

# Multichannel Blind Identification: From Subspace to Maximum Likelihood Methods

LANG TONG, MEMBER, IEEE, AND SYLVIE PERREAU

*Invited Paper*

*A review of recent blind channel estimation algorithms is presented. From the (second-order) moment-based methods to the maximum likelihood approaches, under both statistical and deterministic signal models, we outline basic ideas behind several new developments, the assumptions and identifiability conditions required by these approaches, and the algorithm characteristics and their performance. This review serves as an introductory reference for this currently active research area.*

**Keywords**—Blind equalization, parameter estimation, system identification.

## I. INTRODUCTION

### A. What Is Blind Channel Estimation and Why?

There have been considerable interests from both signal processing and communications communities in the so-called “blind” problem. This is evident from titles of recent publications in both societies’ journals and annual conferences. The basic blind channel estimation problem involves a channel model shown in Fig. 1, where only the observation signal  $y$  is available for processing in the identification and estimation of channel  $\mathcal{H}$ . This is in contrast to the classical input–output system identification and estimation problem where both input  $s$  and observation  $y$  are used.

The impetus behind the increased research activities in blind techniques is perhaps their potential applications in wireless communications, which are currently experiencing explosive growth. For example, the distortion caused by multipath interference affects both transmission quality and efficiency in wireless communications. Conventional designs of receivers that mitigate such distortions require either the knowledge of the channel or the access to the in-

put so that certain “training” signals can be transmitted. The latter is the case in many, if not most, communication systems design. The transmission of training signals obviously decreases communications throughput although, for time invariant channels, the loss is insignificant because only one training is necessary. For time varying channels, however, the loss of throughput becomes an issue. For example, in high-frequency (HF) communications, the time used to transmit training signals can be as much as 50% of the overall transmission. Even the group special mobile (GSM) system for cellular mobile communication has considerable overhead associated training. Yet another example, described by Godard [46] in his pioneer work in blind equalization, is in computer networks where links between terminal and central computers need to be established in an asynchronous way such that, in some instances, training is impossible. Another example is the potential application of blind equalization in high-definition television (HDTV) broadcasting [35]. Outside of the communications arena, blind channel estimation has long been an interest in geoscience. Some of the earlier pioneers in this field include Robinson [99], Wiggins [133], and Donoho [24]. Recent results can be found in [79] and [97]. Blind channel estimation also has applications in image restoration problems. Already, blind estimation techniques have been proposed for image deblurring applications [40], [42].

At first glance, the estimation problem illustrated in Fig. 1 may not seem tractable. How is it possible to distinguish the signal from the channel when neither is known? The essence of blind channel estimation rests on the exploitation of structures of the channel and properties of the input. A familiar case is when the input has known probabilistic description, such as distributions and moments. In such a case, the problem of estimating the channel using the output statistics is related to time series analysis. In communications applications, for example, the input signals may have the finite alphabet property, or sometimes exhibit cyclostationarity. This last property was exploited in [118] to demonstrate the possibility of estimating a nonminimum

Manuscript received May 23, 1997; revised May 8, 1998. This work was supported in part by the National Science Foundation under Contract NCR-9321813 and by the Office of Naval Research under Contract N00014-96-1-0895.

L. Tong is with the School of Electrical Engineering, Cornell University, Ithaca, NY 14853 USA (e-mail: ltong@ee.cornell.edu).

S. Perreau is with the Institute for Telecommunications Research, The Levels, SA 5095 Australia (e-mail: sylvie@spri.levels.unisa.edu.au).

Publisher Item Identifier S 0018-9219(98)06969-2.

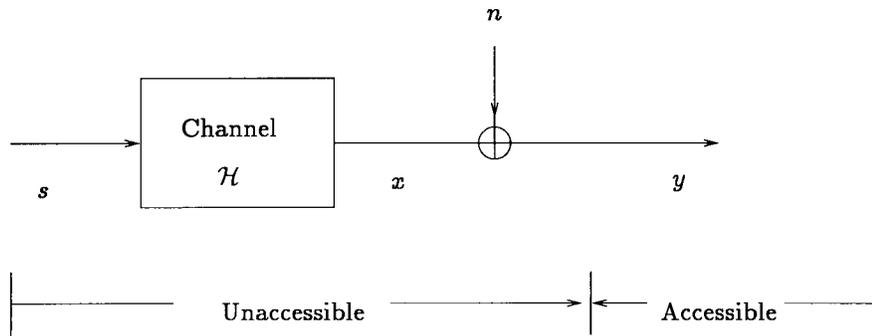


Fig. 1. Schematic of blind channel estimation.

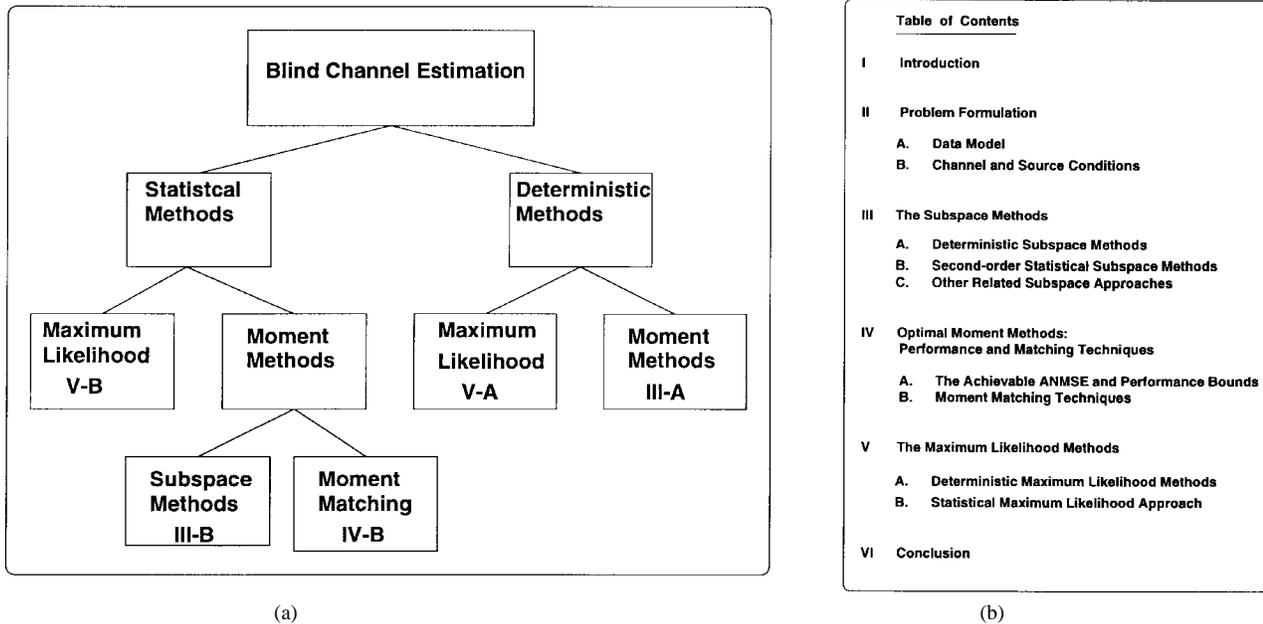


Fig. 2. (a) Classification of blind channel estimators and (b) contents of the paper.

phase channel using only the second-order statistics, which led to the development of many subspace-based blind channel estimation algorithms.

### B. The Goal and the Scope of this Paper

By complementing recent surveys [72], [73], the goal of this paper is to review developments in blind channel identification and estimation within the estimation theoretical framework. We have paid special attention to the issue of identifiability, which is at the center of all blind channel estimation problems. Various existing algorithms are classified into the moment-based and the maximum likelihood (ML) methods. We further divide these algorithms based on the modeling of the input signal. If input  $s$  is assumed to be random with prescribed statistics (or distributions), the corresponding blind channel estimation schemes are considered to be statistical. On the other hand, if the source does not have a statistical description, or although the source is random but the statistical properties of the source are not exploited, the corresponding estimation algorithms are classified as deterministic. Fig. 2 shows a map for different classes of algorithms and the organization of the paper.

Space limitations force us to be brief in discussing some algorithms and, unfortunately, to exclude other important approaches. We have excluded discussions related to the “dual” problem of blind channel estimation, namely, the blind signal estimation, where the goal is to estimate or detect the input signal without knowing the channel. Formulated as blind signal estimation problems are some of the first applications of blind estimation (equalization) method in communications, including the celebrated work of Godard [46], Sato [101], and Trechler and Agee [124]. These earlier works are based on some forms of higher order statistics. Under the multichannel model, direct blind equalization becomes possible using only the second-order statistics, which potentially may have faster convergence rates. It is shown by Liu and Dong [75] that in the absence of noise, a whitening filter is in fact a perfect equalizer. Direct equalization using the second-order statistics was first recognized by Slock [108], and there are a number of new developments [16], [34], [45], [74], [76], [77], [110]. Also not considered here are neural network-based direct blind signal estimation techniques (see [6]).

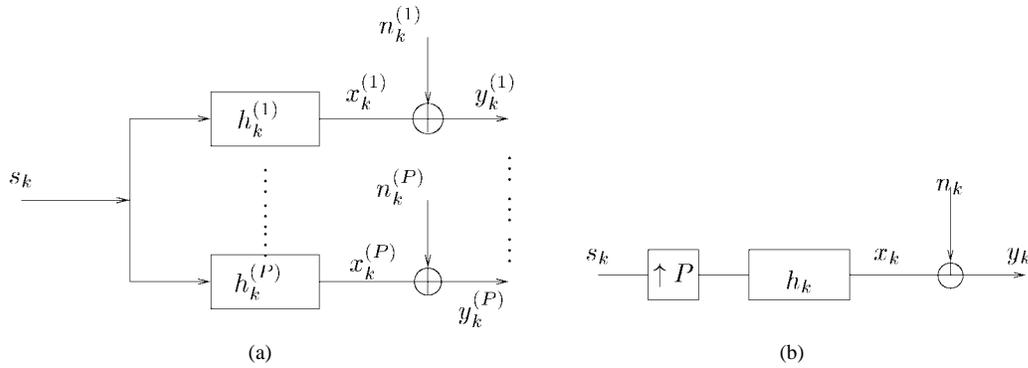


Fig. 3. (a) A multichannel model and (b) a multirate channel model.

In presenting moment-based methods, this paper focuses on the second-order moment techniques that have received considerable research attention since the publication of [118]. Consequently, applications of the algorithms surveyed in this paper are limited to certain types of channels and sources specified by the identifiability condition. Without elaboration, we briefly mention two alternative approaches.

1) *The Higher Order Statistical Approaches:* Many applications may not have the multichannel model considered in this paper. In such a case it may be necessary to exploit higher order statistics. There is an extensive literature dealing with blind channel estimation using higher order statistics in both time and frequency domains. See, for example, [12], [30], [39], [54], [90], [125], and [126] and a tutorial in [81]. For the multiuser case, see [38], [115], and [116].

2) *The Bayesian Approach:* In this paper, the channel is modeled by finite dimensional deterministic unknown parameters. In some applications, however, channels can be modeled as a random vector or a random process. For example, Rayleigh fading channels can be modeled as a Gaussian random process with a certain power spectrum. In such cases we have a Bayesian estimation problem. State-space model of the channel is also used in some applications where the extended Kalman filter can be applied [67]. More recent approaches can be found in [60], [61], and [68].

3) *Notations:* Most notations are standard: vectors and matrices are boldface small and capital letters, respectively; the matrix transpose, the complex conjugate, the Hermitian, and pseudoinverse are denoted by  $(\cdot)^t$ ,  $(\cdot)^*$ ,  $(\cdot)^H$ , and  $(\cdot)^\dagger$ , respectively;  $\otimes$  stands for the Kronecker product;  $\mathbf{I}_p$  is the  $p \times p$  identity matrix;  $E(\cdot)$  is the mathematical expectation.

## II. PROBLEM FORMULATION

In contrast to classical input-output channel identification and estimation problems, the so-called blind channel identification and estimation involves only the channel output. The basic problem considered here is: given the received (perhaps noisy) signal, estimate the channel impulse response.

### A. Channel Models

We consider two equivalent channel models shown in Fig. 3. The multichannel model is natural in applications

involving multiple receivers, whereas the multirate channel model comes directly from communication problems involving linear modulations.

1) *The Multichannel Model:* Considered in this paper is the identification and estimation of a discrete-time linear single-input  $P$ -output channel model shown in Fig. 3(a). Denoting the vector impulse response and its  $Z$ -transform by

$$\mathbf{h}_k \triangleq [h_k^{(1)}, \dots, h_k^{(P)}]^t \stackrel{Z}{\Leftrightarrow} \mathbf{h}(z) \quad (1)$$

we have the following system equations:

$$\mathbf{x}_k = \sum_{i=0}^L \mathbf{h}_i s_{k-i} \quad \mathbf{y}_k = \mathbf{x}_k + \mathbf{n}_k \quad (2)$$

where  $\mathbf{x}_k = [x_k^{(1)}, \dots, x_k^{(P)}]^t$  is the noiseless channel output and  $\mathbf{y}_k$  is the received (noisy) signal.

It is often convenient to consider the channel model for a block of  $q$  consecutive samples. Denoting

$$\begin{aligned} \vec{\mathbf{x}}_q^t(k) &= [\mathbf{x}_k^t, \dots, \mathbf{x}_{k-q+1}^t] \\ \vec{\mathbf{y}}_q^t(k) &= [\mathbf{y}_k^t, \dots, \mathbf{y}_{k-q+1}^t] \end{aligned} \quad (3)$$

$$\begin{aligned} \vec{\mathbf{n}}_q^t(k) &= [\mathbf{n}_k^t, \dots, \mathbf{n}_{k-q+1}^t] \\ \vec{\mathbf{s}}_{q+L}^t(k) &= [s_k, \dots, s_{k-L-q+1}] \end{aligned} \quad (4)$$

$$\vec{\mathbf{h}}^t = [\mathbf{h}_0^t, \dots, \mathbf{h}_L^t] \quad (5)$$

we then have

$$\vec{\mathbf{x}}_q(k) = \mathcal{F}_q(\vec{\mathbf{h}}) \vec{\mathbf{s}}_{q+L}(k) = \mathcal{T}_q(\vec{\mathbf{s}}(k)) \vec{\mathbf{h}} \quad (6)$$

$$\vec{\mathbf{y}}_q(k) = \vec{\mathbf{x}}_q(k) + \vec{\mathbf{n}}_q(k) \quad (7)$$

where the  $qP \times (L+q)$  filtering matrix  $\mathcal{F}_q(\mathbf{h})$  and the data matrix  $\mathcal{T}_q(\vec{\mathbf{s}}(k))$  are defined by

$$\begin{aligned} \mathcal{F}_q(\vec{\mathbf{h}}) &\triangleq \underbrace{\begin{pmatrix} \mathbf{h}_0 & \cdots & \mathbf{h}_L & & \\ & \ddots & & \ddots & \\ & & \mathbf{h}_0 & \cdots & \mathbf{h}_L \end{pmatrix}}_{L+q \text{ columns}} \\ \mathcal{T}_q(\vec{\mathbf{s}}(k)) &\triangleq \begin{pmatrix} s_k & s_{k-1} & \cdots & s_{k-L} \\ s_{k-1} & s_{k-2} & \cdots & s_{k-L-1} \\ \vdots & \vdots & \cdots & \vdots \\ s_{k-q+1} & s_{k-q} & \cdots & s_{k-q-L+1} \end{pmatrix} \otimes \mathbf{I}_P. \end{aligned} \quad (8)$$

We drop the subscript  $q$  whenever its omission does not cause confusion.

2) *Multirate Channel Model*: An equivalent alternative to multichannel representation is the multirate channel model shown in Fig. 3(b). If the input sequence  $s_k$  is wide sense stationary, then  $y_k$  is wide sense cyclostationary with period  $P$ . It is the cyclostationarity of the received signal that makes the identification using second-order statistics possible. See [32] for detailed discussions.

The system equations are given by

$$x_k = \sum_i s_i h_{k-iP} \quad y_k = x_k + n_k. \quad (9)$$

The relation between  $\{h_k, x_k, y_k\}$  in the multirate channel model and  $\{h_k^{(i)}, x_k^{(i)}, y_k^{(i)}\}$  in the multichannel model is given by

$$\{x_k\} = \{x_k^{(1)}, \dots, x_k^{(P)}\} \quad (10)$$

$$\{y_k\} = \{y_k^{(1)}, \dots, y_k^{(P)}\} \quad (11)$$

### B. Channel and Source Conditions

As in classical system identification problems, certain conditions about the channel and the source must be satisfied to ensure identifiability. In the multichannel blind identification problem, two conditions are shared by many different approaches.

1) *Channel Diversity*: What makes identification of the multichannel model different from that of the single channel case is the channel diversity. By diversity we mean that different subchannels have different modes. When they are modeled as finite impulse response channels, this means that they have different zeros, or in other words, they are coprime.

The significance of coprimeness among subchannels can be understood more clearly in the deterministic setting where no stochastic models are used to describe the input sequence. Consider the multichannel model shown in Fig. 3. If the subchannels are not coprime, then there exists a common factor  $c(z)$  such that

$$h^{(i)}(z) = c(z)\tilde{h}^{(i)}(z) \quad i = 1, \dots, P. \quad (12)$$

Consequently, without further information, it is difficult to distinguish whether  $c(z)$  is part of the input signal or part of the channel. In fact, one can replace  $c(z)$  by factors of the input sequence without affecting the observation. Therefore, the identification cannot be made unique (unless other properties of the input sequence and the channel are used).

Another ramification of coprimeness among subchannels is the existence of finite impulse response (FIR) inverse. When there is only one channel it is not possible to reconstruct the input sequence by using an FIR filter. This is not necessarily true in the multichannel case. Indeed, there exists a bank of filters  $\mathbf{g}(z)$  such that

$$\mathbf{g}^t(z)\mathbf{h}(z) = 1 \quad (13)$$

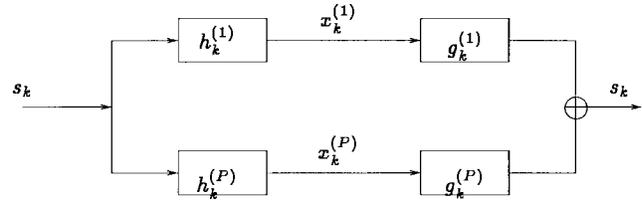


Fig. 4. A multichannel with FIR inverse.

if and only if subchannels are coprime. Equation (13) is the so-called Bezout equation. In other words, given the noiseless channel output  $\mathbf{x}_k$ , the channel input  $s_k$  can be reconstructed using a bank of FIR filters  $g_i(z)$  as shown in Fig. 4. In communication applications, filter  $\mathbf{g}(z)$  is often referred to as the zero-forcing receiver for it completely eliminates the channel distortion. The concept of FIR inverse of a vector FIR channel has been exploited in coding [80].

The coprimeness has several equivalent forms, and they are summarized below.

*Property 1*: Subchannels  $\{h^{(i)}(z)\}$  do not share common zeros:

- P1) if and only if there exists a vector polynomial  $\mathbf{g}(z)$  such that

$$\mathbf{g}^t(z)\mathbf{h}(z) = 1; \quad (14)$$

- P2) if and only if  $\mathcal{F}_q(\vec{\mathbf{h}})$  has full column rank for some  $q$ .

P1) comes from a basic property of ring (e.g., see [31, p. 10]). P2) was shown by Sylvester (in 1840) as a test for coprimeness [31], [63] (also see [29]). This condition plays a particularly important role in the development of many blind identification algorithms based on second-order statistical and deterministic formulations.

2) *Linear Complexity and Persistent Excitation*: Linear complexity measures the predictability of a finite-length deterministic sequence. Persistent excitation measures the richness of the infinite-length signal in the frequency domain. Both concepts are relevant in the deterministic setting of blind channel identification where the source is not assumed to have a probabilistic model. We adopt the following definition given in [10] and [47].

*Definition 1*: The linear complexity of sequence  $\{s_k\}_{k=0}^n$  is defined as the smallest value of  $c$  for which there exists  $\{\lambda_j\}$  such that

$$s_i = -\sum_{j=1}^c \lambda_j s_{i-j} \quad i = c, \dots, n. \quad (15)$$

An infinite sequence  $\{s_k\}$  is weakly persistently exciting of order  $c$  if

$$\rho_1 \mathbf{I}_c > \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=1}^N \begin{pmatrix} s_{k+c} \\ \vdots \\ s_{k+1} \end{pmatrix} (s_{k+c}^* \cdots s_{k+1}^*) > \rho_2 \mathbf{I}_c \quad (16)$$

where  $\rho_i > 0$ .

A connection between linear complexity and persistent excitation can be observed through the sample covariance of the input sequence, which enters the definition of persistent excitation directly. We define a Toeplitz matrix  $\mathbf{S}_c$  by

$$\mathbf{S}_c \triangleq \begin{pmatrix} s_c & s_{c-1} & \cdots & s_0 \\ s_{c+1} & s_c & \cdots & s_1 \\ \vdots & \vdots & \cdots & \vdots \\ s_n & s_{n-1} & \cdots & s_{n-c} \end{pmatrix}. \quad (17)$$

If  $s_k$  has linear complexity  $c$  or greater, then  $\mathbf{S}_c$  has full column rank. Hence, the sample covariance of the vector sequence  $\mathbf{s}(k) = [s_k, s_{k-1}, \dots, s_{k-c}]^T$  has full rank. On the other hand, if  $s_k$  has linear complexity no greater than  $c$ , the sample covariance of  $\mathbf{s}_k$  is rank deficient. For a quasi-stationary sequence, persistent excitation implies that the sample covariance matrix is full rank and its spectrum has at least  $L$  nonzero points (see [78]).

### III. THE SUBSPACE METHODS

Many recent blind channel estimation techniques exploit subspace structures of observation. The key idea is that the channel (or part of the channel) vector is in a one-dimensional subspace of either the observation statistics or a block of noiseless observations. These methods, which are often referred to as subspace algorithms, have the attractive property that the channel estimates can often be obtained in a closed form from optimizing a quadratic cost function

$$\hat{\mathbf{h}} = \arg \min_{\mathbf{h} \in \mathcal{S}} \bar{\mathbf{h}}^H \mathbf{Q} \bar{\mathbf{h}} \quad (18)$$

where  $\mathcal{S}$  is a set that specifies the domain of  $\bar{\mathbf{h}}$ .

Subspace methods can sometimes be considered part of the moment methods. They are attractive because of the closed-form identification. On the other hand, as they rely on the property that the channel lies in a unique direction (subspace), they may not be robust against modeling errors, especially when the channel matrix  $\mathcal{F}(\bar{\mathbf{h}})$  is close to being singular. The second disadvantage is that they are often more computationally expensive.

#### A. Deterministic Subspace Methods

Deterministic subspace methods do not assume that the input source has a specific statistical structure. Perhaps a more striking property of deterministic subspace methods is the so-called finite sample convergence property. Namely, when there is no noise, the estimator produces the exact channel using only a finite number of samples, provided that, of course, the identifiability condition is satisfied. Therefore, these methods are most effective at high SNR and for small data sample scenarios. On one hand, deterministic methods can be applied to a much wider range of source signals; on the other hand, not using the source statistics affects its asymptotic performance, especially when the identifiability condition is close to be violated.

1) *Assumptions and Identifiability:* Deterministic subspace methods assume the following conditions.

*Assumption 1:*

- 1.1) The noise  $\mathbf{n}_k$  is zero mean, white with known covariance  $\sigma^2 \mathbf{I}$ .
- 1.2) The channel has known order  $L$ .

The assumption that  $L$  is known may not be practical. To address this problem, there are three approaches. First, channel order detection and parameter estimation can be performed separately. There are well-known order detection schemes that can be used in practice (e.g., see [5], [96], and [129]). Second, some statistical subspace methods [3] require only the upper bound of  $L$ . Third, channel order detection and parameter estimation can be performed jointly [123]. Similarly, the noise variance  $\sigma^2$  may not be known in practice, but it can be estimated in many ways. For example, the noise variance estimation and channel order detection can be performed jointly using singular values of the estimated covariance matrix [129].

For deterministic methods it is necessary to impose conditions on the input sequence  $s_k$ , which significantly complicates the identifiability condition. Xu *et al.* in [134] and Hua and Wax in [59] gave the necessary and sufficient conditions of identifiability. Here we present only a sufficient condition for identifiability.

*Theorem 1:* Under Assumption 1, the channel  $\mathbf{h}(z)$  (or  $\bar{\mathbf{h}}$ ) can be uniquely identified up to a constant factor from the noiseless observation  $\mathbf{x}_k$  if

- 1) subchannels are coprime;
- 2) the source sequence has linear complexity greater than  $2L$ .

The condition that subchannels are coprime is also necessary for identifiability. It was shown in [134] that it is necessary that the linear complexity of the source, characterized by “modes” in [134], is greater than  $L$ . When the source is a realization of an ergodic process, the linear complexity condition is satisfied automatically, and we have the same identifiability condition as the stochastic formulation. The persistent excitation of the source, along with the coprime condition of subchannels certainly also ensures the identifiability.

2) *The Cross Relation Approach:* The cross relation (CR) approach, a termed coined by Hua [56], wisely exploits the multichannel structure. This algorithm was discovered independently and in different forms by Liu *et al.* [71], [134], Güreli and Nikias [50], [51], Baccala and Roy [7], [8], and Robinson [98]. An adaptive implementation using neural network is presented in [22] and [23].

Consider the noiseless multichannel model involving channels  $i$  and  $j$  shown in Fig. 5. We simply have for all  $i$  and  $j$

$$x_k^{(j)} * h_k^{(i)} - x_k^{(i)} * h_k^{(j)} = 0. \quad (19)$$

In the matrix form, we have

$$[\mathbf{X}_j - \mathbf{X}_i] \begin{pmatrix} \bar{\mathbf{h}}_i \\ \bar{\mathbf{h}}_j \end{pmatrix} = \mathbf{0}, \quad \bar{\mathbf{h}}_i = \begin{pmatrix} h_0^{(i)} \\ \vdots \\ h_L^{(i)} \end{pmatrix} \quad (20)$$

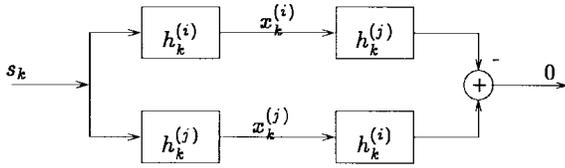


Fig. 5. The cross relation between two channels.

where  $\mathbf{X}_j$  can be constructed from the  $N$  received data samples

$$\mathbf{X}_j = \begin{pmatrix} x_L^{(j)} & x_{L-1}^{(j)} & \cdots & x_0^{(j)} \\ x_{L+1}^{(j)} & x_L^{(j)} & \cdots & x_1^{(j)} \\ \vdots & \vdots & \cdots & \vdots \\ x_{N-1}^{(j)} & x_{N-2}^{(j)} & \cdots & x_{N-L-1}^{(j)} \end{pmatrix}. \quad (21)$$

Of course, one can consider all possible pairs to obtain the following identification equation

$$\mathcal{D}^H(\mathbf{X}) \begin{pmatrix} \bar{\mathbf{h}}_1 \\ \vdots \\ \bar{\mathbf{h}}_P \end{pmatrix} = \mathbf{0} \quad (22)$$

where  $\mathcal{D}(\cdot)$  is the data selection transform [71], [138] as shown in (23), shown at the bottom of the page. It is shown in [134] that, under the identifiability conditions,  $\mathcal{D}^H(\mathbf{X})$  is column-rank deficient by one. Hence, the solution of (22) provides the channel identification up to a scaling factor. When noise presents only  $\{y_k^{(j)}\}$  are available. The Cramér–Rao approach minimizes the following least squares (LS) cost

$$\hat{\mathbf{h}} = \arg \min_{\|\bar{\mathbf{h}}\|=1} \|\mathcal{D}^H(\mathbf{Y})\bar{\mathbf{h}}\|_2^2. \quad (24)$$

Equivalently, the channel estimate  $\hat{\mathbf{h}}$  can be obtained from the singular vector of  $\mathcal{D}(\mathbf{Y})\mathcal{D}^H(\mathbf{Y})$  associated with the smallest singular value. It can be shown further [138] that  $\mathcal{D}(\mathbf{Y})\mathcal{D}^H(\mathbf{Y})$  can be replaced by the sample covariance of  $\bar{\mathbf{y}}(k)$ . By subtracting the noise statistics, a mean-square consistent estimator can be obtained.

a) *Algorithm characteristics:* Unlike statistical methods, the CR method is very effective for small data sample applications at high SNR. Under the condition that subchannels are coprime and linear complexity conditions, observations of  $3L + 1$  samples are sufficient [59]. In simulations, Hua [56] showed that CR method combined with the ML approach offers performance close to the Cramér–Rao lower bound. The main problem of the CR

method is that the channel order  $L$  cannot be over estimated, in contrast to some of the statistical subspace approaches. For finite samples, this algorithm may also be biased.

3) *Noise Subspace Approach:* The (noise) subspace method, proposed by Moulines *et al.* [82], [83], exploits the structure of the filtering matrix  $\mathcal{F}_q(\vec{\mathbf{h}})$  directly. The basic idea is to force the signal space to have the block Toeplitz form of  $\mathcal{F}_q(\vec{\mathbf{h}})$ . The dual of this approach is to force the Toeplitz structure of  $\mathcal{T}_q(\vec{\mathbf{s}}(k))$  presented in [127], thus both can be considered as forms of subspace intersection. See [127] for this connection.

Suppose that  $\mathbf{v} = [v_1, \dots, v_{Pq}]$  is in the orthogonal complement of the range space of  $\mathcal{F}_q(\vec{\mathbf{h}})$ , i.e.,

$$\underbrace{[v_1^*, \dots, v_P^*]}_{\mathbf{v}_1^H} \cdots \underbrace{[v_{(q-1)P+1}^*, \dots, v_{Pq}^*]}_{\mathbf{v}_q^H} \cdot \underbrace{\begin{pmatrix} \mathbf{h}_0 & \cdots & \mathbf{h}_L \\ & \ddots & \\ & & \mathbf{h}_0 & \cdots & \mathbf{h}_L \end{pmatrix}}_{L+q \text{ columns}} = \mathbf{0}. \quad (25)$$

The above can also be written as a linear equation with respect to the channel parameter  $\vec{\mathbf{h}}$ . Specifically, we have

$$\underbrace{[\mathbf{h}_0^H \cdots \mathbf{h}_L^H]}_{L+q \text{ columns}} \underbrace{\begin{pmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_q \\ & \ddots & \\ & & \mathbf{v}_1 & \cdots & \mathbf{v}_q \end{pmatrix}}_{L+q \text{ columns}} = \vec{\mathbf{h}}^H \mathcal{F}_{(L+1)}(\mathbf{v}) = \mathbf{0}. \quad (26)$$

Here,  $\mathbf{v}_k$  is the  $k$ th subvector of  $\mathbf{v}$ :  $\mathbf{v}_k = [v_{(k-1)q+1}, \dots, v_{kq}]^t$ . The above equation can be used to identify the channel vector  $\vec{\mathbf{h}}$  provided that (26) has a unique solution. Moulines *et al.* gave the following theorem.

*Theorem 2 [83]:* Let  $\mathcal{N}(\vec{\mathbf{h}}) = \text{span}\{\mathbf{v}_1, \dots, \mathbf{v}_d\}$  be the orthogonal complement of the column space of  $\mathcal{F}_q(\vec{\mathbf{h}})$ . For any  $\vec{\mathbf{h}}$  and  $\mathbf{h}_1$  satisfying the condition that subchannels are coprime,  $\mathcal{N}(\vec{\mathbf{h}}) = \mathcal{N}(\mathbf{h}_1)$  if and only if  $\vec{\mathbf{h}} = \alpha \mathbf{h}_1$ . Further, for all  $i$ ,  $\vec{\mathbf{h}}$  satisfies the following equation:

$$\mathcal{F}_{(L+1)}^H(\mathbf{v}_i)\vec{\mathbf{h}} = \mathbf{0}. \quad (27)$$

Having the estimated basis  $\{\hat{\mathbf{v}}_i\}$  of the orthogonal complement of  $\mathcal{F}_q(\vec{\mathbf{h}})$ , identification of channel can be accomplished by the following optimization:

$$\hat{\mathbf{h}} = \arg \min_i \sum_i \vec{\mathbf{h}}^H \mathcal{F}_{(L+1)}(\mathbf{v}_i) \mathcal{F}_{(L+1)}^H(\mathbf{v}_i) \vec{\mathbf{h}}. \quad (28)$$

$$\mathcal{D}(\mathbf{X}) \triangleq \begin{pmatrix} \mathbf{X}_2 & \mathbf{X}_3 & \cdots & \mathbf{X}_P & \vdots & \vdots & \vdots \\ -\mathbf{X}_1 & & & & \vdots & \mathbf{X}_3 & \cdots & \mathbf{X}_P & \vdots & \vdots \\ & -\mathbf{X}_1 & & & \vdots & -\mathbf{X}_2 & & & \vdots & \cdots & \vdots \\ & & \ddots & & \vdots & & \ddots & & \vdots & & \mathbf{X}_P \\ & & & -\mathbf{X}_1 & \vdots & & & & \vdots & & -\mathbf{X}_{P-1} \\ & & & & \vdots & & & & \vdots & & \vdots \\ & & & & -\mathbf{X}_1 & \vdots & & & -\mathbf{X}_2 & \vdots & \vdots & \vdots & \vdots \end{pmatrix}^* \quad (23)$$

With the above theorem, the estimation of the channel can now be accomplished by first estimating the orthogonal complement of the  $\mathcal{F}_q(\vec{\mathbf{h}})$ . This can be achieved in a number of ways. One of the frequently used approaches is the signal-noise space decomposition. From the multichannel model, we have

$$\begin{aligned}\hat{\mathbf{R}}_y &= \frac{1}{N} \sum_{k=0}^{N-1} \vec{\mathbf{y}}(k) \vec{\mathbf{y}}^H(k) \xrightarrow{N \rightarrow \infty} \mathbf{R} \\ &= \mathcal{F}_q(\vec{\mathbf{h}}) \hat{\mathbf{R}}_s \mathcal{F}_q^H(\vec{\mathbf{h}}) + \sigma^2 \mathbf{I}.\end{aligned}\quad (29)$$

The singular value decomposition of  $\mathbf{R}$  has the form

$$\mathbf{R} = \mathbf{U} \text{diag}\{\lambda_1^2 + \sigma^2, \dots, \lambda_{L+q}^2 + \sigma^2, \sigma^2, \dots, \sigma^2\} \mathbf{U}^H \quad (30)$$

where  $\lambda_i$  are the singular values of  $\mathcal{F}_q(\vec{\mathbf{h}})$ . If subchannels are coprime, i.e.,  $\mathcal{F}_q(\vec{\mathbf{h}})$  is full column rank, the orthogonal complement of the range space of  $\mathcal{F}_q(\vec{\mathbf{h}})$ , also referred to as the noise subspace, is given by the singular vectors  $\{\mathbf{u}_i\}_{i=L+q+1}^{Pq}$  of  $\mathbf{R}$  associated with the singular value  $\sigma^2$ . Note that when there is no noise  $\mathcal{N}(\mathbf{h})$  can be obtained directly, using only a finite number of data samples from the eigen-decomposition of the data matrix  $\mathbf{Y} = [\vec{\mathbf{y}}(0) \vec{\mathbf{y}}(1), \dots]$ . Incidentally, vectors in noise space can also be viewed as linear prediction error (or the blocking) filters. With this interpretation, Slock presented a linear prediction-based subspace approach [108].

*a) Algorithm characteristics:* There is a strong connection between the CR and the noise subspace approaches. As pointed out in [2], they are different only in their choices of parameterizing the signal and the noise subspaces. For a special but important case when  $P = 2$  [138], these two algorithms are in fact identical. Similar to the CR method, the noise subspace method also requires the knowledge of the channel order  $L$ . Overdetermination of the channel order renders the algorithm ineffective without additional processing. The noise subspace method is also suitable for short data size applications. Although it is a bit more complex than the CR method, it appears to offer improved performance in many simulations. Several extensions have been obtained. Hua *et al.* [57] investigated the minimum noise subspace for channel identification. The multiuser case is presented in [2].

*4) Identification via Least Squares Smoothing:* Although deterministic approaches enjoy the advantage of having fast convergence, they share some common difficulties. For example, the determination of channel order is required and often difficult. Second, the adaptive implementation of these algorithms is not straightforward. Recently, Tong and Zhao proposed approaches based on the least squares smoothing (LSS) of the observation process [121]–[123], [141], [142].

The key idea of LSS rests on the isomorphic relation between the input and the observation spaces. Given the input sequence  $s_k$  and the noiseless observation  $\mathbf{x}_k$ , we

define

$$\begin{aligned}\mathbf{s}_t &\triangleq [s_t, s_{t+1}, \dots] \\ \mathbf{X}_t &\triangleq [\mathbf{x}(t), \mathbf{x}(t+1), \dots]\end{aligned}\quad (31)$$

$$\begin{aligned}\mathcal{S}_{t,w} &\triangleq \text{span}\{\mathbf{s}_t, \mathbf{s}_{t-1}, \dots, \mathbf{s}_{t-w+1}\} \\ \mathcal{X}_{t,w} &\triangleq \text{span}\{\mathbf{x}_t, \mathbf{x}_{t-1}, \dots, \mathbf{x}_{t-w+1}\}\end{aligned}\quad (32)$$

where  $\mathcal{S}_{t,w}$  and  $\mathcal{X}_{t,w}$  are spaces spanned by the past input and (noiseless) observation vectors. It can be shown that when the channel is identifiable, there exists a  $w_o$  such that

$$\mathcal{S}_{t,w} = \mathcal{X}_{t,w}, \quad \forall w > w_o. \quad (33)$$

This implies that the input and observation spaces are identical. We now change the problem of identifying channel using the input subspaces.

For simplicity, consider the case for  $L = 2$ . We have

$$\begin{aligned}\mathbf{X}_t &= \mathbf{h}_0 \mathbf{s}_t + \mathbf{h}_1 \mathbf{s}_{t-1} + \mathbf{h}_2 \mathbf{s}_{t-2} \\ \mathbf{X}_{t+1} &= \mathbf{h}_0 \mathbf{s}_{t+1} + \mathbf{h}_1 \mathbf{s}_t + \mathbf{h}_2 \mathbf{s}_{t-1} \\ \mathbf{X}_{t+2} &= \mathbf{h}_0 \mathbf{s}_{t+2} + \mathbf{h}_1 \mathbf{s}_{t+1} + \mathbf{h}_2 \mathbf{s}_t.\end{aligned}$$

We define a projection space  $\mathcal{Z}$  that satisfies the following two conditions: 1)  $\{\mathbf{s}_{t+2}, \mathbf{s}_{t+1}, \mathbf{s}_{t-1}, \mathbf{s}_{t-2}\} \subset \mathcal{Z}$  and 2)  $\mathbf{s}_t \notin \mathcal{Z}$ . Because of the isomorphic relation between the input and observation spaces, it can be shown that

$$\mathcal{Z} = \mathcal{X}_{t,w} \cup \mathbf{X}_{t+L,-w} \quad (34)$$

which is the space spanned by the past and future observations. The smoothing error  $\tilde{\mathbf{X}}_{t+i|\mathcal{Z}}$  of  $\mathbf{X}_t$  has the following form:

$$\mathbf{E} \triangleq \begin{pmatrix} \tilde{\mathbf{X}}_{t|\mathcal{Z}} \\ \tilde{\mathbf{X}}_{t+1|\mathcal{Z}} \\ \tilde{\mathbf{X}}_{t+2|\mathcal{Z}} \end{pmatrix} = \begin{pmatrix} \mathbf{h}_0 \\ \mathbf{h}_1 \\ \mathbf{h}_2 \end{pmatrix} \tilde{\mathbf{s}}_{t|\mathcal{Z}} = \vec{\mathbf{h}} \tilde{\mathbf{s}}_{t|\mathcal{Z}}. \quad (35)$$

The channel vector can then be obtained from the projection error matrix  $\mathbf{E}$ . A general formulation that does not require the knowledge of the channel order is given in [121] and [123].

*a) Algorithm characteristics:* This approach has two attractive features. First, it converts a channel estimation problem to a linear LSS problem for which there are efficient adaptive implementations [141], [142] using lattice filters. Second, a joint order detection and channel algorithm [121], [123] (J-LSS) can be derived that determines the best channel order and channel coefficients to minimize the smoothing error. J-LSS is perhaps the only deterministic approach that enables channel identification with only the knowledge of the upper bound of the channel order.

## B. Second-Order Statistical Subspace Methods

*1) Assumptions and Identifiability:* In statistical subspace approaches, it is assumed that the source  $s_k$  is a random sequence with known second-order statistics. Although algorithms discussed here can be extended in many different ways, we shall assume the following assumptions in our discussion.

Assumption 2:

- 2.1) The source sequence is zero mean, white, with unit variance.
- 2.2) The noise sequence  $\mathbf{n}_k$ , uncorrelated with  $s_k$ , is zero mean, white, with known covariance  $\sigma^2 \mathbf{I}$ .
- 2.3) The channel order  $L$  is known.

Most algorithms can be extended to cases when the noise is colored but with known correlations. Some of the statistical methods do not require knowledge of the channel order. They require, instead, the upper bound of the channel order.

One of the most important questions is channel identifiability, i.e., given the second-order statistics of  $\mathbf{y}_k$ , can  $\mathbf{h}(z)$  be uniquely determined up to a constant factor? The answer to this question is affirmative provided that the subchannels are coprime.

To illuminate this issue, we present a frequency-domain argument given in [117], but in a slightly different form. Instead of using the multirate channel model we consider the multichannel model where the second-order statistics of the noiseless received signal  $\mathbf{x}_n$  are given by

$$\begin{aligned} r_{ij}(k) &\triangleq E(x_n^{(i)}(x_{n-k}^{(j)})^*) \\ r_{ij}(z) &\triangleq \sum_k r_{ij}(k)z^{-k}. \end{aligned} \quad (36)$$

It can be verified that  $r_{ij}(z)$  is related to the channel by

$$r_{ij}(z) = h_i(z)h_j^*\left(\frac{1}{z^*}\right) \quad \forall i, j. \quad (37)$$

We therefore have

$$\frac{r_{1j}(z)}{r_{jj}(z)} = \frac{h_1(z)}{h_j(z)} \quad \forall j. \quad (38)$$

Hence, those zeros of  $h_1(z)$  not shared by  $h_j(z)$  can be identified from the zeros of  $r_{1j}(z)/r_{jj}(z)$ . Thus, all zeros of  $h_1(z)$ , and  $h_1(z)$  itself, can be identified from the zeros of  $r_{1j}(z)/r_{jj}(z)$  if all channels do not share a common zero. Conversely, if all channels share a common zero  $z_0$ , then one can replace  $z_0$  with  $1/z_0^*$  without affecting  $r_{ij}(z)$  for all  $i$  and  $j$ . Thus the channel is not identifiable. We then have the following necessary and sufficient condition for channel identifiability.

*Theorem 3 [117], [119]:* Under Assumption 2, the channel  $\mathbf{h}(z)$  (or  $\vec{\mathbf{h}}$ ) can be uniquely identified up to a constant factor from the autocorrelation function  $\mathbf{R}_y(k) \triangleq E(\mathbf{y}_n \mathbf{y}_{n-k}^H)$  of the multichannel model [or, equivalently, the autocorrelation function  $r_y(n, k) \triangleq E(y_n y_{n-k}^*)$  in the multirate channel model] if and only if subchannels are coprime.

2) *Identification via Cyclic Spectra:* Having shown that the channel is uniquely determined from the second-order statistics, the next question is how to estimate the channel. Indeed, the arguments leading to the identifiability condition already suggest that channels can be identified from the zeros of the output spectra. Unfortunately, finding zeros of the estimated spectra accurately is difficult because

locations of the zeros can vary significantly with small variations of the estimated autocorrelation. To circumvent such difficulties, a subspace approach was first proposed in [117] and later independently in [41].

We present next an approach based on the multirate channel model. Although an equivalent method can be obtained for the multichannel model, the multirate model exploits the cyclostationary properties of the signal, which are also used in more recent approaches [15], [104] to solve blind channel estimation problems for cases that do not have a clear multichannel representation.

For the multirate channel model, it can be shown that the observation process  $y_k$  is cyclostationary, i.e.,  $r_y(n, k) \triangleq E(y_n y_{n-k}^*)$  is periodic in  $n$  with period  $P$ . Let  $S_y(n, z)$  be the "instantaneous spectrum"

$$S_y(n, z) = \sum_k r_y(n, k)z^{-k}. \quad (39)$$

Since  $S_y(n, z)$  is periodic in  $n$ , it has a Fourier series representation  $\{S^{(k)}(z)\}$  where  $S^{(k)}(z)$  is referred to as the  $k$ th cyclic spectrum. It is easy to show [117] that

$$S_y^{(k)}(z) = h(z)h^*\left[\frac{1}{z^*} \exp\left(j\frac{2\pi}{P}\right)\right] + \delta(k)\sigma^2. \quad (40)$$

Since  $\sigma^2$  is assumed to be known (or can be estimated in practice), we are dealing with the following identification equation in the frequency domain

$$S_y^{(k)}(z) - \sigma^2\delta(k) = h(z)h^*\left[\frac{1}{z^*} \exp\left(jk\frac{2\pi}{P}\right)\right] \quad (41)$$

where  $\delta(k)$  is the unit impulse, and the left-hand side of the above equation is known. A line of arguments identical to multichannel case can be used to show that the channel is identifiable if and only if the multirate channel  $h(z)$  does not have uniformly  $2\pi/P$ -spaced zeros, which is equivalent to the channel diversity condition, i.e., all subchannels in the vector channel model are coprime. To obtain channel identification, observe that, for any  $k_1$  and  $k_2$

$$\begin{aligned} S_x^{(k_1)}(z)h^*\left[\frac{1}{z^*} \exp\left(jk_2\frac{2\pi}{P}\right)\right] \\ - S_x^{(k_2)}(z)h^*\left[\frac{1}{z^*} \exp\left(jk_1\frac{2\pi}{P}\right)\right] = 0. \end{aligned} \quad (42)$$

It is clear that the time domain equivalence of the above identification equation (from the inverse  $Z$  transform of the above) leads to a set of linear equations with respect to the channel coefficients  $h_k$

$$\mathbf{G}(k_1, k_2)\vec{\mathbf{h}} = \mathbf{0} \quad (43)$$

where  $\mathbf{G}(k_1, k_2)$  can be constructed from the cyclic correlations of the received signal. The specific forms of  $\mathbf{G}(k_1, k_2)$  can be found in [117]. Therefore,  $\vec{\mathbf{h}}$  is in the null space of matrix  $\mathbf{G}(k_1, k_2)$ . Combining cyclic spectra for all  $k_1 \neq k_2$ , the intersection of the null spaces of all  $\mathbf{G}(k_1, k_2)$  gives the unique one-dimensional subspace to which the channel vector belongs.

A practical estimation algorithm can be derived from the following optimization:

$$\hat{H}(z) = \arg \min_{h(z)} \sum_{k_1 \neq k_2} \left\| \hat{S}_x^{(k_1)}(z) h^* \left[ \exp \left( j k_2 \frac{2\pi}{T} \right) \frac{1}{z^*} \right] - \hat{S}_x^{(k_2)}(z) h^* \left[ \exp \left( j k_1 \frac{2\pi}{T} \right) \frac{1}{z^*} \right] \right\|^2 \quad (44)$$

where the channel estimate  $\hat{h}$  is given by the quadratic optimization

$$\hat{h} = \arg \min_{\|\mathbf{u}\|_2=1} \sum_{k_1 \neq k_2} \|\hat{\mathbf{G}}(k_1, k_2) \mathbf{u}\|_2^2 \quad (45)$$

where  $\hat{\mathbf{G}}(k_1, k_2)$  is the estimated  $\mathbf{G}(k_1, k_2)$ .

a) *Algorithm characteristics and related work:* This algorithm exploits the complete cyclic statistics of the received and source signals, as well as the FIR structure of the channel model. The disadvantage of this algorithm is that it requires the convergence of source statistics, which means that, even when there is no noise, there is estimation error for any fixed sample size, although the algorithm is mean square consistent.

In related work, Li and Ding [69] developed a frequency domain nonparametric approach that identifies the magnitude and phase response separately from the cyclostationary statistics. Aghajan *et al.* [4] obtained its extension for multiuser scenarios. A similar approach is also presented in [14].

3) *Identification via Filtering Transform:* The first second-order statistical approach to blind channel estimation was proposed in [118], [119]. In this approach, the authors presented a two-step closed-form identification algorithm. The algorithm finds first the filtering matrix  $\mathcal{F}(\vec{\mathbf{h}})$  and then estimates the channel from the estimated filtering matrix.

Considering the time domain channel (2), we have

$$\mathbf{R}_0 \triangleq E(\vec{\mathbf{y}}(k) \vec{\mathbf{y}}^H(k)) = \mathbf{H} \mathbf{H}^H + \sigma^2 \mathbf{I} \quad (46)$$

$$\mathbf{R}_1 \triangleq E(\vec{\mathbf{y}}(k) \vec{\mathbf{y}}^H(k-1)) = \mathbf{H} \mathbf{J} \mathbf{H}^H \quad (47)$$

where  $\mathbf{H}$  is the filtering transform  $\mathcal{F}(\vec{\mathbf{h}})$ , and  $\mathbf{J}$  is the “shifting” matrix with the first lower off-diagonal entries being one and zero elsewhere. It was then shown that  $\mathbf{H}$  can be computed from  $\mathbf{R}_0$  and  $\mathbf{R}_1$ .

a) *Algorithm characteristics and related work:* The implementation of this algorithm requires the channel order and the noise variance, both of which, in principle, can be estimated from the SVD of the estimated covariance matrix  $\mathbf{R}_0$ . While it is consistent, this approach may not perform well for two reasons. First, the algorithm fails to take advantage of the special structure of the filtering transform  $\mathcal{F}(\vec{\mathbf{h}})$ . Second, the performance of such a two-step procedure is often affected by the quality of the estimation in the first step. On the other hand, since no structure in  $\mathcal{F}(\vec{\mathbf{h}})$  is assumed, when a large number of channels is available, this algorithm can be applied to  $\mathbf{y}_k$  [instead of  $\vec{\mathbf{y}}(k)$ ] directly, which may have computational advantages. An extension of this approach to colored source

was presented in [58] and its performance analysis was performed in [92]. The extension to the multiuser case was given in [70].

4) *Identification via Linear Prediction:* Introduced first by Slock [108]–[110], the linear prediction formulation of the multichannel problem plays an important role in the development of several algorithms. We present next one such approach by Abed-Meraim *et al.* [3].

Consider the multichannel model given in (2). The key idea comes from the recognition that the multichannel MA process is also autoregressive. Under the condition that subchannels are coprime there exists a causal  $\mathbf{g}(z)$  such that

$$\mathbf{g}^t(z) \mathbf{h}(z) = 1 \quad \rightarrow \quad s_k = [\mathbf{g}^t(z)] \mathbf{x}_k. \quad (48)$$

Substituting  $s_k = [\mathbf{g}^t(z)] \mathbf{x}_k$  above into the multichannel model, we have

$$\begin{aligned} \mathbf{x}_k &= \underbrace{\mathbf{h}_0 s_k}_{\mathbf{i}_k} + \underbrace{\mathbf{h}_1 ([\mathbf{g}^t(z)] \mathbf{x}_{k-1}) + \dots + \mathbf{h}_L ([\mathbf{g}^t(z)] \mathbf{x}_{k-L})}_{\hat{\mathbf{x}}_k} \\ &= \mathbf{i}_k + \hat{\mathbf{x}}_k. \end{aligned} \quad (49)$$

Since  $\{s_k\}$  is a white sequence,  $s_k$  is orthogonal to all  $\mathbf{x}_{k-j}$  for  $j > 0$ . Since  $\mathbf{g}(z)$  is causal,  $\mathbf{x}_k$  is the summation of the optimal linear prediction  $\hat{\mathbf{x}}_k$  and the innovation (the prediction error) process  $\mathbf{i}_k$ .

The identification of the channel involves two steps.

1) *Identification of  $\mathbf{h}_0$ :* The prediction error covariance is thus given by the covariance of the innovation process, i.e.,

$$\text{Cov}(\mathbf{x}_k - \hat{\mathbf{x}}_k) = \mathbf{h}_0 \mathbf{h}_0^H. \quad (50)$$

Given the autocorrelation function  $\mathbf{R}_x(n)$  of  $\mathbf{x}_k$ , the left-hand side can be computed explicitly using the standard theory in linear prediction. The right-hand side is a rank one matrix made of the vector of the first coefficient of the channel impulse response. Therefore, we can obtain  $\hat{\mathbf{h}}_0 = \mathbf{h}_0 e^{j\theta}$  for some unknown  $\theta$  from the eigenvector of the prediction error covariance associated with the largest eigenvalue.

2) *Identification of  $\mathbf{h}_k$ :* Once  $\hat{\mathbf{h}}_0$  is obtained, from (49), the input sequence can be constructed directly from the innovation sequence  $\mathbf{i}_k = \mathbf{x}_k - \hat{\mathbf{x}}_k$

$$\hat{s}_k = \frac{1}{\|\hat{\mathbf{h}}_0\|^2} \hat{\mathbf{h}}_0^H \mathbf{i}_k = s_k e^{-j\theta}. \quad (51)$$

With the estimated input, we essentially have the standard input–output channel identification problem.

a) *Algorithm characteristics and related work:* The algorithm uses all second-order statistics of the received signal, and it is mean square consistent. It does not require the exact channel order  $L$ , thus it is robust against overdetermination of the channel order. Derived from the noiseless model, the linear prediction idea is no longer valid in the presence of noise. However, when channel parameters are estimated from the autocorrelation functions,

the effect of noise can be lessened by subtracting the terms related to the noise correlation. The main disadvantage of this algorithm is that it is a two-step approach whose performance depends on the accuracy of the estimated  $\mathbf{h}_0$ . When noise presents and  $\|\mathbf{h}_0\|$  is small, performance degradation may be significant. Like all statistical moment methods, the convergence of the source statistics is also required.

The linear prediction-based approaches appear to be rooted in a somewhat surprising result by Liu and Dong [75]. It is shown that, for the multichannel model, a whitening filter is in fact a perfect equalizer, which is not true in the single channel case. Specifically, a finite order  $\mathbf{g}(z)$  is a whitening filter, i.e.,  $v_k = [\mathbf{g}^t(z)]\mathbf{x}_k$  is a white sequence if and only if  $\mathbf{g}^t(z)\mathbf{h}(z) = z^d$  for some  $d$ . In the spectral domain this result is a consequence of the maximum modulus theorem [100, p. 212].

A number of new approaches have been proposed recently. Based on the linear prediction framework, Gorokhov *et al.* proposed a weighted least squares approach in [48]. Ding [20] proposed the outer-product decomposition algorithm (OPDA) that obtains the channel  $\vec{\mathbf{h}}$  directly, hence avoiding the problem of small  $\|\mathbf{h}_0\|$ . Although OPDA was not derived from the linear prediction view point, it has the same identification equation as the multistep linear prediction approach derived by Gesbert and Duhamel [33].

### C. Other Related Subspace Approaches

Space limits our exposition of many channel estimation approaches developed recently. Here we mention two related classes of approaches that can be applied to general subspace methods for improved performance.

1) *Weighted Subspace Approaches*: Subspace approaches usually involve estimating the channel vector (or perhaps part of the channel vector) by optimizing a quadratic cost function

$$\hat{\mathbf{h}} = \arg \min_{\vec{\mathbf{h}} \in \mathcal{S}} \|\mathbf{P}\vec{\mathbf{h}}\|_2^2 \quad (52)$$

where  $\mathbf{P}$  is obtained from the received data. The weighted subspace approaches, successfully used in the direction of arrival estimation in array signal processing (see [128]), employ an additional weighting matrix  $\mathbf{W}$  which is chosen optimally in some ways

$$\hat{\mathbf{h}} = \arg \min_{\vec{\mathbf{h}} \in \mathcal{S}} \vec{\mathbf{h}}^H \mathbf{P}^H \mathbf{W} \mathbf{P} \vec{\mathbf{h}}. \quad (53)$$

The optimal selection of the weighting matrix is, however, nontrivial, and it is often a function of the true channel parameters. A practical solution is to use a consistent estimate of the channel to construct the optimal weighting matrix (see [1], [13], [48], and [66]).

2) *Exploiting Signal Waveforms*: Exploiting side information proves to be an effective way of circumventing the difficulties associated with the ill-conditioning of the channel matrix. Recognizing that in many communication applications the waveforms used in the transmission is often known, Schell *et al.* first proposed a subchannel response

matching approach [102], [103]. Principle component structure of the channel was used in [139]. Ding and Mao presented a knowledge-based approach [21]. In the multiple antenna array setting, while Gunther and Swindlehurst [49] developed an ESPRIT-like subspace approach by parameterizing the channel using physical parameters such as relative delays of multipaths. Applications to IS-136 are reported in [143].

## IV. OPTIMAL MOMENT METHODS: PERFORMANCE AND MATCHING TECHNIQUES

When the source has a statistical model, most subspace methods are part of the moment methods. Specifically, they all can be viewed as estimating channel parameters from the estimated second-order moments  $\hat{\mathbf{r}}$  of the received signal. For the class of consistent estimators, asymptotic normalized mean square error (ANMSE) can be used as a performance measure. Specifically, given the estimated second-order moments  $\hat{\mathbf{r}}(N_s)$  from  $N_s$  observations, the ANMSE of the estimator  $g(\hat{\mathbf{r}}(N_s)) = \hat{\vec{\mathbf{h}}}(N_s)$  is defined by

$$\text{ANMSE}(g) = \lim_{N_s \rightarrow \infty} N_s E \left\{ \|\hat{\vec{\mathbf{h}}}(N_s) - \vec{\mathbf{h}}\|^2 \right\} \quad (54)$$

when the limit exists. By normalization we mean that both the channel and its estimate are normalized to unit 2-norm. Further, to obtain a meaningful MSE, we also assume that the scalar ambiguity of the estimate has been removed. ANMSE( $g$ ) measures the MSE of the consistent estimator  $g$  for a sufficiently large sample size  $N_s$

$$E \left\{ \|\hat{\vec{\mathbf{h}}}(N_s) - \vec{\mathbf{h}}\|^2 \right\} \approx \frac{\text{ANMSE}(g)}{N_s}.$$

Obviously, smaller ANMSE( $g$ ) is desired. When ANMSE( $g$ )  $\rightarrow \infty$ , it implies that the estimator  $g$  does not have convergence at the rate of  $1/N_s$ .

In analyzing the class of blind channel estimators using the second-order moments, we pose the following questions.

- 1) What is the achievable ANMSE among all consistent estimators using consistent estimates of second-order moments?
- 2) What are fundamental limitations to the ANMSE of blind channel estimators using the second-order statistics?
- 3) What is the ANMSE of existing subspace estimators and what are their performance limitations?
- 4) How much potential improvement can be made over the existing subspace based moment estimators?

These questions are addressed in part in [1], [2], [43], [44], [93], [137], [139], and [140].

### A. The Achievable ANMSE and Performance Bounds

The question considered here is the following: given consistent estimates  $\hat{\mathbf{r}}$  of the second-order moments of the observation, what is the minimum ANMSE that can be achieved by an estimator using  $\hat{\mathbf{r}}$ ? The answer to

this problem can be obtained by applying the asymptotic performance analysis of general moment methods [91]. For the case involving real signals Zeng and Tong gave the following theorem. The complex case can be found in [44] and [136].

*Theorem 4 [140]:* Let  $\mathbf{r}$  be the vector consisting of (nonredundant) autocorrelation coefficients. Assume that  $\mathbf{S}(h\vec{b}f)$ , the Jacobian of the autocorrelation vector  $\mathbf{r}$  with respect to the channel vector  $\vec{\mathbf{h}}$ , is full column rank. Let  $\hat{\mathbf{r}}$  be the estimated autocorrelation vector obtained from  $\{\mathbf{y}(i), i = 0, \dots, N_s - 1\}$  with normalized asymptotic covariance  $\Sigma(\vec{\mathbf{h}}) > 0$ . Let  $\hat{\vec{\mathbf{h}}} = g(\hat{\mathbf{r}})$  be a channel parameter estimator such that  $g(\mathbf{r}) = \vec{\mathbf{h}}$ . Then the ANMSE of  $g(\hat{\mathbf{r}})$  is lower bounded by

$$\begin{aligned} \text{ANMSE} &\geq \text{tr} \left\{ \left( \mathbf{S}^T(\vec{\mathbf{h}}) \Sigma^{-1}(\vec{\mathbf{h}}) \mathbf{S}(\vec{\mathbf{h}}) \right)^{-1} \right\} \\ &\geq K \left( \frac{\kappa(\mathbf{S}(\vec{\mathbf{h}}))}{\text{SNR}} \right)^2 \end{aligned} \quad (55)$$

where  $K$  is a constant,  $\kappa(\mathbf{S}(\vec{\mathbf{h}}))$  is the condition number of  $\mathbf{S}(\vec{\mathbf{h}})$ , and the SNR is defined as

$$\text{SNR} \triangleq \frac{1}{P\sigma^2} E \left\{ \sum_{j=1}^P |x_i^{(j)}|^2 \right\} = \frac{1}{P\sigma^2}. \quad (56)$$

Moreover, there exists an estimator  $g_{opt}(\hat{\mathbf{r}})$  that achieves the lower bound  $\text{tr} \{ (\mathbf{S}^T(\vec{\mathbf{h}}) \Sigma^{-1}(\vec{\mathbf{h}}) \mathbf{S}(\vec{\mathbf{h}}))^{-1} \}$ .

From (55), it is clear that the performance of all moment methods are limited by the condition number of the Jacobian  $\mathbf{S}(\vec{\mathbf{h}})$ , which leads to the following question: when is  $\mathbf{S}(\vec{\mathbf{h}})$  singular? This question has a surprisingly simple condition.

*Lemma 1 [140]:*  $\mathbf{S}(\vec{\mathbf{h}})$  is singular if and only if  $\{h_i(z)\}$  share common conjugate reciprocal zeros (CRZ), or equivalently,  $\{h^{(i)}(z), h^{(i)}(1/z^*)\}_{i=1}^M$  share common zeros.

The above condition shows an interesting difference from the condition of identifiability (subchannels are coprime). Note that the violation of the identifiability condition does not imply that no moment algorithm can achieve the ANMSE bound. When subchannels do have common zeros, there are multiple but possibly finite numbers of possible solutions to the identification equation. If one can restrict the parameter set to the neighborhood of the true channel, optimal algorithms with minimum ANMSE do exist [44]. This, of course, is not unique to the multichannel identification.

It is also interesting to compare the performance bound for the CR and the noise subspace method. For the special case when  $P = 2$ , the ANMSE of both CR and the noise subspace methods can be obtained easily if the covariance matrix  $\mathbf{R}_y$  has the Wishart distribution. Under this assumption, it can be shown that

$$\text{ANMSE}_{\text{CR}} = \sum_{k=1}^{2L+1} \frac{\sigma^2(\lambda_k^2 + \sigma^2)}{(\lambda_k^2)^2} \geq K_{\text{CR}} \frac{\kappa^2(\mathcal{F}(\vec{\mathbf{h}}))}{\text{SNR}} \quad (57)$$

where  $\lambda_{\max} = \lambda_1 > \lambda_2 > \dots > \lambda_{2L+1} = \lambda_{\min}$  are the singular values of the  $\mathcal{F}(\vec{\mathbf{h}})$ ,  $\kappa(\mathcal{H}(\vec{\mathbf{h}}))$  is the condition number of  $\mathcal{F}(\vec{\mathbf{h}})$  and  $K_{\text{CR}}$  is a constant. If the source is Gaussian, Abed-Meraim *et al.* obtained a different bound [2].

The above bound shows that the CR and noise subspace methods are limited by the condition number of the channel matrix  $\mathcal{F}(\vec{\mathbf{h}})$  or the locations of channel zeros. Indeed, subspace methods often suffer from the ill-conditioning of the matrix from which they are derived. For example, certain channels have closely located zeros, which causes the ill-conditioning of the channel matrix. This effect was illustrated in Endres *et al.* [25].

### B. Moment Matching Techniques

The moment matching approach is motivated by the existence of a moment method that achieves the minimum ANMSE. Giannakis and Halford investigated the general moment matching approach of the following form:

$$\hat{\mathbf{h}}_{opt} = \arg \min \|\hat{\mathbf{r}} - \mathbf{r}(\vec{\mathbf{h}})\|_{\mathbf{W}}^2 \quad (58)$$

where  $\mathbf{W}$  is a weighting matrix. By choosing appropriate  $\mathbf{W}$ , as a function of  $\vec{\mathbf{h}}$ , the so-called asymptotic best consistent (ABC) estimator achieving the minimum ANMSE was proposed in [43] and [44]. The suboptimal approach with no weighting was investigated in [120].

While moment matching methods have a much more robust performance against channel order selection and the channel condition, they are unfortunately not easy to implement because of the existence of local minima in the optimization. To incorporate the subspace structure into the moment matching approach, Zeng and Tong proposed in [139] the following channel estimation criterion:

$$\hat{\vec{\mathbf{h}}} = \arg \min_{\vec{\mathbf{h}} \in \mathcal{S}, \|\vec{\mathbf{h}}\|=1} \|\hat{\mathbf{r}} - \mathbf{r}(\vec{\mathbf{h}})\|_{\mathbf{W}}^2 \quad (59)$$

where  $\mathcal{S}$  is a linear subspace containing that used in the subspace algorithms. The selection of  $\mathcal{S}$  leads to a method that combines both subspace and moment matches.

## V. THE ML METHODS

One of the most popular parameter estimation algorithms is the ML method. Not only can such methods be derived in a systematic way, but perhaps more importantly, the class of maximum likelihood estimators are usually optimal for large data records as they approximate the minimum variance unbiased estimators. Asymptotically, under certain regularity conditions, the variance of ML estimators approach the Cramér–Rao bound (CRB), which is the lower bound on variance for all unbiased estimators. Unfortunately, unlike subspace based approaches, the ML methods usually cannot be obtained in closed form. Their implementations are further complicated by the existence of local minima. However, ML approaches can be made very effective by including the subspace and other suboptimal approaches as initialization procedures. The general

formulation of the ML estimation can be found in many textbooks (e.g., see [91]).

The problem at hand is to estimate the deterministic (vector) parameter  $\theta$  given the probabilistic model of the observation. Specifically, let  $f(y; \theta)$  be the probability density function of random variable  $Y$  parameterized by  $\theta \in \Theta$ . Given an observation  $Y = y$ ,  $\theta$  is estimated by maximizing

$$\hat{\theta} = \arg \max_{\theta \in \Theta} f(y; \theta) \quad (60)$$

where  $f(y; \theta)$ , when viewed as the function of  $\theta$ , is referred to as the likelihood function.

The ML-based blind channel estimation can be derived based on either the statistical or the deterministic setting depending on the model of the source signal.

**SML** *Statistical ML estimation:* In such a case, the input sequence is assumed to be random with a known distribution. In such a formulation, the only unknown parameter is the channel vector ( $\theta = \vec{\mathbf{h}}$ ). In this case, the dimension of the unknown parameter is fixed with respect to the data size.

**DML** *Deterministic ML estimation:* Here the input sequence is part of the unknown parameters, i.e.,  $\theta = (\vec{\mathbf{h}}, \{s_k\})$ , although one may only be interested in estimating  $\vec{\mathbf{h}}$ . In such a case, the dimension of the parameters increases with the size of the observation.

These two classes of ML estimators are discussed next.

#### A. DML Approach

The DML approach assumes no statistical model for the input sequence  $s_k$ . In other words, both the channel vector  $\vec{\mathbf{h}}$  and the input source vector  $\vec{\mathbf{s}}$  are parameters to be estimated. In this paper, we shall only consider the estimation of the channel.

Consider the multichannel model in (2)

$$\vec{\mathbf{y}} = \mathcal{T}(\vec{\mathbf{s}})\vec{\mathbf{h}} + \vec{\mathbf{n}}. \quad (61)$$

The DML problem can be stated as follows: given  $\vec{\mathbf{y}}$ , estimate  $\vec{\mathbf{h}}$  by

$$\{\hat{\vec{\mathbf{h}}}_{\text{DML}}, \hat{\vec{\mathbf{s}}}_{\text{DML}}\} = \arg \max f(\vec{\mathbf{y}}; \vec{\mathbf{h}}, \vec{\mathbf{s}}) \quad (62)$$

where  $f(\vec{\mathbf{y}}; \vec{\mathbf{h}}, \vec{\mathbf{s}})$  is the density function of the observation vectors  $\vec{\mathbf{y}}$  parameterized by both the channel  $\vec{\mathbf{h}}$  and the input source  $\vec{\mathbf{s}}$ .

When the noise is zero-mean Gaussian with covariance  $\sigma^2 \mathbf{I}$ , the ML estimates can be obtained by the nonlinear least squares optimization

$$\{\hat{\vec{\mathbf{h}}}_{\text{DML}}, \hat{\vec{\mathbf{s}}}_{\text{DML}}\} = \arg \min \|\vec{\mathbf{y}} - \mathcal{T}(\vec{\mathbf{s}})\vec{\mathbf{h}}\|_2^2. \quad (63)$$

1) *Assumptions and Identifiability:* In considering the deterministic model, we assume the following assumptions.

*Assumption 3:*

- 3.1) The noise  $\vec{\mathbf{n}}$  is zero mean, Gaussian, with known covariance  $\sigma^2 \mathbf{I}$ .
- 3.2) The channel has known order  $L$ .

We note that the noise variance can also be considered as part of the parameters. For simplicity and consistency with other approaches it is assumed to be known in our discussion. Note also that the set of assumptions for DML is almost the same as that for the deterministic subspace methods, except that the noise in DML is assumed to be Gaussian. Again, the channel model  $L$  must be known for identifiability reasons.

It is not surprising that the identifiability condition for DML is the same as that for the deterministic second-order moment methods. Specifically, the channel is identifiable if subchannels are coprime and the source has linear complexity greater than  $2L + 1$ . The reason is that, when the noise is Gaussian, all information about the channel in the likelihood function resides in the second-order moments of the observations. Readers are referred to Theorem 1 for sufficient conditions and related discussions.

2) *IQML, TSML, and Other Iterative Methods:* These algorithms are developed by Hua [55], [56] and around the same time by Slock [108]. The iterative quadratic maximum likelihood (IQML) approach, proposed by Bresler and Macovski [11] for estimating superimposed exponential signals, transforms the DML problem into a sequence of quadratic optimization problems for which simple solutions can be obtained. It turns out that IQML has a related form in blind channel estimation using DML. This connection, first pointed out by Slock and Papadias [108], [110], has its root in the linear prediction formulation in both problems.

The joint optimization of the likelihood function in both the channel and the source parameter spaces is difficult. Fortunately, the observation is linear in both the channel and the input parameters. In other words, we have a separable nonlinear LS problem, which allows us to reduce the complexity considerably. The nonlinear LS optimization can be achieved sequentially in one of the following ways:

$$\begin{aligned} \{\hat{\vec{\mathbf{h}}}_{\text{DML}}, \hat{\vec{\mathbf{s}}}_{\text{DML}}\} &= \arg \min_{\vec{\mathbf{h}}} \left\{ \min_{\vec{\mathbf{s}}} \|\vec{\mathbf{y}} - \mathcal{T}(\vec{\mathbf{s}})\vec{\mathbf{h}}\|_2^2 \right\} \quad (64) \\ &= \arg \min_{\vec{\mathbf{s}}} \left\{ \min_{\vec{\mathbf{h}}} \|\vec{\mathbf{y}} - \mathcal{T}(\vec{\mathbf{s}})\vec{\mathbf{h}}\|_2^2 \right\}. \quad (65) \end{aligned}$$

Considering next the optimization in (64), we have

$$\begin{aligned} \hat{\vec{\mathbf{h}}}_{\text{DML}} &= \arg \min_{\vec{\mathbf{h}}} \left\| \underbrace{\left( \mathbf{I} - \mathcal{F}(\vec{\mathbf{h}})\mathcal{F}^\dagger(\vec{\mathbf{h}}) \right)}_{\mathcal{P}(\vec{\mathbf{h}})} \vec{\mathbf{y}} \right\|^2 \\ &= \arg \min_{\vec{\mathbf{h}}} \|\mathcal{P}(\vec{\mathbf{h}})\vec{\mathbf{y}}\|^2 \quad (66) \end{aligned}$$

where  $\mathcal{P}(\vec{\mathbf{h}})$  is a projection transform of  $\vec{\mathbf{y}}$  into the orthogonal complement of the range space of  $\mathcal{F}(\vec{\mathbf{h}})$  or the noise subspace of the observation.

The key of IQML type of algorithms is the parameterization of  $\mathcal{P}(\vec{\mathbf{h}})$ . Hua in [56] obtained  $\mathcal{P}(\vec{\mathbf{h}})$  directly from

the channel vector  $\vec{\mathbf{h}}$ . Fig. 5 provided the clue for such a construction, where it is clear that the channel itself can be used to null the noiseless observation, a process called “blocking” by Slock. Hua’s construction of  $\mathcal{P}(\vec{\mathbf{h}})$  uses the data selection transform defined in (64) to obtain the IQML form

$$\hat{\mathbf{h}}_{\text{DML}} = \arg \min_{\vec{\mathbf{h}}} \vec{\mathbf{h}}^H \mathbf{Y}^H \mathbf{Q}(\vec{\mathbf{h}}) \mathbf{Y} \vec{\mathbf{h}} \quad (67)$$

where  $\mathbf{Q}(\vec{\mathbf{h}})$  can be obtained easily from  $\vec{\mathbf{h}}$ , and  $\mathbf{Y}$  is a matrix constructed from  $\vec{\mathbf{y}}$ . To implement the DML estimation, Hua proposed a two-step approach referred to as the two-step maximum likelihood (TSML) method that 1) uses the CR method to obtain an initial estimate of the channel and 2) substitute the initial estimate into  $\mathbf{Q}(\vec{\mathbf{h}})$  and optimize (67) recursively.

*a) Algorithm characteristics and related work:* A number of IQML type of approaches have been proposed depending on the parameterization of the projection  $\mathcal{P}(\vec{\mathbf{h}})$ . In Slock’s “minimum null-space parameterization” [110], IQML is applied to the blocking filter. A different approach was developed by Harikumar and Bresler [53]. This IQML type of algorithm (not surprisingly) offers more efficient channel estimates when compared with moment methods. Hua demonstrated that TSML is both “high SNR” consistent and efficient. Similarly, Harikumar and Bresler also showed that the CR method used in Hua’s TSML is a coarse approximation of IQML, which ultimately supports Hua’s TSML. The performance comparison with the Cramér–Rao bound has also been obtained in [53], [56], and [85]. As a “dual” to the IQML-type of algorithms, Feder and Catipovic [27] proposed a DML by obtaining first  $\hat{\mathbf{s}}_{\text{DML}}$  by optimizing first the inner term in (65). Since the estimation of the input is obtained first, it suffers from the fact that the dimension of the problem increases with the sample size, which renders this approach not practical for large data size applications. For cases when the input sequence has the finite alphabet property, simplifications can be obtained (see [27]).

*3) DML for Finite Alphabet Input:* Similar to SML with hidden Markov model (HMM), finite alphabet properties can also be incorporated into DML. Because of the finite alphabet property, it is difficult to apply the separation idea in IQML-type approach. Consequently, this class of algorithms, first proposed by Seshadri [105] and Ghosh and Weber [36], iterates between estimates of the channel and the input. At iteration  $k$ , with an initial guess of the channel  $\vec{\mathbf{h}}^{(k)}$ , the algorithm estimates the input sequence  $\vec{\mathbf{s}}^{(k)}$  and the channel  $\vec{\mathbf{h}}^{(k+1)}$  for the next iteration by

$$\vec{\mathbf{s}}^{(k)} = \arg \min_{\vec{\mathbf{s}} \in \mathcal{S}} \|\vec{\mathbf{y}} - \mathcal{F}(\vec{\mathbf{h}}^{(k)}) \vec{\mathbf{s}}\|^2 \quad (68)$$

$$\vec{\mathbf{h}}^{(k+1)} = \arg \min_{\vec{\mathbf{h}}} \|\vec{\mathbf{y}} - \mathcal{T}(\vec{\mathbf{s}}^{(k)}) \vec{\mathbf{h}}\|^2 \quad (69)$$

where  $\mathcal{S}$  is the (discrete) domain of  $\vec{\mathbf{s}}$ . The optimization in (69) is a linear least squares problem whereas the optimization in (68) can be achieved by using the Viterbi

algorithm [28]. The convergence of such approaches are not guaranteed in general.

*a) Algorithm characteristics and related work:* The finite alphabet nature of the input makes the evaluation of the Cramér–Rao lower bound difficult. Paris argued in [86] that, if the input sequence is equally probable, the probability that the above estimate  $\hat{\vec{\mathbf{s}}}$  differs from the ML estimate of  $\vec{\mathbf{s}}$  with known  $\vec{\mathbf{h}}$  diminishes with the noise variance. Similarly, at high SNR, one can expect that the above channel estimate is close to the ML channel estimate with known input. There are many variations in the implementation of the nonlinear LS to reduce the implementation complexity. Seshadri presented “blind” trellis search techniques [106]. Reduced-state sequence estimation [26] was proposed in [36]. The so-called iterative LS with projection (ILSP) proposed by Talwar *et al.* [111], [112] is a relaxation technique that first ignores the finite alphabet property and then projects the estimate to its nearest discrete value. Raheli *et al.* proposed a per-survivor processing technique in [95]. An algebraic approach was presented by Yellin and Porat [135].

## B. Statistical Maximum Likelihood Approach

We consider the statistical model where the source sequence  $s_k$  is random. The formulation of the problem is straightforward in principle. Recalling the multichannel model (2) where we consider a block of  $N$  received vectors  $\vec{\mathbf{y}} \triangleq [\mathbf{y}_{N-1}^t, \dots, \mathbf{y}_0^t]^t$

$$\vec{\mathbf{y}} = \mathcal{F}_N(\vec{\mathbf{h}}) \vec{\mathbf{s}} + \vec{\mathbf{n}} \quad (70)$$

where we have omitted the time index because  $\vec{\mathbf{y}}$  has included all observations. The SML problem can be stated as follows: given  $\vec{\mathbf{y}}$ , estimate  $\vec{\mathbf{h}}$  by

$$\hat{\mathbf{h}}_{\text{SML}} = \arg \max_{\vec{\mathbf{h}}} f(\vec{\mathbf{y}}; \vec{\mathbf{h}}) = \arg \max_{\vec{\mathbf{h}}} \log f(\vec{\mathbf{y}}; \vec{\mathbf{h}}) \quad (71)$$

where  $f(\vec{\mathbf{y}}; \vec{\mathbf{h}})$  is the density function of the observation vectors  $\vec{\mathbf{y}}$  parameterized by  $\vec{\mathbf{h}}$ .

*1) Assumptions and Identifiability:* The SML estimation hinges on the availability and the evaluation of the likelihood function. Although SML applies to more general cases, we shall make the following assumptions in our discussion.

*Assumption 4:*

- 4.1) components of  $\vec{\mathbf{s}}$  and  $\vec{\mathbf{n}}$  are jointly independent;
- 4.2)  $\vec{\mathbf{n}}$  is zero mean Gaussian with covariance  $\sigma^2 \mathbf{I}$ ;
- 4.3) components of  $\vec{\mathbf{s}}$  are independently, identically distributed (i.i.d.) with known probability density function.

Identifiability remains to be an important issue in SML approach. The identifiability condition tells when SML can be applied. A main issue is whether the likelihood function provides sufficient information to distinguish different models. Specifically,  $\vec{\mathbf{h}}$  is identifiable if  $f(\vec{\mathbf{y}}; \vec{\mathbf{h}}) = f(\vec{\mathbf{y}}; \vec{\mathbf{h}}')$  (almost everywhere) implies  $\vec{\mathbf{h}} = e^{j\theta} \vec{\mathbf{h}}'$  for some  $\theta$ . It is not

surprising to see that the class of channels identifiable by SML is larger than that by moments. Obviously, parameters identifiable by moments are identifiable by the likelihood function. It can be shown further that as long as  $\vec{s}$  is non-Gaussian, the linear structure in (70) is uniquely determined by the density function  $f(\mathbf{y}; \vec{\mathbf{h}})$  [62]. Indeed, under the non-Gaussian assumption of the source, subspace methods developed by Giannakis and Mendel [39] can be used to obtain closed-form identification based on higher order cumulants of the observation. When  $\vec{s}$  is Gaussian, all statistical information about the channel is contained in the second-order moment. In such a case, identifiability can be ensured if subchannels are coprime.

*Theorem 5:* Under Assumption 4, the channel parameter  $\vec{\mathbf{h}}$  is identifiable by the likelihood function if and only if one of the following conditions is satisfied:

- 1)  $\mathbf{s}$  is non-Gaussian;
- 2) subchannels are coprime.

2) *The EM Approach:* The SML optimization in (71) is in general difficult because  $f(\vec{\mathbf{y}}; \vec{\mathbf{h}})$  is nonconvex. The expectation-maximization (EM) algorithm [9], [19] can be applied to transform the complicated optimization (71) to a sequence of quadratic optimizations. We shall give an intuitive explanation of this idea. More rigorous development can be found in [19].

If the input is known, e.g.,  $\vec{s} = \xi$ , the ML estimation of  $\vec{\mathbf{h}}$  is a simple LS problem involving maximizing a quadratic cost

$$\log f(\vec{\mathbf{y}}|\vec{s} = \xi; \vec{\mathbf{h}}) = -C\|\vec{\mathbf{y}} - \mathcal{T}(\xi)\vec{\mathbf{h}}\|^2 \quad (72)$$

where  $C$  is a constant. When  $\vec{s}$  is unknown but with known distribution, we should consider maximizing the above cost averaged over all possible input sequences. Given the received signal  $\vec{\mathbf{y}}$ , this average should be performed using the *a posteriori* distribution of  $\vec{s}$

$$J(\vec{\mathbf{h}}) = E(\log f(\vec{\mathbf{y}}|\vec{s} = \xi; \vec{\mathbf{h}})|\vec{\mathbf{y}}). \quad (73)$$

Unfortunately, the computation of  $J(\vec{\mathbf{h}})$  requires the knowledge of the true channel parameter. To circumvent this difficulty, we may consider an approximation of the above cost function using, at iteration  $k$ , the current estimate  $\vec{\mathbf{h}}^{(k)}$

$$Q(\vec{\mathbf{h}}, \vec{\mathbf{h}}^{(k)}) \triangleq E(\log f(\vec{\mathbf{y}}|\vec{s} = \xi; \vec{\mathbf{h}})|\vec{\mathbf{y}}; \vec{\mathbf{h}}^{(k)}) \quad (74)$$

$$= - \int \|\vec{\mathbf{y}} - \mathcal{T}(\vec{s})\vec{\mathbf{h}}\|^2 f(\vec{s}|\vec{\mathbf{y}}; \vec{\mathbf{h}}^{(k)}) d\vec{s} \quad (75)$$

$$= -E(\|\vec{\mathbf{y}} - \mathcal{T}(\vec{s})\vec{\mathbf{h}}\|^2|\vec{\mathbf{y}}; \vec{\mathbf{h}}^{(k)}). \quad (76)$$

It is clear that the maximization of  $Q(\vec{\mathbf{h}}, \vec{\mathbf{h}}^{(k)})$  with respect to  $\vec{\mathbf{h}}$  is in fact a quadratic optimization problem with the new estimate  $\vec{\mathbf{h}}^{(k+1)}$  given by

$$\vec{\mathbf{h}}^{(k+1)} = \left[ E(\mathcal{T}^H(\vec{s})\mathcal{T}(\vec{s})|\vec{\mathbf{y}}; \vec{\mathbf{h}}^{(k)}) \right]^{-1} \cdot E(\mathcal{T}^H(\vec{s})\vec{\mathbf{y}}|\vec{\mathbf{y}}; \vec{\mathbf{h}}^{(k)}). \quad (77)$$

One recognizes that the only difference between the above solution and that when the input is known is the conditional expectation.

Although the above arguments are solid heuristics, it is not clear whether  $\vec{\mathbf{h}}^{(k)} \rightarrow \hat{\vec{\mathbf{h}}}_{\text{SML}}$ . This is indeed the case provided a good initial guess of the channel is available.

*Theorem 6 [19]:* If  $Q(\vec{\mathbf{h}}^{(k+1)}, \vec{\mathbf{h}}^{(k)}) \geq Q(\vec{\mathbf{h}}^{(k)}, \vec{\mathbf{h}}^{(k)})$ , then  $f(\vec{\mathbf{y}}; \vec{\mathbf{h}}^{(k+1)}) \geq f(\vec{\mathbf{y}}; \vec{\mathbf{h}}^{(k)})$ .

The above theorem implies that the maximization of the likelihood function can be achieved by a sequential maximization the ‘‘auxiliary function’’  $Q(\vec{\mathbf{h}}, \vec{\mathbf{h}}^{(k)})$ , which has a closed-form solution

E: Compute:

$$Q(\vec{\mathbf{h}}, \vec{\mathbf{h}}^{(k)}) = E(\|\vec{\mathbf{y}} - \mathcal{T}(\vec{s})\vec{\mathbf{h}}\|^2|\vec{\mathbf{y}}; \vec{\mathbf{h}}^{(k)}).$$

M: Maximization:

$$\vec{\mathbf{h}}^{(k+1)} = \left[ E(\mathcal{T}^H(\vec{s})\mathcal{T}(\vec{s})|\vec{\mathbf{y}}; \vec{\mathbf{h}}^{(k)}) \right]^{-1} \cdot E(\mathcal{T}^H(\vec{s})\vec{\mathbf{y}}|\vec{\mathbf{y}}; \vec{\mathbf{h}}^{(k)}).$$

a) *Algorithm characteristics and related work:* The performance of EM algorithm depends on its initialization, which may be facilitated by moment techniques such as those described in Section III (see [89]). When EM converges globally, the estimate achieves asymptotically the CRB for the case of i.i.d. sequences (which is not the case here). See [18] for the evaluation of CRB when the input is Gaussian.

Various algorithms are implemented either in ‘‘on-line’’ or batch modes. Kaleh and Vallet [64] first applied the EM algorithm to the equalization of communication channels with input sequence having finite alphabet property. By using an HMM they developed a batch (off-line) procedure that includes the so-called forward and backward recursions [94]. The complexity of this algorithm increases exponentially with the channel memory. Shao and Nikias [107] proposed an approximation for calculating the elements of the conditional autocorrelation matrix and correlation vector involved in (77). Such an approximation becomes exact as the block size approaches infinity. At high SNR, further reduction of complexity can be achieved as shown in [87].

To relax the memory requirements and facilitate channel tracking, ‘‘on-line’’ sequential approaches have been proposed in [113], [114], and [130] for general input, and in [65] for input with finite alphabet properties under an HMM formulation. Given the appropriate regularity conditions [113] and a good initialization guess, it can be shown that these algorithms converge (almost surely and in the mean-square sense) to the true channel value. To reduce the implementation complexity associated with the HMM formulation, suboptimal approaches have been proposed in [131] and [132]. The complexity of the implementation in [88], [132] increases linearly with the channel. When SNR approaches infinity, the suboptimal implementation achieves optimality.

## VI. CONCLUSION

In this paper, we have presented some recent developments in blind identification and estimation of single input and multiple output channels. Depending on the application, the problem of blind identification is a problem of exploiting structural information of the channel and properties of its input. Because different applications utilize different structures to specify the unknown parameters, there is great diversity in developing new approaches. Recently, there have been several developments in semiblind channel estimation techniques. Semiblind channel estimation stands for cases when part of the input is accessible. This case is significant in several ways. First, the availability of certain input should improve the performance of any blind channel estimators. de Carvalho and Slock evaluated the CRB for blind channel estimation for both deterministic and Gaussian input cases [17], [18]. Second, previously unidentifiable channel may become identifiable. If zeros are introduced periodically in the input data sequence, Giannakis [37] showed that any FIR channel can be identified, and a subspace algorithm was proposed in [52]. Pal proposed a semiblind channel estimation technique in [84].

While it is clear that these methods have potential applications in many different problems, it is still too early to assess their impact, for most studies have been conducted either in simulation or with real data but in a controlled manner.

## REFERENCES

- [1] K. Abed-Meraim, J. Cardoso, A. Gorokhov, P. Loubaton, and E. Moulines, "On subspace methods for blind identification of single-input multiple-output FIR systems," *IEEE Trans. Signal Processing*, vol. 45, pp. 42–55, Jan. 1997.
- [2] K. Abed-Meraim, P. Loubaton, and E. Moulines, "A subspace algorithm for certain blind identification problems," *IEEE Trans. Inform. Theory*, vol. 43, pp. 499–511, Mar. 1997.
- [3] K. Abed-Meraim, E. Moulines, and P. Loubaton, "Prediction error method for second-order blind identification," *IEEE Trans. Signal Processing*, vol. 45, pp. 694–705, Mar. 1997.
- [4] H. Aghajan, B. Hassibi, B. Khalaj, A. Paulraj, and T. Kailath, "Blind identification of FIR channels with multiple users via spatio-temporal processing," in *Proc. IEEE GLOBECOM*, San Francisco, CA, Nov. 1994, vol. 3, pp. 1899–1903.
- [5] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Automat. Contr.*, vol. AC-19, pp. 716–723, Dec. 1974.
- [6] S. Amari and A. Cichocki, "Adaptive blind signal processing—Neural network approaches," this issue, pp. 2026–2048.
- [7] L. A. Baccala and S. Roy, "A new blind time-domain channel identification method based on cyclostationarity," presented at 26th Conf. Information Sciences and Systems, Princeton, NJ, Mar. 1994.
- [8] —, "A new blind time-domain channel identification method based on cyclostationarity," *IEEE Signal Processing Lett.*, vol. 1, pp. 89–91, June 1994.
- [9] L. E. Baum, T. Petrie, G. Soules, and N. Weiss, "A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains," *Ann. Math. Stat.*, vol. 41, pp. 164–171, 1970.
- [10] R. E. Blahut, *Algebraic Methods for Signal Processing and Communications Coding*. New York: Springer-Verlag, 1992.
- [11] Y. Bresler and A. Macovski, "Exact maximum likelihood parameter estimation of superimposed exponential signals in noise," *IEEE Trans. Audio, Speech, Signal Processing*, vol. ASSP-34, pp. 1081–1089, Oct. 1986.
- [12] D. Brillinger, "The identification of polynomial systems by means of higher-order spectra," *J. Sound Vib.*, vol. 20, pp. 301–313, 1970.
- [13] J.-F. Cardoso, P. Loubaton, and E. Moulines, "On weighted subspace estimates in system identification," in *Proc. 8th IEEE Signal Processing Workshop Statistical and Array Signal Processing*, Corfu, Greece, June 1996, pp. 352–355.
- [14] A. Chevreuil and P. Loubaton, "On the use of conjugate cyclo-stationarity: A blind second-order multi-user equalization method," in *Proc. IEEE Intl. Conf. Acoustics, Speech, Signal Processing*, Atlanta, GA, May 1996, vol. 5, pp. 2439–2442.
- [15] —, "Blind second-order identification of FIR channels: Forced cyclo-stationarity and structured subspace method," in *Proc. 1st IEEE Signal Processing Workshop Signal Processing Advances in Wireless Communication*, Paris, France, Apr. 1997, vol. 1, pp. 121–124.
- [16] S. Choi and R. Liu, "An adaptive system for direct blind multi-channel equalization," in *Proc. 1st IEEE Signal Processing Workshop Signal Processing Advances in Wireless Communication*, Paris, France, Apr. 1997, vol. 1, pp. 85–88.
- [17] E. de Carvalho and D. T. M. Slock, "Cramér–Rao Bounds for semi blind, blind and training sequence based channel estimation," in *Proc. 1st IEEE Signal Processing Workshop on Signal Processing Advances in Wireless Communication*, Apr. 1997, vol. 1, pp. 129–132, Paris, France.
- [18] —, "Maximum-likelihood blind FIR multi-channel estimation with Gaussian prior for the symbols," in *Proc. IEEE Intl. Conf. Acoustics, Speech, Signal Processing*, Munich, Germany, Apr. 1997, vol. 5, pp. 3593–3596.
- [19] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via EM algorithm," *J. Royal Statist. Soc.*, vol. 39, ser. B, 1977.
- [20] Z. Ding, "A blind channel identification algorithm based on matrix outer-product," in *Proc. 1995 IEEE Int. Conf. Communications*, Dallas, TX, pp. 852–856.
- [21] Z. Ding and Z. Mao, "Knowledge based identification of fractionally sampled channels," in *Proc. 1995 Int. Conf. Acoustics, Speech, Signal Processing*, Detroit, MI, pp. 1996–1999.
- [22] G. Dong and R. Liu, "Adaptive blind channel identification," in *Proc. 1995 IEEE Int. Conf. Communications*, Dallas, TX, June 1996, pp. 828–831.
- [23] —, "An orthogonal learning rule for null-space tracking with implementation to blind two-channel identification," *IEEE Trans. Circuits Syst. I*, vol. 45, pp. 26–33, Jan. 1998.
- [24] D. Donoho, "On minimum entropy deconvolution," *Applied Time Series Analysis II*, D. F. Findley, Ed. New York: Academic, 1981.
- [25] T. Endres, B. D. O. Anderson, C. R. Johnson, and L. Tong, "On the robustness of FIR channel identification from second-order statistics," *IEEE Signal Processing Lett.*, vol. 3, pp. 153–155, May 1996.
- [26] M. V. Eyuboğlu and S. U. Qureshi, "Reduced-state sequence estimation with set partitioning and decision feedback," *IEEE Trans. Commun.*, vol. COM-36, pp. 13–20, Jan. 1988.
- [27] M. Feder and J. A. Catipovic, "Algorithms for joint channel estimating and data recovery—Application to underwater communications," *IEEE J. Oceanic Eng.*, vol. 16, pp. 42–55, Jan. 1991.
- [28] G. D. Forney, "The Viterbi algorithm," *Proc. IEEE*, vol. 61, pp. 268–278, Mar. 1972.
- [29] —, "Minimal bases of rational vector spaces, with applications to multivariable linear systems," *SIAM J. Control*, vol. 13, no. 3, pp. 493–520, May 1975.
- [30] B. Friedlander and B. Porat, "Asymptotically optimal estimation of MA and ARMA parameters of non-Gaussian processes from higher-order moments," *IEEE Trans. Automat. Control*, vol. 35, pp. 27–35, Jan. 1990.
- [31] P. A. Fuhrmann, *A Polynomial Approach to Linear Algebra*. New York: Springer-Verlag, 1996.
- [32] W. A. Gardner, *Cyclostationarity in Communications and Signal Processing*. New York: IEEE Press, 1994.
- [33] D. Gesbert and P. Duhamel, "Robust blind identification and equalization based on multi-step predictors," in *Proc. IEEE Intl. Conf. Acoust. Speech, Sig. Proc.*, Munich, Germany, Apr. 1997, vol. 5, pp. 2621–2624.
- [34] D. Gesbert, C. B. Papadias, and A. Paulraj, "Blind equalization of polyphase FIR channels: A whitening approach," Asilomar, CA, Nov. 1997.

- [35] M. Ghosh, "Blind decision feedback equalization for terrestrial television receivers," this issue, pp. 2070–2081.
- [36] M. Ghosh and C. L. Weber, "Maximum-likelihood blind equalization," *Opt. Eng.*, vol. 31, no. 6, pp. 1224–1228, June 1992.
- [37] G. Giannakis, "Filterbanks for blind channel identification and equalization," *IEEE Signal Processing Lett.*, vol. 7, July 1997.
- [38] G. Giannakis, Y. Inouye, and J. Mendel, "Cumulant based identification of multichannel moving-average models," *IEEE Trans. Automat. Contr.*, vol. 34, no. 7, pp. 783–787, July 1989.
- [39] G. Giannakis and J. Mendel, "Identification of nonminimum phase systems using higher order statistics," *IEEE Trans. Acoustics, Speech, Signal Processing*, vol. 37, pp. 7360–7377, Mar. 1989.
- [40] G. B. Giannakis and R. W. Heath, Jr., "Blind identification of multichannel FIR blurs and perfect image restoration," submitted for publication.
- [41] G. B. Giannakis, "Linear cyclic correlation approach for blind identification of FIR channels," in *Proc. 28th Asilomar Conf. Signals, Systems, and Computers*, Pacific Grove, CA, Nov. 1994, pp. 420–423.
- [42] G. B. Giannakis and W. Chen, "Blind blur identification and multichannel image restoration using cyclostationarity," in *Proc. IEEE Workshop Nonlinear Signal and Image Processing*, Halkidiki, Greece, June 1996, pp. 543–546.
- [43] G. B. Giannakis and S. D. Halford, "Performance analysis of blind equalizers based on cyclostationary statistics," in *Proc. 26th Conf. Information Sciences and Systems*, Princeton, NJ, Mar. 1994, pp. 711–716.
- [44] —, "Asymptotically optimal blind fractionally-spaced channel estimation and performance analysis," *IEEE Trans. Signal Processing*, vol. 45, pp. 1815–1830, July 1997.
- [45] G. B. Giannakis and C. Tepedelenlioglu, "Batch and adaptive direct blind equalizers of multiple FIR Channels: A deterministic approach," in *Proc. 30th Asilomar Conf. on Signals, Systems, and Computers*, Pacific Grove, CA, Nov. 1996.
- [46] D. N. Godard, "Self-recovering equalization and carrier tracking in two-dimensional data communication systems," *IEEE Trans. Commun.*, vol. COM-28, pp. 1867–1875, Nov. 1980.
- [47] G. C. Goodwin and K. S. Sin, *Adaptive Filtering, Prediction, and Control*. Englewood Cliffs, NJ: Prentice-Hall, 1984.
- [48] A. Gorokhov, P. Loubaton, and E. Moulines, "Second-order blind equalization in multiple input multiple output systems: A weighted least squares approach," in *Proc. IEEE Intl. Conf. Acoustics, Speech, Signal Processing*, Atlanta, GA, May 1996, vol. 5, pp. 2415–2418.
- [49] J. Gunther and A. Swindlehurst, "Algorithms for blind equalization with multiple antennas based on frequency domain subspaces," in *Proc. IEEE Intl. Conf. Acoustics, Speech, Signal Processing*, Atlanta, GA, May 1996, vol. 5, pp. 2419–2422.
- [50] M. L. Gürelli and C. L. Nikias, "EVAM: An eigenvector-based deconvolution of input colored signals," *IEEE Trans. Signal Processing*, vol. 43, pp. 134–149, Jan 1995.
- [51] M. I. Gürelli and C. L. Nikias, "A new eigenvector-based algorithm for multichannel blind deconvolution of input colored signals," in *Proc. IEEE Intl. Conf. Acoustics, Speech, Signal Proc.*, Minneapolis, MN, Apr. 1993, vol. 4, pp. 448–451.
- [52] S. D. Halford and G. B. Giannakis, "Direct blind equalization for transmitter induced cyclostationarity," in *Proc. 1st IEEE Signal Processing Workshop Signal Processing Advances in Wireless Communication*, Paris, France, Apr. 1997, vol. 1, pp. 117–120.
- [53] G. Harikuma and Y. Bresler, "Analysis and comparative evaluation of techniques for multichannel blind equalization," in *Proc. 8th IEEE Signal Processing Workshop Statistical and Array Signal Processing*, Corfu, Greece, June 1996, pp. 332–335.
- [54] D. Hatzinakos and C. Nikias, "Estimation of multipath channel response in frequency selective channels," *IEEE J. Select. Areas Commun.*, vol. 7, pp. 12–19, Jan. 1989.
- [55] Y. Hua, "Fast maximum likelihood for blind identification of multiple FIR channels," in *Proc. 28th Asilomar Conf. Signals, Systems, and Computers*, Pacific Grove, CA, Nov. 1994.
- [56] Y. Hua, "Fast maximum likelihood for blind identification of multiple FIR channels," *IEEE Trans. Signal Processing*, vol. 44, pp. 661–672, Mar. 1996.
- [57] Y. Hua, K. Abed-Meraim, and M. Wax, "Blind system identification using minimum noise subspace," in *Proc. 8th IEEE Signal Processing Workshop Statistical and Array Signal Processing*, Corfu, Greece, June 1996, pp. 308–311.
- [58] Y. Hua and W. Qiu, "Source correlation compensation for blind channel identification based on second-order statistics," *IEEE Signal Processing Lett.*, vol. 1, pp. 119–120, Aug. 1994.
- [59] Y. Hua and M. Wax, "Strict identifiability of multiple FIR channels driven by an unknown arbitrary sequence," *IEEE Trans. Signal Processing*, vol. SP-44, pp. 756–759, Mar. 1996.
- [60] R. A. Iltis, "A Bayesian maximum-likelihood sequence estimation algorithm for *a priori* unknown channels and symbol timing," *IEEE J. Select. Areas Commun.*, vol. 10, pp. 579–588, Apr. 1992.
- [61] R. A. Iltis, J. J. Shynk, and K. Giridhar, "Bayesian algorithms for blind equalization using parallel adaptive filtering," *IEEE Trans. Commun.*, vol. 42, pp. 1017–1032, Feb.–Apr. 1994.
- [62] A. Kagan, Y. V. Linnik, and C. R. Rao, *Characterization Problems in Mathematical Statistics*. New York: Wiley, 1973.
- [63] T. Kailath, *Linear Systems*. Englewood Cliffs, NJ: Prentice-Hall, 1980.
- [64] G. K. Kaleh and R. Vallet, "Joint parameter estimation and symbol detection for linear or nonlinear unknown dispersive channels," *IEEE Trans. Commun.*, vol. 42, pp. 2406–2413, July 1994.
- [65] V. Krishnamurthy and J. B. Moore, "On-line estimation of hidden Markov model parameters based on Kullback–Leibler information measure," *IEEE Trans. Signal Processing*, vol. 41, pp. 2557–2573, Aug. 1993.
- [66] M. Kristensson and B. Ottersten, "Statistical analysis of a subspace method for blind channel identification," in *Proc. IEEE Intl. Conf. Acoustics, Speech, Signal Processing*, Atlanta, GA, May 1996, vol. 5, pp. 2435–2438.
- [67] R. L. Lawrence and H. Kaufman, "The Kalman filter for the equalization of digital communication channel," *IEEE Trans. Commun.*, vol. COM-19, Dec. 1971.
- [68] G. Lee, S. B. Gelfand, and M. P. Fitz, "Bayesian techniques for blind deconvolution," *IEEE Trans. Commun.*, vol. 44, pp. 826–835, July 1996.
- [69] Y. Li and Z. Ding, "A new nonparametric cepstral method for blind channel identification from cyclostationary statistics," presented at 27th Asilomar Conf. Signals, Systems, and Computers, Asilomar, CA, Oct. 1993.
- [70] Y. Li and K. J. R. Liu, "On blind MIMO channel identification using second-order statistics," in *Proc. 1996 Conf. Information Science and Systems*, Princeton, NJ, vol. 2, pp. 1166–1171.
- [71] H. Liu, G. Xu, and L. Tong, "A deterministic approach to blind identification of multichannel FIR systems," presented at 27th Asilomar Conference on Signals, Systems, and Computers, Asilomar, CA, Oct. 1993.
- [72] H. Liu, G. Xu, L. Tong, and T. Kailath, "Recent developments in blind channel equalization: From cyclostationarity to subspaces," *Signal Processing*, vol. 50, pp. 83–99, 1996.
- [73] R. Liu, "Blind signal processing: An introduction," in *Proc. 1996 Intl. Symp. Circuits and Systems*, Atlanta, GA, vol. 2, pp. 81–83.
- [74] —, "A fundamental theorem on direct blind equalization of multiple-channel information transmission," in *Proc. 1996 Intl. Symp. Circuits and Systems*, vol. 1, pp. 705–708.
- [75] R. Liu and G. Dong, "A fundamental theorem for multiple-channel blind equalization," *IEEE Trans. Circuits Syst. I*, vol. 44, pp. 472–473, May 1997.
- [76] R. Liu and H. Luo, "Some results on multiple-channel blind equalization," in *Proc. '97 Euro. Conf. Circuit Theory and Design*, pp. 1347–1351.
- [77] —, "Direct blind separation of independent non-Gaussian signals with dynamic channels," presented at 5th IEEE Int. Workshop Cellular Neural Networks and their Applications, Apr. 1998.
- [78] L. Ljung, *System Identification: Theory for the User*. Englewood Cliffs, NJ: Prentice-Hall, 1987.
- [79] H. Luo and Y. Li, "Application of blind channel identification techniques to prestack seismic deconvolution," this issue, pp. 2082–2089.
- [80] J. Massey and M. Sain, "Inverse of linear sequential circuits," *IEEE Trans. Comput.*, vol. C-17, pp. 330–337, 1968.
- [81] J. M. Mendel, "Tutorial on higher-order statistics (spectra) in signal processing and system theory: Theoretical results and some applications," *Proc. IEEE*, vol. 79, pp. 278–305, Mar. 1991.
- [82] E. Moulines, P. Duhamel, J. F. Cardoso, and S. Mayrargue, "Subspace-methods for the blind identification of multichannel

- FIR filters," presented at ICASSP'94 Conf., Adelaide, Australia, Apr. 1994.
- [83] —, "Subspace-methods for the blind identification of multichannel FIR filters," *IEEE Trans. Signal Processing*, vol. 43, pp. 516–525, Feb. 1995.
- [84] D. Pal, "Fractionally spaced semiblind decision feedback equalization of wireless channels," in *Proc. 1993 IEEE Intl. Conf. Communications*, Geneva, Switzerland, pp. 1139–1143.
- [85] C. B. Papadias, "Methods for blind equalization and identification of linear channels," Ph.D. dissertation, École Nationale Supérieure des Télécommunications, Paris, France, Mar. 1995.
- [86] B. P. Paris, "Self-adaptive maximum-likelihood sequence estimation," presented at IEEE GLOBECOM, 1993.
- [87] S. Perreau, "A sub-optimal off-line approach for maximum likelihood blind equalization of a multiusers system," Univ. Connecticut, Storrs, Tech. Rep. 97-03, May 1997.
- [88] S. Perreau, L. White, and P. Duhamel, "A new decision-feedback equalizer incorporating fixed-lag smoothing," submitted for publication.
- [89] —, "A reduced computation multichannel adaptive equalizer based on HMM," in *Proc. 8th IEEE Signal Processing Workshop Statistical and Array Signal Processing*, Corfu, Greece, June 1996, pp. 156–159.
- [90] B. Porat and B. Friedlander, "Blind equalization of digital communication channels using high-order moments," *IEEE Trans. Signal Processing*, vol. 39, pp. 522–526, 1991.
- [91] B. Porat, *Digital Processing of Random Signals*. Englewood Cliffs, NJ: Prentice-Hall, 1993.
- [92] W. Qiu and Y. Hua, "Performance analysis of a blind channel identification method based on second-order statistics," in *Proc. 21st Asilomar Conf. Signals, Systems, and Computers*, Pacific Grove, CA, Nov. 1995, vol. 2, pp. 1458–1462.
- [93] —, "Performance analysis of the subspace method for blind channel identification," *Signal Processing*, vol. 50, no. 71–81, 1996.
- [94] L. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proc. IEEE*, vol. 77, no. 2, pp. 257–285, Feb. 1989.
- [95] R. Raheli, A. Polydoros, and C. K. Tzou, "Per-survivor processing: A general approach to MLSE in uncertain environments," *IEEE Trans. Commun.*, vol. 43, pp. 354–364, Feb.–Apr. 1995.
- [96] J. Rissanen, "Modeling by shortest data description," *Automatica*, vol. 14, pp. 465–471, 1978.
- [97] E. Robinson, "Predictive deconvolution and wavelet analysis," presented at Intl. Conf. Communications, Computing, Control, Signal Processing, Stanford, CA, June 1996.
- [98] E. A. Robinson, "T. Tomographic deconvolution of echograms," in *Communications, Computation, Control and Signal Processing: A Tribute to Thomas Kailath*, A. Paulraj, V. Roychowdhury and C. Schaper, Eds. Norwell, MA: Kluwer, 1997.
- [99] —, "Predictive decomposition of seismic traces," *Geophysics*, vol. 22, pp. 767–778, 1957.
- [100] W. Rudin, *Real and Complex Analysis*. New York: McGraw-Hill, 1986.
- [101] Y. Sato, "A method of self-recovering equalization for multilevel amplitude-modulation," *IEEE Trans. Commun.*, vol. COM-23, pp. 679–682, June 1975.
- [102] S. Schell and D. Smith, "Improved performance of blind equalization using prior knowledge of transmitter filter," in *Proc. 1994 IEEE MILCOM*, Fort Monmouth, NJ, vol. 1, pp. 128–132.
- [103] S. V. Schell, D. L. Smith, and S. Roy, "Blind channel identification using subchannel response matching," presented at 26th Conf. Information Sciences and Systems, Princeton, NJ, Mar. 1994.
- [104] E. Serpedin and G. Giannakis, "Blind channel identification and equalization with modulation induced cyclostationarity," presented at 31st Annu. Conf. Information Sciences and Systems, Baltimore, MD, Mar. 1997.
- [105] N. Seshadri, "Joint data and channel estimation using fast blind trellis search techniques," in *Proc. Globecom'90*, 1991, pp. 1659–1663.
- [106] —, "Joint data and channel estimation using blind trellis search techniques," *IEEE Trans. Commun.*, vol. 42, pp. 1000–1011, Feb.–Apr. 1994.
- [107] M. Shao and C. L. Nikias, "An ML/MMSE estimation approach to blind equalization," in *Proc. ICASSP'94 Conf.*, Adelaide, Australia, vol. 4, pp. IV–569–572.
- [108] D. Slock, "Blind fractionally-spaced equalization, perfect reconstruction filterbanks, and multilinear prediction," presented at ICASSP'94 Adelaide, Australia.
- [109] —, "Blind joint equalization of multiple synchronous mobile users using oversampling and/or multiple antennas," presented at 28th Asilomar Conf. Signals, Systems, Computers, Monterey, CA, Oct. 1994.
- [110] D. Slock and C. B. Papadias, "Further results on blind identification and equalization of multiple FIR channels," in *IEEE Proc. Intl. Conf. Acoustics, Speech, Signal Processing*, Detroit, MI, Apr. 1995, pp. 1964–1967.
- [111] 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, pp. 29–31, Feb. 1994.
- [112] —, "Blind separation of synchronous co-channel digital signals using an antenna array. I. Algorithms," *IEEE Trans. Signal Processing*, vol. 44, pp. 1184–1197, May 1996.
- [113] D. M. Titterington, A. F. M. Smith, and U. E. Makov, *Statistical Analysis of Finite Mixture Distributions* (Wiley Series in Probability and Mathematical Statistics). New York: Wiley, 1985.
- [114] D. M. Titterington, "Recursive parameter estimation using incomplete data," *J. Royal Statist. Soc. B*, vol. 46, no. 2, pp. 257–267, 1984.
- [115] L. Tong, "Identification of multichannel MA parameters using higher-order statistics," submitted for publication.
- [116] L. Tong, Y. Inouye, and R. Liu, "A finite-step global convergence algorithm for the parameter estimation of multichannel MA processes," *IEEE Trans. Signal Processing*, vol. 40, pp. 2547–2558, Oct. 1992.
- [117] L. Tong, G. Xu, B. Hassibi, and T. Kailath, "Blind identification and equalization of multipath channels: A frequency domain approach," *IEEE Trans. Inform. Theory*, vol. 41, pp. 329–334, Jan. 1995.
- [118] L. Tong, G. Xu, and T. Kailath, "A new approach to blind identification and equalization of multipath channels," presented at 25th Asilomar Conf., Pacific Grove, CA, Nov. 1991.
- [119] —, "Blind identification and equalization based on second-order statistics: A time domain approach," *IEEE Trans. Inform. Theory*, vol. 40, pp. 340–349, Mar. 1994.
- [120] L. Tong and H. Zeng, "Blind channel identification using cyclic spectra," in *Proc. 26th Conf. Information Sciences and Systems*, Princeton, NJ, Mar. 1994, pp. 711–716.
- [121] —, "Joint order detection and blind channel estimation by least squares smoothing," *IEEE Trans. Signal Processing*, to be published.
- [122] L. Tong and Q. Zhao, "Blind channel estimation by least squares smoothing," in *Proc. 1998 IEEE Intl. Conf. Acoustics, Speech, Signal Processing*, Seattle, WA, vol. 5, pp. 2121–2124.
- [123] —, "Joint order detection and channel estimation by least squares smoothing," presented at 30th Conf. Information Sciences and Systems, Princeton, NJ, Mar. 1998.
- [124] J. R. Treichler and B. G. Agee, "A new approach to multipath correction of constant modulus signals," *IEEE Trans. Audio, Speech, Signal Processing*, vol. ASSP-31, pp. 459–472, Apr. 1983.
- [125] J. K. Tugnait, "Identification of nonminimum phase linear stochastic systems," *Automatica*, vol. 22, pp. 454–464, 1986.
- [126] —, "Blind equalization and estimation of digital communication FIR channels using cumulant matching," *IEEE Trans. Commun.*, vol. 43, pp. 1240–1245, Feb.–Apr. 1995.
- [127] A. van der Veen, S. Talwar, and A. Paulraj, "A subspace approach to blind space-time signal processing for wireless communication systems," *IEEE Trans. Signal Processing*, vol. 45, pp. 173–190, Jan 1997.
- [128] M. Viberg, "Subspace fitting concepts in sensor array processing," Ph.D. dissertation, Linköping Univ., Linköping, Sweden, Oct. 1989.
- [129] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Trans. Audio, Speech, Signal Processing*, vol. ASSP-33, pp. 387–392, Apr. 1985.
- [130] E. Weinstein, M. Feder, and A. Oppenheim, "Sequential algorithms for parameter estimation based on the Kullback–Leibler information measure," *IEEE Trans. Signal Processing*, vol. SP-38, pp. 1652–1654, Sept. 1990.
- [131] L. B. White and V. Krishnamurthy, "Adaptive blind equalization of FIR channels using hidden Markov Models," presented at Int. Conf. Communications, Geneva, June 1993.

- [132] L. B. White, S. Perreau, and P. Duhamel, "Reduced computation blind equalization for FIR channel input Markov models," in *Proc. Int. Conf. Communications*, Seattle, WA, June 1995, pp. 993–997.
- [133] R. A. Wiggings, "Minimum entropy deconvolution," *Geoplotation*, vol. 16, pp. 21–35, 1978.
- [134] G. Xu, H. Liu, L. Tong, and T. Kailath, "A least-squares approach to blind channel identification," *IEEE Trans. Signal Processing*, vol. 43, pp. 2982–2993, Dec. 1995.
- [135] D. Yellin and B. Porat, "Blind identification of FIR systems excited by discrete-alphabet inputs," *IEEE Trans. Signal Processing*, vol. 41, pp. 1331–1339, Mar. 1993.
- [136] H. Zeng, "Blind estimation: Performance analysis and new algorithms," Ph.D. dissertation, Univ. Connecticut, Storrs, May 1997.
- [137] H. Zeng and L. Tong, "Some new results on blind channel estimation: Performance and algorithms," presented at 27th Conf. Information Sciences and Systems, Baltimore, MD, Mar. 1995.
- [138] H. Zeng and L. Tong, "Connections between the least-squares and subspace approaches to blind channel estimation," *IEEE Trans. Signal Processing*, vol. 44, pp. 1593–1596, June 1996.
- [139] ———, "Blind channel estimation using the second-order statistics: Algorithms," *IEEE Trans. Signal Processing*, vol. 45, Aug. 1997.
- [140] ———, "Blind channel estimation using the second-order statistics: Asymptotic performance analysis," *IEEE Trans. Signal Processing*, vol. 45, Aug. 1997.
- [141] Q. Zhao and L. Tong, "Adaptive blind channel estimation by least squares smoothing," submitted for publication.
- [142] ———, "Fast convergent adaptive blind channel estimation and deconvolution," presented at 30th Conf. Information Sciences and Systems, Princeton, NJ, Mar. 1998.
- [143] M. Zoltowski and D. Tseng, "Blind channel identification for narrowband digital communication based on parametric modeling of channel impulse response," in *Proc. 1997 Allerton Conf. Communications, Control, and Computing*, Urbana, IL, pp. 503–512.

**Lang Tong** (Member, IEEE), for a photograph and biography, see this issue, p. 1906.



**Sylvie Perreau** was born in France in 1969. In 1993 she graduated from the École Nationale Supérieure de l'Électronique et des Applications and obtained the M.Sc. degree from the University of Cergy-Pontoise. In 1997, she received the Ph.D. degree in signal processing from Telecom Paris.

In 1997 she worked at the University of Connecticut, Storrs. She is now with the Institute for Telecommunications Research in Adelaide, Australia. Her research interests are in blind equalization, interference mitigation, and blind array processing.