UC Riverside
UC Riverside Previously Published Works

Title
Fast maximum likelihood for blind identification of multiple FIR channels

Permalink
https://escholarship.org/uc/item/12z491mk

Journal
IEEE Transactions on Signal Processing, 44(3)

ISSN
1053-587X

Author
Hua, Yingbo

Publication Date
1996-03-01

Peer reviewed

Fast Maximum Likelihood for Blind Identification of Multiple FIR Channels

Yingbo Hua, *Senior Member, IEEE*

Abstract— This paper develops a fast maximum likelihood method for estimating the impulse responses of multiple FIR channels driven by an arbitrary unknown input. The resulting method consists of two iterative steps, where each step minimizes a quadratic function. The two-step maximum likelihood (TSML) method is shown to be high-SNR efficient, i.e., attaining the Cramér-Rao lower bound (CRB) at high SNR. The TSML method exploits a novel orthogonal complement matrix of the generalized Sylvester matrix. Simulations show that the TSML method significantly outperforms the cross-relation (CR) method and the subspace (SS) method and attains the CRB over a wide range of SNR. This paper also studies a Fisher information (FI) matrix to reveal the identifiability of the M-channel system. A strong connection between the FI-based identifiability and the CR-based identifiability is established.

I. INTRODUCTION

BLIND identification of multiple FIR channels is an important problem arising in many areas including mobile communications, multisensor signal analysis, and multisensor image restoration. Although for some applications the unknown (or inaccessible) input to the FIR channels is known to have certain statistical or/and algebraic characteristics, for some others the unknown input could be virtually arbitrary such as nonstationary, non-Gaussian, and colored. Even in mobile communications, when a fast varying channel needs to be identified within a very short period of time, any known statistical characteristics (such as whiteness) of the unknown input becomes hardly useful since a too-short data sequence cannot yield a reliable statistical average. Therefore, under certain practical conditions, the input has to be assumed to be virtually arbitrary.

Note that if a system is time invariant during a period when a long enough data sequence is available and *a priori* statistical information about its input is reliable, then the statistics of the input should be exploited. Examples of using statistical knowledge of the input include the second-order statistic (SOS)-based methods [3]–[5], [32] and the higher order statistics (HOS)-based methods [6]–[8]. However, this paper addresses the situation where the available data sequence is relatively short or the system is fast varying.

The author is with the Department of Electrical and Electronic Engineering, The University of Melbourne, Parkville, Australia (e-mail: yhua@ee.mu.oz.au).

Blind identification of multiple FIR channels driven by virtually arbitrary input was recently studied in [1] and [2]. They utilized a fact that for such an M-channel system, the output of a channel convolved with the impulse response of another channel is equal to the output of the first channel convolved with the impulse response of the second channel. This is what they called a cross-relation property, and hence, their method will be called the cross-relation (CR) method. Apparently, in order to emphasize the deterministic-input model under which the CR method was developed, the CR method was called a deterministic approach in [1]–[2]. However, the CR method does not necessarily require the deterministic assumption. In fact, the CR method is another SOS-based method where the SOS of the channel outputs is estimated in a unique way. For short data sequences, the CR method has been shown in [1], [2], and [16] to significantly outperform the previously developed SOS-based methods [3]–[5]. Another SOS-based method called the subspace (SS) method, was recently reported in [15] and was also shown to be more accurate than those in [3]–[5]. A common feature of the CR and SS methods is that for a virtually arbitrary input and in the absence of noise, both methods yield the exact channel identification.

This paper presents a further investigation into the M-channel system driven by unknown deterministic input. A fast maximum likelihood method that consists of two iterative steps, where each step minimizes a quadratic function, is developed. The two-step maximum likelihood (TSML) method is shown to be high-SNR efficient, i.e., attaining the Cramér-Rao lower bound (CRB) at relatively high SNR. The TSML method performs significantly better than the CR and SS methods. The first step of the TSML method coincides in a natural way with the CR method, and the second step of the TSML method can be viewed as a weighted CR method.

A discussion of a Fisher information (FI) matrix, which yields information on the channel identifiability, is also shown in this paper. Using the FI matrix to study the channel identifiability is in contrast with several existing approaches [1]–[5]. A strong relation between the FI-based identifiability and the CR-based identifiability [1], [2] is established in this paper. In a subsequent paper [18], the channel identifiability is further investigated using a concept called strict identifiability.

Using the ML principle for this blind indentification problem was also recently studied in [13] and [14]. In [13], an iterative algorithm was developed based on the expectation-maximization (EM) principle whose convergence critically depends on a good initial estimate of the channel impulse response. In [14], a two-channel system was studied by fol-

lowing an idea called iterative quadratic maximum likelihood (IQML) [20]–[27]. Although the independently developed TSML method shown in this paper is similar to the IQML method in [14], the former applies to any number of channels (except the single channel case, of course), which is a significant generalization of the latter. Other ML methods such as [28] and [29] deal with a differently formulated problem, where the finite input alphabet is exploited, and an expensive search procedure is required.

Despite the similarity between the TSML and the IQML, we prefer to name our algorithm TSML because further iteration (beyond the second iteration) of the TSML method (or the IQML algorithms shown in [20]–[27]) requires more computations and does not even asymptotically improve the estimation accuracy (or increasing the likelihood function).

The rest of paper is organized as follows. The channel model and its properties are shown in Section II, where an orthogonal complement matrix of the generalized Sylvester matrix is introduced, which lays a foundation for the rest of this work. The FI matrix of the channel model is studied in Section III, where an exact equivalence between the FI-based identifiability and the CR-based identifiability is established. The TSML method is developed in Section IV, where its consistency and efficiency are proved. An efficient implementation of the TSML method is described in Section V. The simulation comparisons of the CR, SS, and TSML methods are shown in Section VI.

II. THE CHANNEL MODEL AND ITS PROPERTIES

This paper considers the following discrete M-channel FIR system:

$$y_1(k) = s(k) * h_1(k) + w_1(k)$$
$$y_2(k) = s(k) * h_2(k) + w_2(k)$$
$$\vdots$$
$$y_M(k) = s(k) * h_M(k) + w_M(k) \qquad (2.1)$$

where

$y_i(k)$ output of the channel i at time k
$s(k)$ common input
$h_i(k)$ finite impulse response (FIR) of the channel i
$w_i(k)$ noise.

It is assumed that the maximum order of the M channels is L, and the number of available output samples of each channel is N. The symbol $*$ denotes convolution. All quantities in (2.1), except the integers, are complex valued unless stated otherwise. Note that the channel order L will be assumed to be known throughout this paper.

It is easy to verify that (2.1) can be expressed as follows:

$$y = H_M s + w \qquad (2.2)$$

where y is the output vector of all channels

$$y = [y_1^T \quad y_2^T \quad \cdots \quad \cdots \quad y_M^T]^T \qquad (2.3)$$

with

$$y_i = [y_i(0) \quad y_i(1) \quad \cdots \quad \cdots \quad y_i(N-1)]^T \qquad (2.4)$$

and the superscript T denotes the transpose. w is the noise vector of all channels. s is the input vector

$$s = [s(-L) \quad s(-(L-1)) \quad \cdots \quad s(0) \quad s(1) \quad \cdots \quad s(N-1)]^T. \qquad (2.5)$$

H_M is known as a generalized Sylvester matrix [11]

$$H_M = \begin{bmatrix} H_{(1)} \\ H_{(2)} \\ \vdots \\ H_{(M)} \end{bmatrix} \qquad (2.6)$$

where $H_{(i)}$ is the $N \times (N+L)$ Sylvester matrix of the ith channel response, which is shown in (2.7) at the bottom of the page.

The matrix H_M has a number of important properties. These properties, along with an orthogonal complement matrix of the generalized Sylvester matrix, will be shown later in this section.

To describea property of channels, the notion of "zeros" will be needed. A zero z_0 of the channel i is defined by

$$H_i(z_0) = \sum_{k=0}^{L} h_i(k) z_0^{-k} = 0$$

where $H_i(z)$ is known as the transfer function of the ith channel. Another common notion of linear systems will be needed. This notion is called "modes." A mode in this paper is defined to be a finite sequence of the form $m(k) = k^n z^k$ for some complex number z and a nonnegative integer n, where z is referred to as its root and n the order. A sequence $\{s(k), k = -L, \cdots, N-1\}$ of length $N+L$ is said to have p modes if $s(k)$ can be written as a linear combination of p modes of length $N+L$, i.e., for $k = -L, \cdots, N-1$

$$s(k) = \sum_{i=1}^{p} c_i m_i(k)$$

where for each (nonzero) mode, all the coefficients of the corresponding lower order modes *may* or *may not* be zero. For example, the sequence $c_1 k^2 z_1^k + c_2 k z_1^k + c_3 z_1^k + c_4 z_2^k$ is said to have four modes as long as c_1 and c_4 are nonzero. A

$$H_{(i)} = \begin{bmatrix} h_i(L) & h_i(L-1) & \cdots & & \cdots & h_i(0) & & \\ & h_i(L) & h_i(L-1) & \cdots & & \cdots & h_i(0) & \\ & & \ddots & \ddots & \ddots & \ddots & \ddots & \\ & & & h_i(L) & h_i(L-1) & \cdots & \cdots & h_i(0) \end{bmatrix}. \qquad (2.7)$$

channel is *said to be* associated with a mode $m(k)$ if the root of the mode is a zero of the channel. The following facts on modes are easy to show:

    a) Any m (distinct) modes of length r are independent of each other provided $m \leq r$.

    b) Any sequence of length r can be expressed as a linear combination of no more than r modes of length r.

    c) With probability one, an arbitrary sequence of length r has r modes of length r (although, in practice, a sequence of length r may have less than r dominant modes of length r).

    d) A mode of root z and order n at time $k+1$ is a linear combination of modes of the same root z and orders 0, 1, $\cdots$, and n, at time k, i.e.

$$(k+1)^n z^{k+1} = z \sum_{i=0}^{n} \frac{n!}{(n-i)!i!} k^i z^k.$$

(This relation is needed for Lemma 1 in [18].)

The definitions of "modes" vary in the literature; see [19, p. 168], [1], and [2]. The modes defined in this paper can be viewed as independent basis functions to describe any finite length sequence (including any finite length realization of white or colored random processes).,

For the first reading, readers may now choose to skip the following Lemmas 1–6 but read Theorem 1 and its discussions in the end of this section. Theorem 1 provides an orthogonal complement matrix of the generalized Sylvester matrix H_M, which is essential in relating the FI matrix to the CR method and in developing the TSML method.

For the following lemmas, we need to define two matrices: a p-channel system matrix H_p, which is simply H_M with M replaced by p, and a "companion" p-channel system matrix $\overline{H}_p$ defined by

$$\overline{H}_p = \begin{bmatrix} \overline{H}_{(1)} \\ \overline{H}_{(2)} \\ \vdots \\ \overline{H}_{(p)} \end{bmatrix} \tag{2.8}$$

where $\overline{H}_{(i)}$ is the top-left $(N-L) \times N$ submatrix of $H_{(i)}$.

Lemma 1:

$$\mathrm{rank}(H_p) = N + L - c \quad \text{for} \quad N \geq L - c$$
$$\mathrm{rank}(\overline{H}_p) = N - c \quad \text{for} \quad N \geq 2L - c$$

where c is the number of zeros shared by the channels 1 to p. (Note that $c = L$ when $p = 1$). Hence, H_p and $\overline{H}_p$ have full column rank if and only if there is no common zero among the p channels, provided $N \geq 2L$.

Proof: See proof of Theorem 1 in [11].   ◆

We now define two matrices of modes: $M_{1:p}$ and $\overline{M}_{1:p}$, which are $(N+L) \times c$ and $N \times c$, respectively. Each column of $M_{1:p}$ or $\overline{M}_{1:p}$ is a mode sequence with root equal to a zero shared by the channels 1 to p. Let the modes $m_i(k), i = 1, 2, \cdots, c$, be shared by the p channels, and then, we have (2.9), which appears at the bottom of the page, and $\overline{M}_{1:p}$ is the top N rows of $M_{1:p}$.

Lemma 2:

$$\mathrm{null}(H_p) = \mathrm{range}(M_{1:p}) \quad \text{for} \quad N \geq L - c$$
$$\mathrm{null}(\overline{H}_p) = \mathrm{range}(\overline{M}_{1:p}) \quad \text{for} \quad N \geq 2L - c.$$

Proof: It is easy to verify that $H_p M_{1:p} = 0$ and $\mathrm{rank}(M_{1:p}) = c$, and hence, $\mathrm{rank}(H_p) + \mathrm{rank}(M_{1:p}) = N + L$, which is the total number of the columns of H_p (where Lemma 1 is used). The same can be verified for the second equation. ◆

Lemma 3: Let each column of a $(N+L)$-row matrix $M_{|i}$ be a mode sequence with root *not* equal to any zero of the channel i, and let $\overline{M}_{|i}$ be the top N rows of $M_{|i}$ and $\overline{\overline{M}}_{|i}$ be the top $N-L$ rows of $M_{|i}$. Then

$$\mathrm{range}(H_{(i)} M_{|i}) = \mathrm{range}(\overline{M}_{|i})$$
$$\mathrm{range}(\overline{H}_{(i)} \overline{M}_{|i}) = \mathrm{range}(\overline{\overline{M}}_{|i}).$$

Proof: If all modes in $M_{|i}$ are of the first order with the distinct roots: $z_1, z_2, \cdots, z_t$, it is easy to verify that

$$H_{(i)} M_{|i} = \overline{M}_{|i} \, \mathrm{diag}(z_1^L H_i(z_1) \quad z_2^L H_i(z_2) \quad \cdots \quad z_t^L H_i(z_t)).$$

If any of the modes in $M_{|i}$ is of higher order, one can similarly verify that

$$H_{(i)} M_{|i} = \overline{M}_{|i} T$$

where T is a full-rank square matrix consisting of the channel transfer functions and their derivatives at the zeros, and the highest order of the derivatives is equal to the highest order of modes minus one. A similar proof can be found for the second equation in the lemma.   ◆

Lemma 4:

$$\mathrm{range}\begin{pmatrix} 0 \\ \vdots \\ 0 \\ \overline{M}_{1:p|p+1} \end{pmatrix} \subset \mathrm{range}\begin{pmatrix} H_{(1)} \\ \vdots \\ H_{(p)} \\ H_{(p+1)} \end{pmatrix}$$

where $\overline{M}_{1:p|p+1}$ consists of the modes shared by the channels 1 to p but not by the channel $p+1$. The symbol $\subset$ denotes "belong to."

$$M_{1:p} = \begin{bmatrix} m_1(0) & m_2(0) & \cdots & m_c(0) \\ m_1(1) & m_2(1) & \cdots & m_c(1) \\ \vdots & \vdots & \cdots & \vdots \\ m_1(N+L-1) & m_2(N+L-1) & \cdots & m_c(N+L-1) \end{bmatrix} \tag{2.9}$$

Proof: It is easy to verify that

$$H_{p+1}M_{1:p|p+1} = \begin{bmatrix} H_{(1)}M_{1:p|p+1} \\ \vdots \\ H_{(p)}M_{1:p|p+1} \\ H_{(p+1)}M_{1:p|p+1} \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \overline{M}_{1:p|p+1}T \end{bmatrix}$$

where Lemmas 2 and 3 are applied. Since T is nonsingular, Lemma 4 is proved. ♦

Lemma 5: For $N \geq L$, the columns of H_p are independent of those of $C_{1:p}$ defined as

$$C_{1:p} = \begin{bmatrix} 0 & \cdots & \cdots & 0 \\ \overline{M}_{1:p} & & & \\ & \ddots & & \\ & & \ddots & \\ & & & \overline{M}_{1:p} \end{bmatrix}$$

whose dimension is $pN \times (p-1)c$, and henc rank$(H_p, C_{1:p}) = N + L + (p - 2)c$.

Proof: Suppose that there is a nonzero vector t such that $H_p t$ belongs to range$(C_{1:p})$. Since the top N elements of $H_p t$ must be zero, i.e., $H_{(1)}t = 0$, then t must be a linear combination of the L modes of the channel 1 (Lemma 2 for $p = 1$). However, for $i > 1$, a nonzero $H_{(i)}t$ (other elements in $H_p t$) must belong to range$(\overline{M}_{1:p|i})$ (Lemmas 2 and 3). However, range$(\overline{M}_{1:p|i})$ is independent of range$(\overline{M}_{1:p})$. Therefore, no nonzero $H_p t$ belongs to range$(C_{1:p})$. ♦

We now define the matrix G_p as follows:

$$G_2^H = [-\overline{H}_{(2)} \quad \overline{H}_{(1)}]$$

$$G_{p+1}^H = \begin{bmatrix} \begin{array}{c|c} G_p^H & 0 \\ \hline -\overline{H}_{(p+1)} & \overline{H}_{(1)} \\ -\overline{H}_{(p+1)} & \overline{H}_{(2)} \\ \ddots & \vdots \\ -\overline{H}_{(p+1)} & \overline{H}_{(p)} \end{array} \end{bmatrix}.$$

(2.10)

The following lemma shows a relation between G_p and H_p.

Lemma 6: For $p = 2$ and $N \geq L+1$, or $p > 2$ and $N \geq 2L$,

$$\text{null}(G_p^H) = \text{range}(H_p \quad C_{1:p}).$$

Proof: One can easily verify that

$$\overline{H}_{(i)}H_{(j)} = \overline{H}_{(j)}H_{(i)}$$

which is in fact the commutativity of convolution. Then it is easy to verify that range$(H_p, C_{1:p})$ belongs to null(G_p^H), i.e., $G_p^H(H_p, C_{1:p}) = 0$. What is left to show is that range$(H_p, C_{1:p})$ is the *complete* null space of G_p^H. For $p = 2$ and $N \geq L + 1$, it is easy to see that rank$(G_2^H) = N - L$ and rank$(H_2, C_{1:2}) = N + L$ (Lemma 5), and therefore, null$(G_2^H) = $ range$(H_2, C_{1:2})$. (Note that the number of columns of G_2^H is $2N$.) Now, it suffices to show that the lemma is true for G_{p+1} assuming it is true for G_p. Denote

a null vector of G_{p+1}^H by

$$t = \begin{bmatrix} t_a \\ t_{p+1} \end{bmatrix} = \begin{bmatrix} t_1 \\ \vdots \\ t_p \\ t_{p+1} \end{bmatrix}$$

such that the equation

$$G_{p+1}^H t = 0$$

is equivalent to the following:

$$G_p^H t_a = 0$$
$$-\overline{H}_{(p+1)}t_i + \overline{H}_{(i)}t_{p+1} = 0 \quad \text{for} \quad i = 1, \cdots, p.$$

We now search for the complete solution space of these two equations. The solution space of the first equation (for t_a) is now assumed to be what is shown in the lemma. To find the complete solution space of the second equation (for $t_p + 1$), all we need to do is to find a solution for t_{p+1} (if it exists) for each nonzero t_a from range$(H_p, C_{1:p})$ *and* all solutions for t_{p+1} when $t_a = 0$. Note that range$(C_{1:p}) = $ range$(C'_{1:p}) + $ range$(C''_{1:p})$, where $C'_{1:p}$ is $C_{1:p}$ with all modes shared by the channel $p + 1$, and $C''_{1:p}$ is $C_{1:p}$ with no modes shared by the channel $p + 1$. We now consider the following cases.

Case a: If $t_a = H_p c$ for any c, then $t_{p+1} = H_{(p+1)}c$ is a solution (easy to verify)

Case b: If $t_a = C''_{1:p}c$ for some c, then $t_1 = 0$, and each t_i, for $i, > 1$ must be a linear combination of the modes shared by the channels 1 to p but not by the channel $p + 1$. Then, t_{p+1} must satisfy

$$-\overline{H}_{(1)}t_{p+1} = 0$$

and

$$-\overline{H}_{(p+1)}t_i + \overline{H}_{(i)}t_{p+1} = 0 \quad \text{for} \quad i \geq 2.$$

The former implies that t_{p+1} is a linear combination of the L modes of the channel 1 (Lemma 2 for $p = 1$), and then $\overline{H}_{(i)}t_{p+1}$ is a linear combination of the modes not shared by the channel i (Lemmas 2 and 3). However, $-\overline{H}_{(p+1)}t_i$ is a linear combination of the modes shared by the channels 1 to p (*including the channel i*) but not the channel $p + 1$. Therefore, for nonzero $t_a = C''_{1:p}c$, there is no solution for t_{p+1} (provided $N \geq 2L$ of course)

Case c: If $t_a = C'_{1:p}c$, then $t_{p+1} = 0$ is a solution (easy to verify).

Case d: If $t_a = 0$, the solution space for t_{p+1} is null$(\overline{H}_p)$ which is equal to range$(\overline{M}_{1:p})$ (Lemma 2). However, the solution subspace $t_a = 0$ and all t_{p+1} from range$(\overline{M}_{1:p|p+1})$ is a part of the solution space of Case a (Lemma 4). Hence, we only need to keep range$(\overline{M}_{1:p+1})$ for t_{p+1} in this case.

Combining Cases a–d yields the complete solution space of $G_{p+1}^H t = 0$, which is range$(H_{p+1}, C_{1:p+1})$. The proof is completed. ♦

Theorem 1: Provided that all channels do not share a common zero, and $N \geq 2L$ (for $M = 2$, only $N > L$ is required), an *orthogonal complement* matrix of the generalized Sylvester matrix H_M is G_M, which is defined in (2.10), i.e.

$$P_G + P_H = I$$

where I is the identity matrix, and P_G and P_H denote the orthogonal projection matrices onto range(G_M) and range(H_M), respectively.

Proof: This theorem follows directly from Lemma 6, where $C_{1:M}$ is empty. ♦

The $MN \times \frac{1}{2}M(M-1)(N-L)$ orthogonal complement matrix G_M is expressed directly in terms of the system impulse response and, hence, is very useful for simplifying the ML estimation of the channel impulse responses, which will be shown in Section IV. Furthermore, G_M will be instrumental in establishing a link between the FI matrix and the CR method, which will be shown in Section III.

For $M = 2$, the form of G_M was also shown in [14]. However, for $M > 2$, the form of G_M is much more complicated, and its property as orthogonal complement of H_M is much more involved to prove.

There are still open questions associated with Theorem 1. For $M = 2$, the condition $N \geq L + 1$ is not only sufficient but also necessary (easy to show). However, for $M > 2$, is the condition $N \geq 2L$ also necessary? We have not found an answer to this question except for the case where $M = 3$. For $M = 3$, the answer is yes, as is shown next. Assume that $M = 3$. Then, G_M is of $3N \times 3(N - L)$, H_M is of $3N \times (N + L)$, rank(H_M) $= N + L$ (from Lemma 1 with $c = 0$ and $N \geq L$), and then, left-nullity(H_M) $= 3N - (N + L)$. Then, a necessary condition for G_M to be an orthogonal complement matrix of H_M is columns(G_M) $\geq$ left-nullity(H_M), i.e., $3(N - L) \geq 3N - (N + L)$, or equivalently, $N \geq 2L$.

Another question is about the uniqueness of G_M. Clearly, G_M does not always have independent columns, i.e., G_M is "fat." Searching for a "leaner" orthogonal complement matrix is still a challenge (deleting any columns of G_M does not seem to yield another orthogonal complement matrix).

III. FI MATRIX

Before a fast ML method is developed in the next section, this section presents a FI matrix for the M-channel system. For this purpose, it is assumed that the channel noise is white complex circular Gaussian, and the input sequence is unknown and deterministic. It then follows that the channel output vector y is Gaussian distributed with the mean vector $m = H_M s$ and the covariance matrix $R_y = \sigma_w^2 I$, where σ_w^2 is the noise variance. Except for σ_w^2, the unknown parameters in this system are described by the $2(ML + M + N + L) \times 1$ vector a

$$a = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_{2(ML+M+N+L)} \end{bmatrix} = \begin{bmatrix} \text{Re}(h) \\ \text{Re}(s) \\ \text{Im}(h) \\ \text{Im}(s) \end{bmatrix} \qquad (3.1)$$

where s is the input vector defined in (2.5), and h the system impulse response vector defined as

$$h = [h_1^T \quad h_2^T \quad \cdots \quad h_M^T]^T$$

and h_i is the ith channel impulse response vector

$$h_i = [h_i(0) \quad h_i(1) \quad \cdots \quad h_i(L)]^T.$$

It is well known (easy to show) that the unknown noise variance σ_w^2 is decoupled from (or does not affect the CRB of) the parameters in the mean vector m. It is also easy to show that the FI matrix [12] associated with a (containing all unknown parameters in m) is given by

$$F_{i,k} = \frac{2}{\sigma_w^2} \text{Re} \left\{ \frac{\partial m^H}{\partial \alpha_i} \frac{\partial m}{\partial \alpha_k} \right\}. \qquad (3.2)$$

From here, one can verify that the $2(ML + M + N + L) \times 2(ML + M + N + L)$ FI matrix can be further expressed as

$$F = \frac{2}{\sigma_w^2} \begin{bmatrix} \text{Re}(F_c) & -\text{Im}(F_c) \\ \text{Im}(F_c) & \text{Re}(F_c) \end{bmatrix} \qquad (3.3)$$

where F_c is a $(ML + M + N + L) \times (ML + M + N + L)$ *complex* FI matrix defined as

$$F_c = Q^H Q \qquad (3.4)$$

with

$$Q = [I_M \otimes S \quad H_M] \qquad (3.5)$$

$$S = \begin{bmatrix} s(0) & s(-1) & \cdots & s(-L) \\ s(1) & s(0) & \cdots & s(-L+1) \\ \vdots & \vdots & \vdots & \vdots \\ s(N-1) & s(N-2) & \cdots & s(N-L-1) \end{bmatrix} \qquad (3.6)$$

in which I_M is the $M \times M$ identity matrix, and $\otimes$ is the Kronecker product. In fact, $I_M \otimes S = \text{diag}(S, \cdots, S)$. The (i, j)th element of F_c corresponds to the ith and jth elements of the complex parameter vector defined as

$$a_c = \begin{bmatrix} \alpha_{c,1} \\ \alpha_{c,2} \\ \vdots \\ \alpha_{c,ML+M+N+L} \end{bmatrix} = \begin{bmatrix} h \\ s \end{bmatrix}. \qquad (3.7)$$

It is easy to show that if F_c^{-1} exists (this requires deleting certain rows and columns from F_c, to be shown later), then the inverse of the FI matrix is given by

$$F^{-1} = \frac{\sigma_w^2}{2} \begin{bmatrix} \text{Re}(F_c^{-1}) & -\text{Im}(F_c^{-1}) \\ \text{Im}(F_c^{-1}) & \text{Re}(F_c^{-1}) \end{bmatrix}. \qquad (3.8)$$

The CRB [12, p. 79] on the variance of each parameter is given by a corresponding element on the diagonal of F^{-1}. Note that this CRB is a function of the input sequence.

The FI matrix not only provides the CRB but also the information on the channel identifiability. It is well known [12, p. 70] that the ML method yields parameter estimates with variances equal to the CRB at high SNR (with all other variables such as N fixed). Therefore, if the FI matrix is invertible, i.e., the CRB exists, then all parameters associated with the FI matrix are identifiable as the SNR approaches

infinity. Based on this observation, it follows that the nullity of the FI matrix is equal to the total number of parameters which are not identifiable, or in other words,

$$\text{nullity}(F) = \text{degrees of uncertainty.}$$

In terms of F_c, it can be written as

$$\text{nullity}(F_c) = \text{complex degrees of uncertainty} \qquad (3.9)$$

where a complex degree corresponds to a complex unknown parameter.

The following theorem reveals several relations between nullity(F_c) and the system conditions.

Theorem 2:

a) nullity(F_c) ≥ 1.

b) nullity(F_c) > 1 if $\{s(k), k = -L, \cdots, N-1\}$ has less than $L+2$ modes, the M channels share a common zero, or $N \leq L+1$.

c) nullity(F_c) $= 1$ if $\{s(k), k = -L, \cdots, N-1\}$ has $2L+1$ or more modes, the M channels do not share a common zero, and $N \geq 3L+1$.

d) If nullity(F_c) $= d$, let F_c' be F_c without d rows and d columns of F_c corresponding to d nonzero parameters in a, and then, nullity(F_c') $= 0$.

Proof: Recall (3.4) and (3.5). Part a is easy to show by verifying that $[h', s'] = [h, -s]$ is always a solution to

$$Q \begin{bmatrix} h' \\ s' \end{bmatrix} = [I_M \otimes S \quad H_M] \begin{bmatrix} h' \\ s' \end{bmatrix} = 0. \qquad (3.10)$$

Part d also follows from an observation of (3.10): deleting d columns of Q corresponding to d nonzero elements of $[h^T, s^T]$ leads to no solution to (3.10). A direct proof of Parts b and c has been found but is omitted here due to the following Theorem 3 and some related results in [1], [2], [18]. ◆

Part a) of Theorem 2 implies that there is always a complex degree of uncertainty in the M-channel system (which is not surprising of course). Based on this, we make a definition of the FI based system identifiability:

Definition 1—FI Identifiability: The M-channel system is said to be FI identifiable if nullity($F_c = 1$). $= 1$.

With this definition, Part b) provides a necessary channel identifiability (ID) condition, and Part c gives a sufficient ID condition. Comparing the FI-based ID conditions with those based on the CR method [1], [2] implies a close connection between the FI-based identifiability and the CR-based identifiability. To establish a strong connection between them, the CR-based identifiability is now reviewed.

The authors of [1] and [2] exploited a (cross-relation) fact that the output of a channel convolved with the impulse response of another channel is equal to the output of the second channel convolved with the impulse response of the first channel, i.e.,

$$y_i(k) * h_j(k) = y_j(k) * h_i(k).$$

Based on this, they showed that $h' = h$ must be a solution to the following (cross-relation) equation in the absence of noise:

$$Y_M h' = 0 \qquad (3.11)$$

where

$$Y_M = \begin{bmatrix} Y_{M-1} & & & & 0 \\ Y_{(M)} & & & & -Y_{(1)} \\ & Y_{(M)} & & & -Y_{(2)} \\ & & \ddots & & \vdots \\ & & & Y_{(M)} & -Y_{(M-1)} \end{bmatrix}$$
$$(3.12)$$

with

$$Y_2 = [Y_{(2)} \quad -Y_{(1)}]$$

and

$$Y_{(i)} = \begin{bmatrix} y_i(L) & y_i(L-1) & \cdots & y_i(0) \\ y_i(L+1) & y_i(L) & \cdots & y_i(1) \\ \vdots & \vdots & \vdots & \vdots \\ y_i(N-1) & y_i(N-2) & \cdots & y_i(N-L-1) \end{bmatrix}.$$
$$(3.13)$$

(The CR equation (3.11) differs from that in [1] and [2] only in the order of rows of the matrix Y_M.) Clearly, if (3.11) has a unique solution in the absence of noise, then the M-channel system is identifiable in terms of h up to a constant scalar. Based on this, we make a formalization of the identifiability discussed in [1] and [2].

Definition 2—CR Identifiability: The M-channel system is said to be CR identifiable if (3.11) has a unique solution in the absence of noise.

Now, a strong relation between the FI identifiability and the CR identifiability can be demonstrated below.

Theorem 3: Provided $N \geq 2L$ (or $N \geq L+1$ for two channels), the FI-based identifiability is equivalent to the CR-based identifiability.

Proof: The proof consists of two cases:

Case a)   There is a common zero among channels.

Case b)   There is no common zero among channels.

For Case a), we show that the system is not identifiable based on either the FI matrix or the CR equation. For Case b), we first show that if the system is not CR identifiable, it must not be FI identifiable and then show that if the system is not FI identifiable, it must not be CR identifiable.

Case a): Suppose that there is a common zero among the M channels. Then, the M-channel system can be easily shown to be not identifiable based on either the CR equation or the FI matrix. The former was shown in [1] and [2], and the latter can be easily shown by observing that there are more than one independent solutions to (3.10), where H_M does not have a full column rank (Lemma 1).

Case b): Now, suppose that there is no common zero among the M channels. The following identity (easy to verify) will be needed:

$$Y_M h' = G_M'^H y \qquad (3.14)$$

where G_M' is constructed from h'. First, we further assume that the system is not CR identifiable. It then follows from (3.11) and (3.14) that for some h' independent of h

$$Y_M h' = G_M'^H y = G_M'^H H_M s = 0. \qquad (3.15)$$

However, one can easily verify that

$$G_M'^H H_M = G_M^H H_M'. \tag{3.16}$$

Hence, (3.15) implies

$$G_M^H H_M' s = 0 \tag{3.17}$$

which means that $H_M' s$ is in null(G_M^H), and hence, by Theorem 1 (provided $N \geq 2L$ or $N \geq L+1$ for two channels)

$$H_M' s = -H_M s' \tag{3.18}$$

for some s'. It is easy to verify that (3.18) is equivalent to (3.10). Therefore, the system must not be FI identifiable.

Still for Case b), we now assume that the system is not FI identifiable (and there is no common zero). Then, (3.10) or, equivalently, (3.18) must have the solution $(h, -s)$ and another independent solution (h', s'). Given the fact that H_M has a full column rank (Lemma 1), (3.18) implies that h' is independent of h. Using this (h', s') in (3.18) implies (3.17) and then (3.15) and, therefore, that the system must not be CR identifiable. The proof is completed.   ◆

IV. FAST MAXIMUM LIKELIHOOD

We now develop a fast (conditional) ML method for estimating the channel impulse responses. Assuming that the channel noise is white complex circular Gaussian, it is easy to show that the maximum likelihood (ML) estimation of h from y is obtained by

$$\max_{h} y^H P_H y \tag{4.1}$$

or equivalently (under the condition of Theorem 1)

$$\min_{h} y^H P_G y. \tag{4.2}$$

It is also well known that (4.1) or (4.2) is the least square error estimation where the error is the difference between the measured channel output and the estimated channel output. For short notation, H_M and G_M will be replaced by H and G, respectively. Once the ML estimate of h is available, the ML estimate of s is known to be $(H^H H)^{-1} H^H y$. The estimation of s (i.e., equalization in the context of communications) will not be further addressed.

It is clear that the projection matrix $P_H = H(H^H H)^{-1} H^H$ is invariant to any complex constant scale on H or h, and hence, the solution to (4.1) or (4.2) is not exactly unique. This was predicted by Theorem 2. In the sequel, it is assumed that the sufficient ID condition as stated in Theorem 2 is satisfied so that the channels are identifiable uniquely up to an arbitrary complex scalar. If we subject (4.1) or (4.2) to $\|h\| = 1$, this will lead to a solution that has an arbitrary scalar of unit amplitude.

A fast ML estimation can now be developed as follows. Rewrite the function in (4.2) into

$$y^H P_G y = y^H G(G^H G)^+ G^H y \tag{4.3}$$

where $^+$ denotes the pseudoinverse. One can verify that

$$G^H y = Y h \tag{4.4}$$

where Y is a short form of Y_M defined in (3.12). Hence, (4.2) becomes

$$\min_{h} h^H Y^H (G^H G)^+ Y h. \tag{4.5}$$

This expression suggests an iterative two-step estimation procedure as shown below:

The TSML Method

Step 1: Minimize $h^H (Y^H Y) h$ with $\|h\| = 1$ to yield h_c.
Step 2: Minimize $h^H (Y^H (G_c^H G_c)^+ Y) h$ with $\|h\| = 1$ to yield h_e, where G_c is constructed from h_c according to (2.10).

At a high computational level, the TSML method is indeed similar to the IQML methods developed for exponential signal processing [20]–[27]. An important novelty here is the construction of G. As will be seen later, the TSML method differs significantly from the existing IQML algorithms at a lower computational level.

The TSML method establishes a natural connection between the CR method [1], [2] and the ML principle, i.e., Step 1 of the TSML method coincides with the CR method. The CR method was developed based on an algebraic insight into the M-channel system. It is interesting to know that this insight also relates to the classical concept of ML estimation through a simple weighting matrix $(G_c^H G_c)^+$. Readers may recall a similar relation between Prony's method and the ML estimation [21] in the context of exponential signal processing. In contrast to [21], we do not suggest further iterations beyond Step 2. This is because further iteration requires a lot more computations and is not guaranteed to improve the accuracy even asymptotically.

Consistency and Efficiency

As shown in [1] and [2], in the absence of noise, Step 1 of the TSML method yields the exact solution h (up to a constant), i.e., h is the unique null vector of the covariance matrix $R_Y = Y^H Y$ (a second-order statistic of the channel outputs). This is called high-SNR consistency of Step 1. Furthermore, if the noise is white and the data length N is infinite, one can verify that

$$R_Y = R_S + R_W \tag{4.6}$$

where R_S is the signal term without the effect of noise, and R_W is the noise term proportional to an identity matrix. In this large-sample case, the least eigenvector of R_Y is also equal to h. This is called large-sample consistency of Step 1.

In the following, we show that Step 2 of the TSML method is high-SNR consistent as well as high-SNR efficient (attaining the CRB). One can verify that in the absence of noise

$$Y = G^H (I_M \otimes S) \tag{4.7}$$

and then

$$\text{range}(Y) = \text{range}(G^H) \tag{4.8}$$

and hence,

$$\text{null}(Y^H (G^H G)^+ Y) = \text{null}(Y^H Y). \tag{4.9}$$

Therefore, Step 2 of the TSML method is high-SNR consistent. Note that this is true despite the fact that $G^H G$ may *not* be of full rank for $M > 2$. To show the high-SNR efficiency of the TSML method, it suffices to verify that at high SNR, the TSML method yields the same (up to the first order approximation) estimates as the exact ML method (e.g., see [12, p. 70]). We now need to verify the following two equations:

$$h_e^H Y^H (G_c^H G_c)^+ Y h_e = h_e^H Y^H (G^H G)^+ Y h_e \qquad (4.10)$$

$$h_m^H Y^H (G_m^H G_m)^+ Y h_e = h_m^H Y^H (G^H G)^+ Y h_m \qquad (4.11)$$

where

m ML estimates,
c consistent estimates,
e estimates by the TSML method.

Note that the minimizer h_e of the left-hand term of (4.10) is the TSML estimate, and the minimizer h_m of the left-hand term of (4.11) is the ML estimate. The minimizer h_e of the right-hand term of (4.10) is obviously the same as the minimizer h_m of the right-hand term of (4.11). What is left to show is that both (4.10) and (4.11) hold to the first-order approximation as SNR approaches infinity. To show that (4.10) holds at high SNR, we assume that L (which is the maximum order of the M-channel system) is known. Then, the weight matrix $(G_c^H G_c)^+$ is a continuous function of the noise (despite the operation of pseudoinverse). Hence, one can write $(G_c^H G_c)^+ = W + \Delta W$, where $W = (G^H G)^+$ is the ideal weight, and ΔW approaches zero as SNR goes to infinity. One can also write $h_e = h + \Delta h_e$, where Δh_e approaches zero as SNR goes to infinity. By substituting these in the left-hand term of (4.10), one can easily verify that the resulting term becomes the right-hand term of (4.10) plus higher order terms of ΔW and Δh_e. Therefore, (4.10) holds to the first-order approximation at high SNR. Equation (4.11) can be similarly proven.

V. IMPLEMENTATION OF THE TSML METHOD

The previous formulation of the TSML method is given at a high computational level. The computational efficiency of the TSML method still largely depends on how it is implemented. Note that the matrices G and Y are sparse, and the pseudoinverse $(G^H G)^+$ (with a known rank) needs to be computed. In this section, we discuss what we think is the most efficient implementations in terms of flops. (A flop is defined [9] as a complex addition and a complex multiplication. The number of flops is also roughly equal to the number of multiplications in a typical algorithm, and hence, one can count the flops simply by counting the multiplications.) This section should be useful for users of the TSML method.

Implementation of Step 1 (CR Method)

In Step 1, we need to compute the $M(L + 1) \times M(L + 1)$ matrix product $Y^H Y$ and its least eigenvector. The least eigenvector can be obtained with an order of $M^2 (L + 1)^2$

flops. The product $Y^H Y$ can be expressed as

$$Y^H Y = \begin{bmatrix} \overline{R_Y(1,1)} & R_Y(2,1) & \cdots & R_Y(M,1) \\ R^Y(2,1)^H & \overline{R_Y(2,2)} & \cdots & R_Y(M,2) \\ \vdots & \vdots & \ddots & \vdots \\ R_Y(M,1)^H & R_Y(M,2)^H & \cdots & \overline{R_Y(M,M)} \end{bmatrix}$$

$$(5.1)$$

where for $i = 1, \cdots, M$ and $j = 1, \cdots, i$

$$R_Y(i,j) = -Y_{(i)}^H Y_{(j)}$$

$$\overline{R_Y(i,i)} = -\sum_{k=1}^{M} R_Y(k,k).$$

For $i = j, Y_{(i)}^H Y_{(j)}$ consists of the following multiplications:

$$y_i^*(k) y_j(k-p) \quad \text{for} \quad k = 0,1,\cdots, N-1-p$$
$$p = 0,1,\cdots, L.$$

However, for $i \neq j, Y_{(i)}^H Y_{(j)}$ consists of both the above multiplications and the following:

$$y_i^*(k-p) y_j(k) \quad \text{for} \quad k = 0,1,\cdots, N-1-p$$
$$p = 1,2,\cdots, L.$$

Based on the above analysis, one can show that for $N \gg L \gg 1$, the computational flops required by Step 1 (equivalently by the CR method) are $NM^2 L$. Note that the computation of the CR method is dominated by obtaining the covariance matrix $R_Y = Y^H Y$.

Implementation of Step 2 (Direct Form)

In Step 2, we need to compute the product $G^H G$, the pseudoinverse $(G^H G)^+$, the product $Y^H (G^H G)^+ Y$, and then the least eigenvector of $Y^H (G^H G)^+ Y$. It is easy to verify that the product $G^H G$ consists of $\overline{H}_{(i)} \overline{H}_{(j)}^H, i = 1, \cdots, M, j = i, \cdots, M$, and each $\overline{H}_{(i)} \overline{H}_{(j)}^H$ consists of

$$h_i(k) h_j^*(p) \quad \text{for} \quad k = 0,1,\cdots, L \quad p = k, \cdots, L.$$

To obtain the pseudoinverse $(G^H G)^+$, one can compute the eigendecomposition $G^H G = E_1 \Sigma_1 E_1^H$, where the zero eigenvalues are removed, Σ_1 is $(MN - N - L) \times (MN - N - L)$, and E_1 is $\frac{1}{2}(N-L)M(M-1) \times (MN - N - L)$. Then, one computes

$$Y^H (G^H G)^+ Y = (E_1^H Y)^H \Sigma_1^{-1} (E_1^H Y) \qquad (5.2)$$

where zeros in Y should be exploited. A detailed flop count of the above computations shows that Step 2 of the TSML method in the direct-form implementation is dominated by the eigendecomposition of $G^H G$, requiring $O\{N^3 M^6\}$ flops.

Implementation of Step 2 (Indirect Form)

Alternatively, one can use the following expression:

$$Y^H (G^H G)^+ Y = Y^H G^+ G^{H+} Y$$
$$= Y^H G^H (GG^H)^+ (GG^H)^+ GY$$
$$= Y^H G^H (GG^H)^{+2} GY. \qquad (5.3)$$

The product GG^H requires the same number of flops as $G^H G$. However, GG^H has the dimension $NM \times NM$, which is smaller than that of $G^H G$ when

$$M \geq \frac{2N}{N-L} + 1 \approx 3.$$

Let the eigendecomposition of GG^H be $E_2 \Sigma_2 E_2^H$, where the zero eigenvalues are removed, Σ_2 is $(NM-N-L) \times (NM-N-L)$, and E_2 is $NM \times (NM-N-L)$. Then, one can write the squared pseudoinverse: $(GG^H)^{+2} = E_2 \Sigma_2^{-2} E_2^H$. Using this in (5.3) yields

$$Y^H (G^H G)^+ Y = Y^H G^H E_2 \Sigma_2^{-2} E_2^H GY. \qquad (5.4)$$

To compue the above matrix, the following structure should be exploited:

$$GY = \begin{bmatrix} \overline{R_{HY}(1,1)} & R_{HY}(2,1) & \cdots & R_{HY}(M,1) \\ R_{HY}(1,2) & \overline{R_{HY}(2,2)} & \cdots & R_{HY}(M,2) \\ \vdots & \vdots & \ddots & \vdots \\ R_{HY}(1,M) & R_{HY}(2,M) & \cdots & \overline{R_{HY}(M,M)} \end{bmatrix}$$

$$(5.5)$$

where for $i = 1, \cdots, M$ and $j = 1, \cdots, M$

$$R_{HY}(i,j) = \overline{H}_{(i)}^H Y_{(j)}$$

$$\overline{R_{HY}(i,i)} = -\sum_{\substack{k=1 \\ k \neq i}}^{M} R_{HY}(k,k).$$

Furthermore, each $\overline{H}_{(i)}^H Y_{(j)}$ requires the multiplications $h_i^*(k) y_j(p), k = 0, 1, \cdots, L$ and $p = 0, 1, \cdots, N-1$.

A flop count shows that the eigendecomposition of GG^H dominates the indirect-form implementation of Step 2, requiring $O\{N^3 M^3\}$ flops.

Total Flops Required by the TSML Method

Combining the previous results shows that an efficient implementation of the TSML method requires $O\{NM^2 L\}$ flops for Step 1 (the CR method) and $O\{N^3 M^3\}$ flops for Step 2.

VI. SIMULATIONS

Although the TSML method is not meant for very long data sequences since the number of unknowns grows linearly with the data length, for applications where the data sequence is unavoidably short (e.g., in a rapid-changing environment), the TSML method should be used to achieve the optimum accuracy.

We have run simulations to compare the performance of the TSML method against those of the CR method [1], [2] and the SS method [15]. Based on simulations, we have observed the following:

1) When the channel condition is relatively ill (e.g., there are closely located zeros among channels so that the matrix H_M is ill conditioned), the TSML method always outperforms the CR and SS methods at relatively high SNR and approaches the CRB at relatively high SNR.

2) When the channel condition is relatively good, both the CR and SS methods attain the CRB at relatively high SNR, and the TSML method does not provide further improvement of accuracy at relatively high SNR.

3) Both the TSML method and the SS method in general performs better than the CR method.

4) When the first step of the TSML method is replaced by the SS method, the TSML method in general outperforms the SS method and approaches the CRB over a much wider range of SNR.

In our simulations, we considered a simple two-channel system, where each channel is a second-order FIR with the transfer function

$$\begin{aligned} H_i(z) &= h_i(0) + h_i(1)z^{-1} + h_i(2)z^{-2} \\ &= 1 - 2\cos(\theta_i)z^{-1} + z^{-2} \\ &= (1 - e^{j\theta_i} z^{-1})(1 - e^{-j\theta_i} z^{-1}) \qquad i = 1, 2. \end{aligned}$$

This implies that $M = 2$ and that $L = 2$. It is clear that θ_i represents the angular position of a corresponding zero on the unit circle. We will let $\theta_1 = \theta$ and $\theta_2 = \theta_1 + \delta$. Then, δ represents a distance between the zeros of the two channels. When δ is small, the channel is considered ill conditioned. (Common zero occurs when δ is zero.) In addition, note that when θ is $-\delta/2, \pi - \delta/2$ or $2\pi - \delta/2$, the two channels are identical and, of course, have common zeros. The channel outputs were corrupted by additive Gaussian white noise of variance σ_w^2. This system was driven by one realization of a white binary (1 or -1) process of variance $\sigma_s^2 = 1$. The number of the output samples from each channel was chosen to be $N = 30$. The SNR of the M-channel system was defined by

$$\text{SNR(dB)} = 10 \log_{10} \left(\frac{E\{\|\mathbf{y}\|^2\}}{E\{\|\mathbf{w}\|^2\}} \right)$$

which can be shown to be

$$\text{SNR(dB)} = 20 \log_{10} \left(\frac{\|\mathbf{h}\| \sigma_s}{\sqrt{M} \sigma_s} \right).$$

The performances of the CR, SS, and TSML methods were measured by the mean-square-error in decibels:

$$\text{MSE(dB)} = 20 \log_{10} \left(\frac{1}{\|\mathbf{h}\|} \sqrt{\frac{1}{N_r} \sum_{i=1}^{N_r} \|\hat{\mathbf{h}}_i - \mathbf{h}\|^2} \right)$$

where $\hat{\mathbf{h}}_i$ denotes the ith run estimate of $\mathbf{h}$, and the first element of $\hat{\mathbf{h}}_i$ is normalized to be one (the same as $\mathbf{h}$). N_r denotes the number of runs and was chosen to be 100. For each run, 100 independent realizations of the channel noise were used. (Each noise realization was used only once in our simulation.)

The CRB on the MSE was defined by

$$\text{CRB(dB)} = 20 \log_{10} \left(\frac{1}{\|\mathbf{h}\|} \sqrt{\text{tr}_h(\mathbf{F}^{-1})} \right)$$

where $\mathbf{F}$ is the FI matrix (defined by (3.2)) with the row and column corresponding to $h_1(0)$ deleted, and $\text{tr}_h(\mathbf{F}^{-1})$ denotes the sum of the diagonal elements of $\mathbf{F}^{-1}$ associated with $\mathbf{h}$. Note that $h_1(0)$ is normalized to
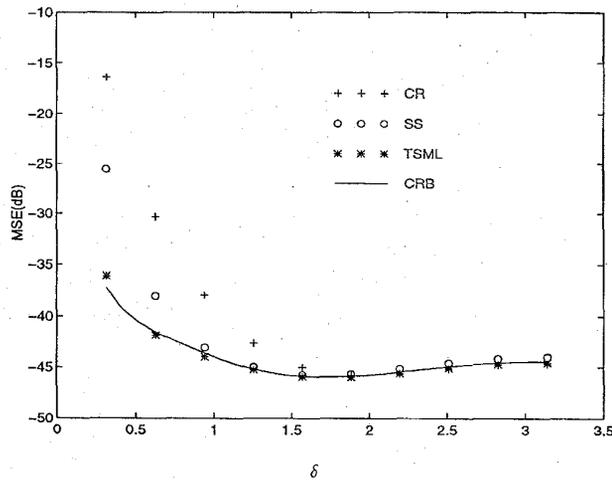
Fig. 1.   Performances of the CR, SS, and TSML methods versus δ, where SNR = 45 dB, and $\theta = 0$.

Fig. 2.   Performances of the CR, SS, and TSML methods versus θ, where SNR = 45 dB, and $\delta = \pi/11$.

be one for the caculation of the MSE and, hence, should be treated as a known parameter in the calculation of the FI matrix. In this example, both the input and the channel responses are real valued, and the unknown parameters are $h_1(1), h_1(2), h_2(0), h_2(1), h_2(2), s(-L), \cdots, s(N)$. Therefore, $\mathrm{tr}_h(F^{-1})$ is simply the sum of the first five diagonal elements of F^{-1}.

Fig. 1 shows the performances of the CR, SS, and TSML method against the channel condition (δ) at a relatively high SNR (45 dB). In the region of poor channel condition (when d is near zero), the TSML method performed better than both the CR and SS methods. In the region of good channel condition (when δ is near π), however, the three methods had comparable performances, and all approached the CRB. The TSML method always approached the CRB closely.

Fig. 2 shows the performances of the CR, SS, and TSML methods against the angular location of the first channel zero (θ) for a poor channel condition (small δ) and a relatively high SNR (45 dB). The TSML method performed consistently better than the other two methods. In the regions around the three points $\theta = -\delta/2, \pi - \delta/2$ or $2\pi - \delta/2$, the channel is very ill conditioned, and the differences between the three methods are very large.

Fig. 3 shows the performances of the CR, SS, and TSML methods against the SNR for a good channel condition (δ is π, and θ is $\pi/10$). The three methods had comparable performances, and all approached the CRB even at a very low SNR. The SNR threshold in this case is very good due to the good channel condition. The CR method and the TSML method are not distinguishable in this figure.

Fig. 4 shows the performances of the CR, SS, and TSML methods against the SNR for a poor channel condition (δ is $\pi/10$, and θ is $\pi/10$). The TSML method consistently performed better than the CR method. The TSML method also performed better than the SS method for relatively high SNR, although in the low SNR region, the SS method is more robust. The latter phenomenon is due to the poor initial estimates by the first step (the CR method) of the TSML
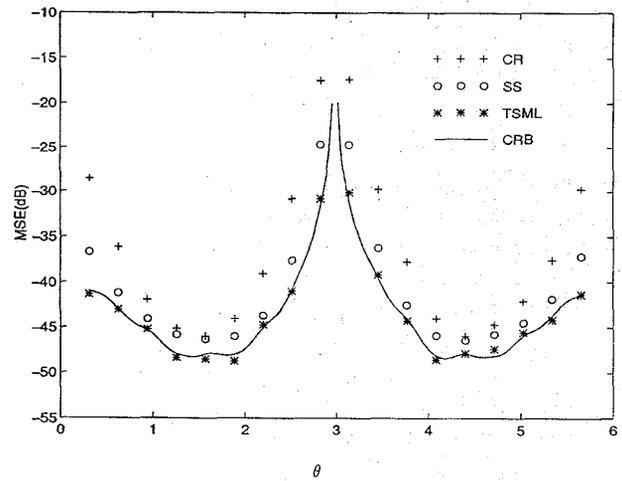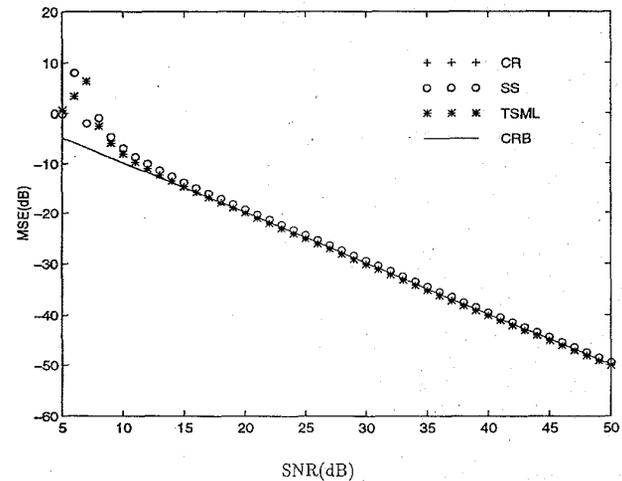
Fig. 3.   Performances of the CR, SS, and TSML methods versus SNR, where $\delta = \pi$, and $\theta = \pi/10$.

method. Nevertheless, in the region where the MSE is small, the SNR required by the TSML method (for a fixed MSE) is about 5 dB lower than the SS method and about 15 dB lower than the CR method.

Fig. 5 illustrates the situation where the first step of the TSML method is replaced by the SS method. It can be seen that the modified TSML method becomes much more robust to noise and outperforms the SS method over a much wider SNR range.

Finally, it should be mentioned that the TXK method [3] performs very poorly for relatively short data. The MES of the TXK method was too large to fit in these figures and, hence, was omitted in this paper. This phenomenon is consistent with the reports shown in [1], [2], [15], [16].

VII. CONCLUSIONS

The TSML method is a computationally efficient alternative to achieve the high-SNR optimum accuracy for this blind
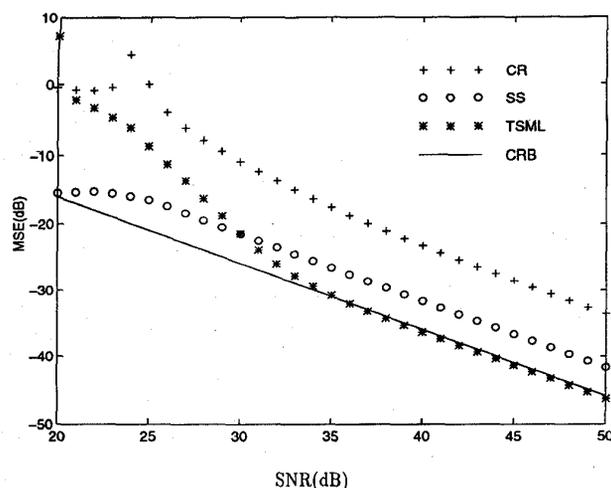
Fig. 4. Performances of the CR, SS, and TSML methods versus SNR, where $\delta = \pi/10$, and $\theta = \pi/10$.
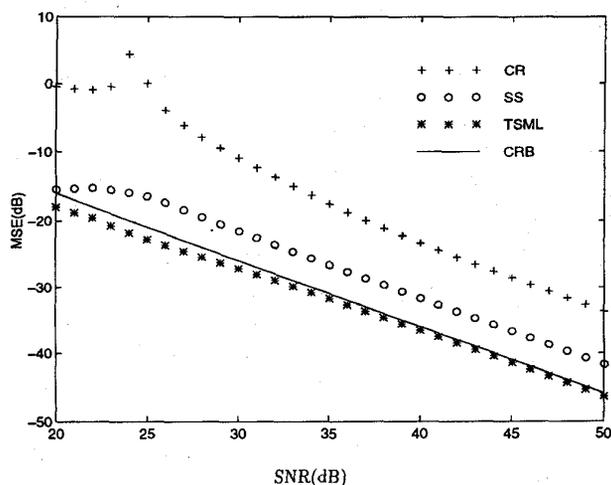
Fig. 5. Performances of the CR, SS, and TSML methods versus SNR, where $\delta = \pi/10$ and $\theta = \pi/10$, and the first step of the TSML method was replaced by the SS method.

system identification problem. Simulations have also supported that the TSML method performs significantly better than the CR and SS methods and approaches the CRB for a wide range of SNR. Replacing the first step of the TSML method by the SS method leads to an even more robust technique. The TSML method also provides an interesting link between the CR method and the ML principle. In studying the channel identifiability, the FI-based approach provides a new and interesting angle with respect to existing approaches. It is a pleasing result that the FI-based identifiability is equivalent to the CR-based identifiability under a mild condition. The orthogonal complement matrix G of the generalized Sylvester matrix H not only has played a critical role in developing the TSML method and understanding the FI matrix and the CR method but could also be useful in further investigation of the M-channel system. The role of the matrix G in understanding

the SS method [15] needs further research. Some initial results are shown in [31].

REFERENCES

[1] G. Xu, H. Liu, L. Tong, and T. Kailath, "A deterministic approach to blind channel identification," IEEE Trans. Signal Processing, to be published.

[2] H. Liu, G. Xu, and L. Tong, "A deterministic approach to blind channel identification of multi-channel FIR systems," in Proc. IEEE ICASSP'94, Adelaide, Australia, vol. 4, Apr. 1994, pp. 581–584.

[3] 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, no. 2, pp. 340–349, Mar. 1994.

[4] 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, no. 1, pp. 329–333, Jan. 1995.

[5] Y. Li and Z. Ding, "Blind channel identification based on second-order cyclostationary statistics," in Proc. IEEE ICASSP'93, Minneapolis, MN, Apr. 1993, pp. 81–84.

[6] G. B. Giannakis and J. M. Mendel, "Identification of nonminimum phase systems using higher order statistics," IEEE Trans. Acoust., Speech, Signal Processing, vol. 37, no. 3, pp. 360–377, Mar. 1989.

[7] B. Porat and B. Friedlander, "Blind equalization of digital communication channel using higher-order moments," IEEE Trans. Signal Processing, vol. 39, no. 2, pp. 522–526, Feb. 1991.

[8] J. K. Tugnait, "Fractionally spaced blind equalization and estimation of FIR channels," in Proc. Int. Conf. Commun., Geneva, Swizerland, June 1993, pp. 428–432.

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

[10] G. W. Stewart, "An updating algorithm for subspace tracking," IEEE Trans. Signal Processing, vol. 40, no. 6, pp. 1535–1541, June 1992.

[11] R. R. Bitmead, S.-Y. Kung, B. D. O. Anderson, and T. Kailath, "Greatest common division via generalized Sylvester and Bezout matrices," IEEE Trans. Automat. Contr., vol. 23, no. 6, pp. 1043–1047, Dec. 1978.

[12] H. L. Van Trees, Detection, Estimation, and Modulation Theory, Part I. New York: Wiley, 1968.

[13] M. Shao and C. L. Nikias, "An ML/MMSE estimation approach to blind equalization," in Proc IEEE ICASSP'94, Adelaide, Australia, vol. 4, Apr. 1994, pp. 569–572.

[14] D. T. M. Slock, "Blind fractionally-space equalization, perfect-reconstruction filter bank and multichannel linear prediction," in Proc. IEEE ICASSP'94, Adelaide, Australia, vol. 4, Apr. 1994, pp. 585–588.

[15] 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, no. 2, pp. 516–525, Feb. 1995.

[16] W. Qiu and Y. Hua, "Performance analysis of a blind channel identification method based on second-order statistics," in Proc. 28th Asilomar Conf. Signal, Syst. Comput., Nov. 1994.

[17] H. Yang and Y. Hua, "On performance of the cross relation method for blind channel identification," Signal Processing, to be published.

[18] Y. Hua and M. Wax, "Strict identifiability of multiple FIR channels driven by an unknown arbitrary sequence," IEEE Trans. Signal Processing, vol. 44, pp. 756–759, this issue.

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

[20] R. Kumaresan, L. L. Scharf, and A. K. Shaw, "An algorithm for pole–zero modeling and spectral analysis," IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-34, pp. 637–640, June 1986.

[21] Y. Bresler and A. Macovski, "Exact maximum likelihood parameter estimation of superimposed signals in noise," IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-34, pp. 1081–1089, Oct. 1986.

[22] A. G. Evans and R. Fischl, "Optimal least squares time-domain synthesis of recursive digital filters," *IEEE Trans. Audio Electroacoust.*, vol. AU-21, pp. 61–65, Feb. 1973.

[23] K. Steiglitz and L. E. McBride, "A technique for the identification of linear systems," *IEEE Trans. Automat. Contr.*, vol. AC-10, pp. 461–464, Oct. 1965.

[24] J. H. McClellan and D. Lee, "Exact equivalence of the Steiglitz–McBride iteration and IQML," *IEEE Trans. Acoust., Speech, Signal Processing,* vol. 39, pp. 509–512, Feb. 1991.

[25] M. Clark and L. L. Scharf, "On the complexity of IQML algorithms," *IEEE Trans. Signal Processing*, vol. 40, pp. 1811–1813, July 1992.

[26] Y. Hua, "The most efficient implementation of the IQML algorithm," *IEEE Trans. Signal Processing*, vol. 42, no. 8, pp. 2203–2204, Aug. 1994.

[27] P. Stoica and K. C. Sharman, "Maximum likelihood methods for direction-of-arrival estimation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, pp. 7, pp. 1132–1143, July 1990.

[28] S. Talwar, M. Viberg, and A. Paulraj, "Blind estimation of multiple co-channel digital signals using an antenna array," *IEEE Signal Processing Lett.*, vol. 1, no. 2, pp. 29–31, Feb. 1994.

[29] N. Seshadri, "Joint data and channel estimation using blind trelis search techniques," *IEEE Trans. Commun.*, vol. 42, no. 2/3/4, pp. 1000–1011, Feb. 1994.

[30] Y. Hua, "Fast maximum likelihood for blind channel identification of multiple FIR channels," in *Proc. 28th Asilomar Conf. Signals, Syst., Comput.,*, Oct 1994, pp. 415–419.

[31] Y. Hua, H. Yang, M. Wax, and M. Zhou, "Blind system identification using multiple sensors," in *Proc. IEEE ICASSP'95*, Detroit, MI, May 1995, pp. 3171–3174.

[32] Y. Hua, H. Yang, and W. Qiu, "Source correlation compensation for blind channel identification based on second order statistics," *IEEE Signal Processing Lett.*, vol. 1, no. 8, pp. 119–120, Aug. 1994.

Yingbo Hua (S'86–M'88–SM'92) was born in China in 1960. He received the B.S. degree from the Nanjing Institute of Technology (currently Southeast University), Nanjing, China, in February 1992 and the M.S. and Ph.D. degrees from Syracuse University, Syracuse, NY, USA, in 1983 and 1988, respectively.

From 1988 to 1989, he was a Postdoctoral Research Fellow at Syracuse University. In February 1990, he joined the University of Melbourne, Melbourne, Australia, where he was a Lecturer from 1990 to 1992, a Senior Lecturer from 1993–1995, and has been an Associate Professor and Reader since January 1996. He has won several large grants from the Australian Government and leads a group of research fellows and research students with the Australian Cooperative Research Centre on Sensor Signal and Information Processing. He has published over 35 journal papers and 55 conference papers in the areas of spectral estimation, array processing, radar imaging, and system identification. He is an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING and is a member of a number of international technical committees.