. 48, pp. 3153–3161, Nov. 2000.
- [5] G. B. Giannakis, Y. Hua, P. Stoica, and L. Tong, *Signal Processing Advances in Wireless and Mobil Communication Volume I: Trends in Channel Identification and Equalization*. Englewood Cliffs, NJ: Prentice-Hall, 2001.
- [6] A. Grace, *Optimization Toolbox, for Use with Matlab*. Natick, MA: The MathWorks, Inc., 1994.
- [7] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 2nd ed. Baltimore, MD: Johns Hopkins Univ. Press, 1989.
- [8] T. P. Krauss and M. D. Zoltowski, "Bilinear approach to multiuser second-order statistics-based blind channel estimation," *IEEE Trans. Signal Processing*, vol. 48, pp. 2473–2486, Sept. 2000.
- [9] C. L. Lawson and R. J. Hanson, *Solving Least Squares Problems*. Englewood Cliffs, NJ: Prentice-Hall, 1974.
- [10] L. Ljung, *System Identification, Theory for the Users*. Englewood Cliffs, NJ: Prentice-Hall, 1987.
- [11] E. Moulines, P. Duhamel, J. Cardoso, and S. Mayargue, "Subspace methods for blind identification of multichannel FIR filters," *IEEE Trans. Signal Processing*, vol. 43, pp. 516–525, Feb. 1995.
- [12] E. Serpedin and G. B. Giannakis, "Blind channel identification and equalization with modulation-induced cyclostationarity," *IEEE Trans. Signal Processing*, vol. 46, pp. 1930–1944, July 1998.
- [13] A. Scaglione, B. Giannakis, and S. Barbarossa, "Redundant filterbank precoders and equalizers. Part I: Unification and optimal design. Part II: Blind channel estimation, synchronization, and direct equalization," *IEEE Trans. Signal Processing*, vol. 47, pp. 1988–2022, July 1999.
- [14] E. Serpedin, A. Chevreuil, G. B. Giannakis, and P. Loubaton, "Blind channel and carrier frequency offset estimation using periodic modulation," *IEEE Trans. Signal Processing*, vol. 48, pp. 2389–2405, Aug. 2000.
- [15] M. Tsatsanis and G. B. Giannakis, "Transmitter induced cyclostationarity for blind channel equalization," *IEEE Trans. Signal Processing*, vol. 45, pp. 1785–1794, July 1997.
- [16] L. Tong, G. Xu, and T. Kailath, "A new approach to blind identification and equalization of multipath channel," in *Proc. 25th Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, Nov. 1991.

Ching An Lin received the B.S. degree from the National Chiao Tung University, Hsinchu, Taiwan, R.O.C., in 1977, the M.S. degree from the University of New Mexico, Albuquerque, in 1980, and the Ph.D. degree from the University of California, Berkeley, in 1984, all in electrical engineering.

He was with the Chung Shan Institute of Science and Technology from 1977 to 1979 and with Integrated Systems Inc. from 1984 to 1986. Since June, 1986, he has been with the Department of Electrical and Control Engineering, the National Chiao Tung University, where he is a Professor. His current research interests are in control and signal processing.



Jwo Yuh Wu was born in Tainan, Taiwan, R.O.C., on July 2, 1973. He received the B.S. and the M.S. degrees, both in electrical and control engineering, in 1996 and 1998, from the National Chiao Tung University, Hsinchu, Taiwan, where he is currently pursuing the Ph.D. degree in electrical and control engineering. His current research interests are in signal processing.

Mr. Wu received the ZyXEL Communications Corporation scholarship in 2001.