Blind System Identification

KARIM ABED-MERAIM, WANZHI QIU, MEMBER, IEEE, AND YINGBO HUA, SENIOR MEMBER, IEEE

Blind system identification (BSI) is a fundamental signal processing technology aimed at retrieving a system's unknown information from its output only. This technology has a wide range of possible applications such as mobile communications, speech reverberation cancellation, and blind image restoration. This paper reviews a number of recently developed concepts and techniques for BSI, which include the concept of blind system identifiability in a deterministic framework, the blind techniques of maximum likelihood and subspace for estimating the system's impulse response, and other techniques for direct estimation of the system input.

Keywords— *Blind techniques, equalization, multichannels, source separation, system identification.*

I. INTRODUCTION

Blind system identification (BSI) is a fundamental signal processing technology aimed at retrieving a system's unknown information from its output only. This technology is particularly suitable for applications where all the available data are generated from an unknown system driven by an unknown input. The word "blind" simply means that the system's input is not available to (cannot be seen by) the signal processor. The task of BSI (Fig. 1) is to identify the input $s(.)$ and/or the system function H from the output $y(.)$. Note that if either the system function or the input signal is known, it becomes a more standard and simpler problem.

The notion of BSI (or the like, such as blind deconvolution) has become well known since the early 1980's. During the 1990's, there has been an increasing research interest devoted to BSI. Unlike most of the work in the 1980's, the work in the 1990's tends to explore to a higher degree the diversities inherent in multiple-output systems. The multiple-output systems arise from multisensor systems, multichannel data acquisition, or fractional sampling systems. A huge number of research articles recently produced by the signal processing community contain a significant amount of new knowledge that can be applied by many other communities, such as the seismic community, speech community, and medical community.

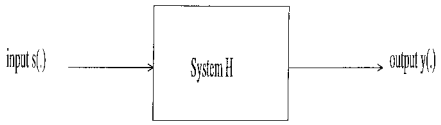

Fig. 1. A system with unknown input and unknown system function.

In this paper, we will first give an overview of BSI and its applications and then review some recently developed basic concepts and techniques of BSI. Other well-established concepts and techniques will also be discussed.

The rest of this paper is organized as follows. In the next section, we present briefly some areas of applications for BSI, which include data communications, speech recognition, image restoration, and seismic signal processing. Section III reviews the basic concepts and techniques of BSI. Following the problem formulation, we highlight the channel identifiability conditions, discuss several important results concerning system identifiability, and then present the main and most recently developed techniques for BSI. Section IV reviews some other concepts and techniques, particularly the methods based on high-order statistics (HOS's), and discusses the case of systems with multiple inputs.

II. AREAS OF APPLICATION

Fig. 1 depicts the BSI problem, where $s(\cdot)$ is the input signal and H is the system function. When the system is linear and time invariant, the system output can be described by the following convolution model:

$$y(\cdot) = s(\cdot) * h(\cdot) + w(\cdot) \tag{1}$$

where $*$ denotes the convolution. Because of (1), the blind identification of $s(.)$ or $h(.)$ is also known as blind deconvolution.

The need for BSI arises from a number of applications, which include the following.

Manuscript received November 25, 1996; revised May 1, 1997. This work was supported by the Australian Research Council and by the Australian Cooperative Research Center for Sensor Signal and Information Processing.

K. Abed-Meraim and Y. Hua are with the Department of Electrical and Electronic Engineering, University of Melbourne, Parkville, Victoria 3052 Australia (e-mail: abed@ee.mu.oz.au; hua@ee.mu.oz.au).

W. Qiu was with the Department of Electrical and Electronic Engineering, University of Melbourne, Parkville, Victoria 3052 Australia. He is now with the Motorola Australian Research Centre, Botany 2019, NSW Australia.

Publisher Item Identifier S 0018-9219(97)05723-X.

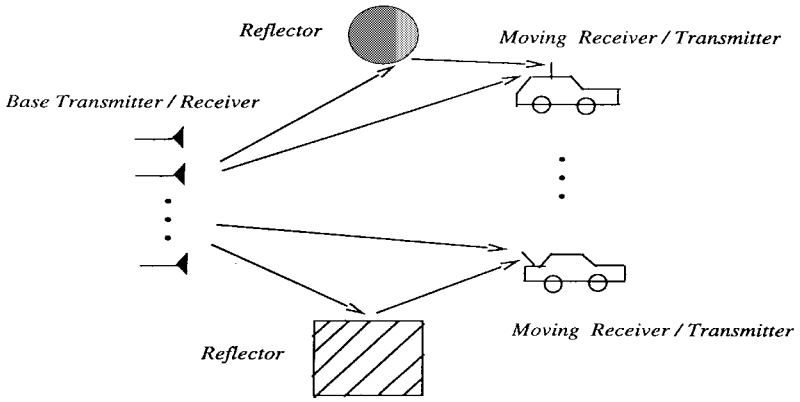

Fig. 2. Mobile communication.

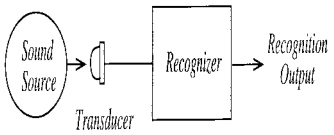

Fig. 3. Speech recognition.

A. Data Communications

In data communications, an unknown channel frequency response with finite bandwidth often causes intersymbol interferences (ISI's). To eliminate the ISI's without knowing the input data, the channel response needs to be identified "blindly."

BSI is particularly important for mobile communications, where severe ISI can arise from the time-varying multipath fading that commonly exists in a mobile communications environment (Fig. 2). The varying channel characteristics must be identified and equalized in real time to maintain the correct flow of information. The channel identification and equalization technique currently used requires a major fraction of the channel capacity to send a training sequence over the channel. It should be noted that while the density of mobile users in a given city area is likely to increase dramatically, the number of radio channels in that area remains constant. Although many techniques, such as code division multiple access (CDMA), can be used to increase the channel capacity, the fraction of the channel capacity currently used for channel identification and equalization is very considerable. To save this fraction of channel capacity, blind channel identification is an attractive approach. Using the blind channel identification techniques, the receiver can identify the channel characteristics and equalize the channel based on the received signal. No training sequence is needed, which saves the channel capacity.

B. Speech Recognition and Reverberation Cancellation

Speech recognition (Fig. 3) is useful in many areas [9]. It can be used for controlling the actions of a machine and for entering and retrieving data. In speech recognition, the available signal for the recognizer is the convolution of the original speech signal, the impulse response of the transducer, and the impulse response of the surrounding environment.

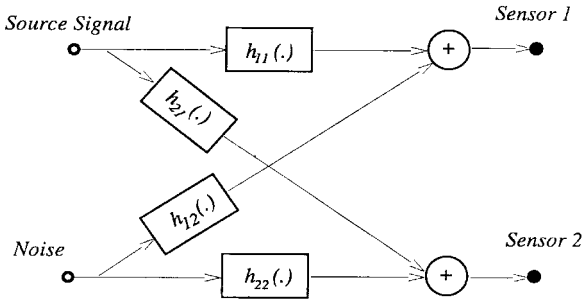

Fig. 4. Speech reverberation and noise cancellation.

Training is now commonly used for the recognizer to obtain knowledge of the transducer. It is obvious, however, that the transducers as well as their surrounding environments vary tremendously [9]. Telephone handsets vary in degrees of distortion, spectral shaping, and response level. Microphones are built by a variety of manufacturers and are located at various positions on the handset, with openings of different sizes, and lie in different points within the sound field around the mouth. A recognizer that is well-trained for one particular transducer in one particular environment might perform very badly when another transducer is used or when the same transducer is used in a different environment. Therefore, it is desirable to build the recognizer without specifying the characteristics of the transducer. The BSI is useful here in identifying the impulse response of the transducer and in enabling recovery of the original speech signal.

Speech reverberation cancellation is necessary when the original speech signal is corrupted by the acoustics of the surrounding environment, which is known as reverberation (Fig. 4). Because the acoustics of the surrounding environment are dependent on the geometry and materials of the room and the speaker's location, the reverberation is ever changing. Since the original speech signal is unobservable and the acoustics of the surrounding environment are unknown, BSI can be used in adaptive cancellation of the reverberation.

C. Image Restoration

BSI is needed for image restoration in many applications such as astronomy, remote sensing, and medical imaging

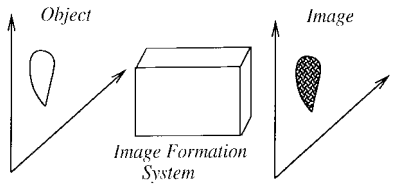

Fig. 5. Image formation.

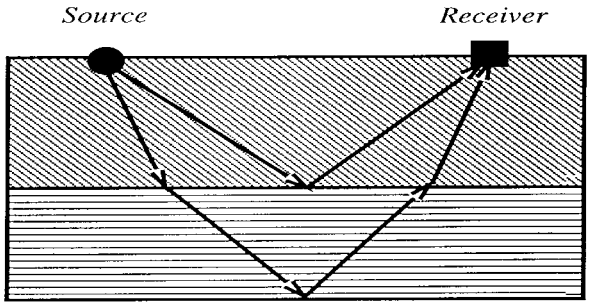

Fig. 6. Exploration seismology.

[11], [12]. In this situation, we have a system that represents blurring effects caused by, for example, camera motion during exposure or inaccurately focused lenses. The original image constitutes the system input. The system output is a blurred version of the original image, i.e., an image degraded by the convolution with the point spread function of the blurring system (Fig. 5). Unfortunately, in many practical situations [5], [11], the point-spread function of the blurring system is unknown and little information is available about the original image. Given the blurred image, the task is to restore it "blindly."

D. Seismic Signal Processing

In exploration seismology (Fig. 6), a charge of dynamite is exploded in the earth and a geophone is used to receive the reflected and/or diffracted signal [4], [6]. This signal is used to estimate the reflection coefficients, which are associated with the impulse responses of the various layers in the earth and unravel its physical characteristics. Here, the received signal is made up of echoes produced at the different earth layers in response to the excitation, which is ordinarily in the form of a short-duration pulse. This again is a BSI problem due to the fact that the exact waveform of the excitation responsible for the generation of the received signal is usually unknown.

III. Some Basic Concepts and Techniques

While the applications of BSI are wide ranging, some basic concepts and techniques are common for all applications. In this section, we provide an introduction to these concepts and techniques.

A. Problem Formulation

Consider a mathematical model where the input and the output are discrete, the system operator H is linear and shift invariant, the system is driven by a single-input sequence $s(k)$ and yields M output sequences $y_1(k), \cdots, y_M(k)$, and the system has finite impulse responses (FIR's) $h_i(k)$, $k = 0, 1, \cdots, L$, and $i = 1, \cdots, M$. Such a system model can be described as follows (in the absence of noise):

$$\begin{cases} y_1(k) = s(k) * h_1(k) \\ y_2(k) = s(k) * h_2(k) \\ \quad \vdots \\ y_M(k) = s(k) * h_M(k) \end{cases} \tag{2}$$

where $*$ denotes linear convolution. This multiple-FIR-channels system obviously is a useful model to describe the case where a single unknown source and multiple spatially or/and temporally distributed sensors exist. The requirement of a single unknown source may seem restrictive but it is common in many applications, as described in the previous section. In the case of multiple sources, which are not separable by simple operations such as bandpass filtering of the channel outputs, the model of multiple-input and multiple-output systems must be considered. This topic will be further discussed in Section IV.

The requirement of FIR is general enough for practical applications, as an infinite impulse response (IIR) system can be well approximated by FIR provided the model order L is large enough. If the order L is too large, IIR would be a better model in that fewer system parameters are necessary. Nevertheless, FIR often leads to simple development of signal processing algorithms. The estimation of L is still a challenging research topic, but a simple technique of estimating L will be referred to in Section III-C.

In this paper, our focus is on the one-dimensional case. The two-dimensional extension of most concepts and techniques shown in this paper remains a research topic [116], [117].

The analysis of the model (2) can be made convenient by using its matrix form as follows. We stack all channel outputs into a single vector, i.e., let

$$\mathbf{y} = [\mathbf{y}_1^T \quad \mathbf{y}_2^T \quad \cdots \quad \mathbf{y}_M^T]^T \tag{3}$$

with

$$\mathbf{y}_i = [y_i(0), \cdots, y_i(N-1)]^T \tag{4}$$

and the superscripts ^{T} and ^{H} denote the transpose and the conjugate transpose, respectively. Then we have

$$\mathbf{y} = \mathbf{H}_M \mathbf{s} \tag{5}$$

where $\mathbf{s}$ is the input vector

$$\mathbf{s} = [s(-L), s(-L+1), \cdots, s(N-1)]^T. \tag{6}$$

$\mathbf{H}_M$ is known as a generalized Sylvester matrix [15]

$$\mathbf{H}_M = \begin{bmatrix} \mathbf{H}_{(1)} \\ \vdots \\ \mathbf{H}_{(M)} \end{bmatrix} \tag{7}$$

where $\mathbf{H}_{(i)}$ is the $N \times (N+L)$ Sylvester matrix of the ith channel response

$$\mathbf{H}_i = \begin{bmatrix} h_i(L) & \cdots & h_i(0) & \cdots & 0 \\ \vdots & \ddots & & \ddots & \vdots \\ 0 & \cdots & h_i(L) & \cdots & h_i(0) \end{bmatrix}. \quad (8)$$

Equation (5) may be subject to further conditions depending on applications. For example, the input sequence may be generated from convolution between a completely unknown sequence and a common channel response. For a simple presentation of some basic concepts and techniques of BSI, however, no further constraints will be imposed on the model (5) in the next section.

B. Identifiability

In general, a system is considered to be completely identifiable if all unknown parameters in the system can be uniquely determined by the available data. It is clear from (5), however, that in the absence of noise, a given output $\mathbf{y}$ can only at best imply a unique input $\mathbf{s}$ and a unique system matrix $\mathbf{H}_M$ up to an unknown scalar. This unknown scalar has to be determined by further knowledge available about the model. For example, if a sample of the input sequence is known, then this scalar is uniquely determined. Given the model (5), we define identifiability as follows.

The system (5) is said to be *identifiable* if a given output $\mathbf{y}$ implies a unique input $\mathbf{s}$ and a unique system matrix $\mathbf{H}_M$ up to an unknown scalar.

A detailed study on the identifiability is shown in [22]. The conditions required for the system to be identifiable relate to the notions of "zeros" and "modes," which are described below.

1) Zeros: The ith channel of the system is said to have a "zero" z_0 if the ith channel's transfer function $H_i(z)$ is zero at $z = z_0$, i.e.,

$$\begin{aligned} H_i(z_0) &= \sum_{k=0}^{L} h_i(k) z_0^{-k} \\ &= 0. \end{aligned}$$

It is clear that a channel of order L must have L zeros and, on the other hand, that the L zeros completely describe the channel up to a scalar.

2) Modes: A mode is any sequence of the form $m(k) = k^n z^k$ where n is an integer and z is a complex number. The integer n is also referred to as the order of the mode, and the complex number z is the root of the mode. A finite input sequence $\{s(k), k = -L, \cdots, N-1\}$ of length $N+L$ is said to have p modes if $s(k)$ can be written as a linear combination of p modes of length $N+L$, i.e., for $k = -L, \cdots, N-1$

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

where for each (nonzero) mode, all the coefficients of the corresponding lower order modes may or may not be zero. For example, the sequence $c_1 k^2 z_1^k + c_2 k z_1^k + c_3 z_1^k + c_4 z_2^k$ is said to have four modes as long as c_1 and c_4 are nonzeros. A channel is said to be associated with a mode $m(k)$ if the root of the mode is a zero of the channel. A simple test of the number of modes is as follows. It can be said that $s(k)$ contains $p+1$ or more modes if and only if the $(p+1) \times (N+L-p)$ Hankel matrix

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

has full row rank where $N+L-p \geq p+1$.

Given the notions of zero and mode, the identifiability conditions can be stated as follows.

3) Necessary Identifiability Conditions [17]: The system is identifiable only if the following conditions are met:

- all channels do not share a common zero, i.e., there is no z_0 such that $H_i(z_0) = 0$ for all i;
- the number of modes in the input sequence is larger than or equal to $L+2$;
- $N \geq L+2$ (where N is the number of samples from each channel).

4) Sufficient Identifiability Condition [17]: The system is identifiable if the following conditions are true:

- all channels do not share a common zero;
- the number of modes in the input sequence is larger than or equal to $2L+1$;
- $N \geq 3L+1$.

It is possible that the sufficient identifiability condition can be relaxed. However, more research in this area is required (see [22] and [23]). The identifiability conditions shown above essentially ensure the following intuitive requirements.

- All channels in the system must be different enough from each other. They cannot be identical, for example.
- The input sequence must be complex enough. It cannot be zero, a constant, or a single sinusoid, for example.
- There must be enough output samples available. A set of available data cannot yield sufficient information on a larger set of unknown parameters, for example.

The identifiability can also be discussed with respect to a given identification technique [19] or in the case where certain statistics of the system outputs are assumed to be given [16]. Furthermore, the conditions of identifiability can be traded off for each other. For example, if the number of available output samples is infinite, the input is non-Gaussian stationary random process. Hence, the HOS's of the output samples can be exactly computed and the system can be identified exactly, even when the channels share common zeros. Note that all zeros of a single channel system are "shared by all channels." Such a system can be identified exactly if the HOS's of the output are known exactly [20] or if the second-order statistics (SOS's) of the output are known exactly and the common zeros are inside the unit circle (minimum-phase condition [21]).

C. Estimation of System Function

We now introduce two powerful techniques for identifying the system function $h_i(k)$ of the model formulated in Section III-A. They are the maximum-likelihood (ML) method and the subspace method.

1) ML Method: ML is a classic approach applicable to any parameter estimation problem where the probability density function (PDF) of the available data is known. Assuming that the system output vector is corrupted by additive circular white Gaussian noise vector $\mathbf{w}$,[1] the system output vector becomes

$$\mathbf{y} = \mathbf{H}_M\mathbf{s} + \mathbf{w} \tag{9}$$

and the PDF of $\mathbf{y}$ is given by

$$p(\mathbf{y}) = \frac{1}{\pi^N\sigma^{2N}} \exp\left(-\frac{1}{\sigma^2}\|\mathbf{y} - \mathbf{H}_M\mathbf{s}\|^2\right) \tag{10}$$

where σ^2 is the variance of each complex element of $\mathbf{w}$, and $\|.\|$ denotes two-norm. The ML estimates of $\mathbf{H}_M$ and $\mathbf{s}$ are given by those arguments that maximize the PDF $p(\mathbf{y})$, i.e.,

$$\begin{aligned}(\mathbf{H}_M, \mathbf{s})_{\mathrm{ML}} &= \arg\max_{\mathbf{H}_M,\mathbf{s}} p(\mathbf{y}) \\ &= \arg\min_{\mathbf{H}_M,\mathbf{s}} \|\mathbf{y} - \mathbf{H}_M\mathbf{s}\|^2\end{aligned} \tag{11}$$

where proper constraints on $\mathbf{H}_M$ and $\mathbf{s}$ are imposed. Note that such ML criterion is equivalent to the least-square (LS) criterion, for which the knowledge of the PDF of $\mathbf{y}$ is not necessary. For any given $\mathbf{H}_M$, the ML estimate of $\mathbf{s}$ that minimizes the quadratic function $\|\mathbf{y} - \mathbf{H}_M\mathbf{s}\|^2$ is known to be

$$\mathbf{s}_{\mathrm{ML}} = (\mathbf{H}_M^H\mathbf{H}_M)^{-1}\mathbf{H}_M^H\mathbf{y}. \tag{12}$$

(Under the necessary identifiability condition, the matrix $\mathbf{H}_M$ is known to have full column rank [22].) Using this estimate in (11) yields

$$(\mathbf{H}_M)_{\mathrm{ML}} = \arg\min_{\mathbf{H}_M} \|(\mathbf{I} - \mathbf{P}_H)\mathbf{y}\|^2 \tag{13}$$

where $\mathbf{P}_H$ is the orthogonal projection matrix onto the range of $\mathbf{H}_M$, i.e.,

$$\mathbf{P}_H = \mathbf{H}_M\left(\mathbf{H}_M^H\mathbf{H}_M\right)^{-1}\mathbf{H}_M^H. \tag{14}$$

Although the minimization in (13) is computationally much more efficient than that in (11), it is still highly nonlinear. Therefore, the computation of (13) has to be iterative in nature. Many iterative optimization approaches such as [25], [26], and [108] can be applied to compute (13). Below, however, we introduce a more elegant technique.

[1] A complex random vector $\mathbf{w}$ is called circular white Gaussian if $\begin{bmatrix}\Re e(\mathbf{w}) \\ \Im m(\mathbf{w})\end{bmatrix}$ is white Gaussian.

Theorem 1 [17]: Define

$$\mathbf{G}_2^H = [-\overline{\mathbf{H}}_2 \quad \overline{\mathbf{H}}_1]$$

$$\mathbf{G}_q^H = \left[\begin{array}{ccc|c} & \mathbf{G}_{q-1}^H & & \mathbf{0} \\ \hline -\overline{\mathbf{H}}_{(q)} & & & \overline{\mathbf{H}}_1 \\ & \ddots & & \vdots \\ & & -\overline{\mathbf{H}}_{(q)} & \overline{\mathbf{H}}_{(q-1)} \end{array}\right] \tag{15}$$

where $q = 3, \cdots, M$ and $\overline{\mathbf{H}}_{(i)}$ is the top-left $(N-L)\times N$ submatrix of $\mathbf{H}_{(i)}$. Then, provided that all channels do not share a common zero and $N \geq 2L$ (for $M = 2$, only $N > L$ is required), an orthogonal complement matrix of the generalized Sylvester matrix $\mathbf{H}_M$ is $\mathbf{G}_M$, i.e.

$$\mathbf{P}_G + \mathbf{P}_H = \mathbf{I} \tag{16}$$

where $\mathbf{I}$ is the identity matrix, and $\mathbf{P}_G$ and $\mathbf{P}_H$ denote the orthogonal projection matrixes onto range $(\mathbf{G}_M)$ and range $(\mathbf{H}_M)$, respectively.

Under the condition of this theorem, the minimization of (13) becomes

$$\begin{aligned}\mathbf{h}_{\mathrm{ML}} &= \arg\min_{\mathbf{h}} \|\mathbf{P}_G\mathbf{y}\|^2 \\ &= \arg\min_{\mathbf{h}} \mathbf{y}^H\mathbf{P}_G\mathbf{y} \\ &= \arg\min_{\mathbf{h}} \mathbf{y}^H\mathbf{G}_M(\mathbf{G}_M^H\mathbf{G}_M)^{\#}\mathbf{G}_M^H\mathbf{y}\end{aligned} \tag{17}$$

where $\mathbf{h}$ is the vector of all channels' impulse responses [i.e., $\{h_i(k), k = 0, 1, \cdots, L, i = 1, \cdots, M\}$] and the superscript $^{\#}$ denotes a Moore–Penrose pseudoinverse with rank $NM - N - L$ [17].

The following is a matrix form of the commutativity property of linear convolution:

$$\mathbf{G}_M\mathbf{y} = \mathbf{Y}_M\mathbf{h} \tag{18}$$

where $\mathbf{Y}_M$ is defined by

$$\mathbf{Y}_2 = [\mathbf{Y}_{(2)} - \mathbf{Y}_{(1)}]$$

$$\mathbf{Y}_q = \left[\begin{array}{ccc|c} & \mathbf{Y}_{q-1} & & \mathbf{0} \\ \hline \mathbf{Y}_{(q)} & & & -\mathbf{Y}_{(1)} \\ & \ddots & & \vdots \\ & & \mathbf{Y}_{(q)} & -\mathbf{Y}_{(q-1)} \end{array}\right] \tag{19}$$

with $q = 3, \cdots, M$ and

$$\mathbf{Y}_{(i)} = \begin{bmatrix} y_i(L) & \cdots & y_i(0) \\ \vdots & & \vdots \\ y_i(N-1) & \cdots & y_i(N-L-1) \end{bmatrix}. \tag{20}$$

Combining (17) with (18) yields

$$\mathbf{h}_{\mathrm{ML}} = \arg\min_{\mathbf{h}} \mathbf{h}^H\mathbf{Y}_M^H(\mathbf{G}_M^H\mathbf{G}_M)^{\#}\mathbf{Y}_M\mathbf{h}. \tag{21}$$

This expression suggests the following two-step ML (TSML) method.

Step 1) $\mathbf{h}_c = \arg\ \min_{\|\mathbf{h}\|=1} \mathbf{h}^H \mathbf{Y}_M^H \mathbf{Y}_M \mathbf{h}$.

Step 2) $\mathbf{h}_e = \arg\ \min_{\|\mathbf{h}\|=1} \mathbf{h}^H \mathbf{Y}_M^H (\mathbf{G}_c^H \mathbf{G}_c)^{\#} \mathbf{Y}_M \mathbf{h}$, where $\mathbf{G}_c$ is $\mathbf{G}_M$ constructed from $\mathbf{h}_c$ according to (15).

The first step comes from (21) by setting the weighting matrix $(\mathbf{G}_M^H \mathbf{G}_M)^{\#}$ to an identity matrix. It can be shown that Step 1) of the algorithm yields the exact estimate of $\mathbf{h}$ in the absence of noise (or when the noise is white and the data length is infinite) and that Step 2) of the algorithm yields the optimum (ML) estimate of $\mathbf{h}$ at a relatively high signal-to-noise ratio (SNR).

The computations required by the TSML method largely depend on how the matrix multiplications, matrix inverse, and minimum eigenvectors are numerically computed. A novel implementation reported in [18] can reduce the computations of the TSML method to the order of N^2 or even N. The order N implementation exploits a property that the orthogonal complement matrix $\mathbf{G}_M$ becomes a band-limited block Toeplitz matrix after a certain row and column permutation.

The first step of the TSML method is known to coincide with a method based on a "cross-relation" (CR) property of the single-input–multiple-output system. This cross relation is as follows:

$$y_i(k) * h_j(k) = y_j(k) * h_i(k). \tag{22}$$

This is a linear equation satisfied by every pair of channels. By collecting all possible pairs of M channels, one can easily establish a set of linear equations. In matrix form, this set of equations can be expressed as

$$\mathbf{Y}_M \mathbf{h} = 0 \tag{23}$$

where $\mathbf{Y}_M$ turns out to be the same as defined by (19). In the presence of noise, the solution to (23) can be naturally replaced by the LS solution

$$\mathbf{h}_{\mathrm{CR}} = \arg \min_{\|\mathbf{h}\|=1} \mathbf{h}^H \mathbf{Y}_M^H \mathbf{Y}_M \mathbf{h}. \tag{24}$$

The CR method is clearly the first step of the two-step ML method and can be viewed as one-step ML (although it does not achieve the ML estimate, even when the SNR is relatively high).

The CR method is named as the LS method in [19] because it represents the LS solution to the CR equation (22). This should not confuse the LS interpretation of the TSML method, as the TSML method minimizes the LS errors between the original data $\mathbf{y}$ and the data model $\mathbf{H}_M \mathbf{s}$.

Last, we note that the CR equation (22) also yields information on L. A method for estimating L based on (22) is available in [19].

2) Channel Subspace (CS) Method: The CS method reformulates the same system (2) as follows:

$$\mathbf{y}(n) = \mathbf{H}_M \mathbf{s}(n) \tag{25}$$

where $n = 0, 1, \cdots, N-W$ and W is an integer "window parameter." This window parameter determines the length of each output vector $\mathbf{y}(n)$ as follows:

$$\mathbf{y}(n) = [\mathbf{y}_1^T(n), \cdots, \mathbf{y}_M^T(n)]^T \tag{26}$$

with

$$\mathbf{y}_i(n) = [y_i(n), \cdots, y_i(n+W-1)]^T. \tag{27}$$

The input vector sequence $\mathbf{s}(n)$ accordingly is defined by

$$\mathbf{s}(n) = [s(n-L), \cdots, s(n+W-1)]^T. \tag{28}$$

$\mathbf{H}_M$ is a generalized Sylvester matrix with the dimension $MW \times (W+L)$, as defined in (7) and (8), except for the dimensions of matrixes. The previous formulation (5) treats the system outputs as a large single vector, while the current formulation (25) treats the system outputs as a sequence of small vectors. The following theorem provides the foundation for the CS method.

Theorem 2 [15]: Define another $MW \times (W+L)$ generalized Sylvester matrix $\mathbf{H}'_M$ corresponding to another impulse response vector $\mathbf{h}'$. If $W \geq L+1$ and the M channels do not share a common zero, then the relations

$$\text{range } (\mathbf{H}'_M) \subset \text{range } (\mathbf{H}_M)$$

and

$$\text{range } (\mathbf{H}_M) \subset \text{range } (\mathbf{H}'_M)$$

hold if and only if $\mathbf{h}'$ is proportional to $\mathbf{h}$.

This theorem implies that under the specified conditions, the range space or, equivalently, the null space of $\mathbf{H}_M$ determines the channel impulse response uniquely up to a scalar. To compute the range or null space of $\mathbf{H}_M$, the CS method constructs the following covariance matrix:

$$\mathbf{R}_y = \frac{1}{N-W+1} \sum_{n=0}^{N-W} \mathbf{y}(n)\mathbf{y}^H(n). \tag{29}$$

This matrix has the following structure:

$$\mathbf{R}_y = \mathbf{H}_M \mathbf{R}_s \mathbf{H}_M^H \tag{30}$$

where

$$\mathbf{R}_s = \frac{1}{N-W+1} \sum_{n=0}^{N-W} \mathbf{s}(n)\mathbf{s}^H(n). \tag{31}$$

It is clear that as long as $\mathbf{R}_s$ is full rank, the range or null space of $\mathbf{H}_M$ can be computed from the eigendecomposition of $\mathbf{R}_y$. Let the eigendecomposition of $\mathbf{R}_y$ be

$$\mathbf{R}_y = \mathbf{E}\mathbf{\Lambda}\mathbf{E}^H \tag{32}$$

where $\mathbf{E}$ and $\mathbf{\Lambda}$ are matrixes of eigenvectors and eigenvalues, i.e.,

$$\mathbf{E} = [\mathbf{e}_1, \cdots, \mathbf{e}_{MW}]$$
$$\mathbf{\Lambda} = \text{diag } (\lambda_1, \cdots, \lambda_{MW}).$$

It can be shown that under the sufficient conditions of identifiability and the choice of window parameter $W = L+1$ (or, more generally, if the window parameter satisfies $(N-L+1)/2 \geq W \geq L+1$, the input has $L+W$ or more modes, and the channels do not share a common zero), the eigenvalues can be arranged as

$$\lambda_1 \geq \cdots \geq \lambda_{L+W} > \lambda_{L+W+1} = \cdots = \lambda_{MW} = 0$$

and, therefore

$$\text{range } ([\mathbf{e}_1, \cdots, \mathbf{e}_{L+W}]) = \text{range } (\mathbf{H}_M) \perp \text{range } ([\mathbf{e}_{L+W+1}, \cdots, \mathbf{e}_{MW}]).$$

Using this relationship and the previous theorem, we know that there is a unique $\mathbf{h}$ up to a scalar such that

$$\mathbf{E}_n^H \mathbf{H}_M = 0 \tag{33}$$

where

$$\mathbf{E}_n = [\mathbf{e}_{L+W+1}, \cdots, \mathbf{e}_{MW}].$$

Computing the solution to (33) yields the desired estimation. To cope with noise on the system outputs, the CS method computes the least-square error (LSE) solution to (33), i.e.,

$$\mathbf{h}_{\text{CS}} = \arg \min_{||\mathbf{h}||=1} ||\mathbf{E}_n^H \mathbf{H}_M||^2.$$

Note that the estimate $\mathbf{h}_{\text{CS}}$ is also a consistent estimator, i.e., $\mathbf{h}_{\text{CS}}$ is equal to the exact $\mathbf{h}$ up to a scalar if the additive noise on the channel outputs is white and the number of output samples is infinite. The performance of the CS method can be nearly as good as the ML method unless the channels have near common zeros [81]. The computation of the CS method is also on the order of N if N is very large (i.e., dominating M, L, and W). Further results and developments on the CS method can be found in [78]–[83].

D. Direct Estimation of the Input

For the aforementioned methods, the focus has been on the estimation of the system response instead of the input. Given the system response, the system input can be found by using (12). An alternative blind estimation scheme estimates the input directly from the multichannel system outputs [28]. The choice between direct system function estimation and direct input sequence estimation depends on the application. For applications such as seismology, where the system function carries the desired information, the direct system function estimation is preferred. For applications such as communication, where the input carries the desired information, the direct input estimation may have more advantages.

We now introduce three techniques for directly estimating the input sequence $s(k)$, $k = -L, \cdots, N-1$. They are the input subspace (IS) method, mutually referenced equalizers (MRE) method, and linear prediction (LP) method.

1) IS Method: Consider the noise-free data matrix given by

$$[\mathbf{y}(0), \cdots, \mathbf{y}(N-W)] = \mathbf{H}_M \mathbf{S}_{W+L}$$

where

$$\mathbf{S}_{W+L} = \begin{bmatrix} s(-L) & s(-L+1) & \cdots & s(N-W-L) \\ s(-L+1) & s(-L+2) & \cdots & s(N-W-L+1) \\ \vdots & \vdots & & \vdots \\ s(W-1) & s(W) & \cdots & s(N-1) \end{bmatrix}.$$

$\mathbf{S}_{W+L}$ is a Hankel matrix with the dimension $(W+L) \times (N-W+1)$. The subspace defined by the rows of $\mathbf{S}_{W+L}$ is now called the *input subspace*.

It is easy to verify that if $\mathbf{H}_M$ has full column rank,[2] the data matrix has the same row span as $\mathbf{S}_{W+L}$. Let $\mathbf{V}_o$ be the orthogonal matrix to the row space of $\mathbf{S}_{W+L}$, i.e., $\mathbf{S}_{W+L}\mathbf{V}_o = 0$. Then, using a property of Hankel matrixes [28]–[30], it is shown that given the null space of $\mathbf{S}_{W+L}$, the null space of $\mathbf{S}_{W+L-1}$ can be constructed as

$$\begin{bmatrix} \mathbf{V}_o & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_o \end{bmatrix}$$

where $\mathbf{0}$ is a $1 \times (N-2W-L+1)$ vector of zeros. Note that in the above matrix, the upper and lower $\mathbf{V}_o$ overlap in rows. More generally, the null spaces of $\mathbf{S}_r$, $r = W+L-k+1$, $k = 1, \cdots, W+L$ denoted by $\mathbf{V}(r)$ have the following form:

$$\mathbf{V}(r) = \underbrace{\begin{bmatrix} \mathbf{V}_o & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{V}_o \end{bmatrix}}_{k \text{ blocks}}. \tag{34}$$

This yields, in particular, the following overdetermined linear equations:

$$\mathbf{S}_1 \mathbf{V}(1) = 0 \tag{35}$$

where

$$\mathbf{S}_1 = [s(-L), \cdots, s(N-1)]$$

which uniquely determines the input sequence $\mathbf{S}_1$ under some mild assumption, as shown in Theorem 3.

[2]This requires $MW \geq L+W$, so that $\mathbf{H}_M$ is vertical or square. In the special case where $W \geq L$, this condition is shown to be equivalent to the absence of common zeros between the scalar polynomials $H_i(z)$; see [15].

Theorem 3 [28]–[30]: Assume that there exists W such that $\mathbf{H}_M$ is full column rank and that the input sequence $\{s(n)\}_{-L\le n\le N-1}$ contains more than $W+L+1$ modes,[3] then $\mathbf{S}_1$ up to a scalar ambiguity is the unique nontrivial solution of the overdetermined linear equations (35).

In the presence of additive noise, the LSE to (35) is computed. The IS algorithm can be summarized as follows.

- Calculate the null space of $\mathbf{S}_{W+L}$ from the data matrix.
- Construct $\mathbf{V}(1)$ as in (34).
- Estimate $\mathbf{S}_1$ by minimizing the quadratic criterion, i.e., $\arg\ \min_{\mathbf{s}} ||\mathbf{s}\mathbf{V}(1)||^2$.

An efficient implementation of the IS algorithm is available in [31].

2) MRE Method: This method, proposed in [32]–[34], estimates directly the channel inverse or rather a set of zero-forcing channel inverses called MRE. The existence of zero-forcing equalizers relies on the left-inversibility of the system matrix $\mathbf{H}_M$, which is ensured when $\mathbf{H}_M$ is full column rank.

The simple idea of the MRE evolves as follows. Assume two $MW\times 1$ multichannel equalizers $\mathbf{g}_i$ and $\mathbf{g}_{i+1}$, which satisfy

$$\begin{aligned}\mathbf{g}_i^H\mathbf{y}(n)&=s(n-i)\\ \mathbf{g}_{i+1}^H\mathbf{y}(n)&=s(n-i-1)\end{aligned}\tag{36}$$

where $\mathbf{y}(n)$ refers to the noise-free output vector defined in (26). Then, an obvious assertion is

$$\mathbf{g}_i^H\mathbf{y}(n)=\mathbf{g}_{i+1}^H\mathbf{y}(n+1),\qquad i=0,\cdots,W+L-2\tag{37}$$

where the outputs of the i-delay and $(i+1)$-delay equalizers are said to be referenced to each other, up to a sample duration delay.

It is shown in [32] that the $W+L-1$ relations given in (37) are sufficient for $\mathbf{g}_0,\cdots,\ \mathbf{g}_{W+L-1}$ to be ideal zero-forcing equalizers in the noiseless case. We have the following theorem.

Theorem 4 [32]: Let W be a positive integer such that $\mathbf{H}_M$ and $\mathbf{S}_{L+W}$ are full column rank and full row rank, respectively. Let $\mathbf{g}_0,\cdots,\mathbf{g}_{W+L-1}$ be complex-valued vectors of size MW. Assume that the independent relationships $\mathbf{g}_i^H\mathbf{y}(n)=\mathbf{g}_{i+1}^H\mathbf{y}(n+1)$ for all n and i are satisfied. Rewrite the vectors in matrix form as $\mathbf{G}=[\mathbf{g}_0,\cdots,\mathbf{g}_{W+L-1}]$. Then we have

$$\mathbf{G}^H\mathbf{H}_M=\alpha\mathbf{I}\ \text{for some complex number}\ \alpha.\tag{38}$$

In practice, the matrix $\mathbf{G}$ is estimated by minimizing the quadratic criterion

$$J(\mathbf{G})=\sum_{n=0}^{N-W}||\mathbf{E}(n)||^2$$

where

$$\mathbf{E}(n)=[\mathbf{I},\mathbf{0}]\mathbf{G}^H\mathbf{y}(n)-[\mathbf{0},\mathbf{I}]\mathbf{G}^H\mathbf{y}(n+1)$$

under a suitable constraint, e.g., Trace $(\mathbf{G}^H\mathbf{G})=1$. $\mathbf{I}$ and $\mathbf{0}$ are the $(W+L-1)\times(W+L-1)$ identity matrix and the $(W+L-1)\times 1$ null vector, respectively. Note that the role of the constraint is to avoid not only the trivial solution $\mathbf{G}=0$ but also other nonzero blocking matrixes, i.e., matrixes such that $\mathbf{G}^H\mathbf{H}_M=0$, which correspond to $\alpha=0$ in (38). In [32] and [34], linear or quadratic constraints are considered to keep the minimization procedure as simple as possible. Different choices for the constraint lead to different equalization performances and also different implementation issues [32], [34].

3) LP Method: For this method, we have to assume that the inputs are zero-mean temporally decorrelated (white) random variables. This requirement seems to hold well in data communications where pseudo-random noise is used to code symbols. Also, as for all the previous methods, we assume that Z-transform polynomials $H_i(z)$ do not share common zeros.

For simplicity, consider first the noise-free observation

$$\mathbf{y}(n)=[\mathrm{H}(z)]s(n),\ \mathrm{H}(z)\stackrel{\text{def}}{=}[H_1(z),\cdots,H_M(z)]^T$$

where $\mathbf{y}(n)$ corresponds to the $M\times 1$ vector defined in (26) for $W=1$ and $[\mathbf{H}(z)]s(n)$ denotes the output of the filter $\mathrm{H}(z)$ excited by $s(n)$. The basic idea behind the LP approach is to recognize that the moving-average (MA) process $\mathbf{y}(n)=[\mathrm{H}(z)]s(n)$ is also a *finite order* autoregressive (AR) process. This property is related to the generalized Bezout identity (see, for example, [21]). It is known [21] that since $H_i(z)$ are coprime, there exists a $1\times M$ polynomial vector $G(z)$ such that

$$\mathrm{G}(z)\mathrm{H}(z)=1.\tag{39}$$

By applying $\mathrm{G}(z)$ to $\mathbf{y}(n)$, it follows that

$$[\mathrm{G}(z)]\mathbf{y}(n)=s(n).\tag{40}$$

$\mathrm{H}(z)$ can be *exactly* inverted by a FIR causal filter. This relation is the key behind all subsequent derivations.

In order to proceed, some additional notations and definitions are in order. Denote by $\mathcal{H}_{n-1}(y)$ the span of the past of $\mathbf{y}$ up to time $n-1$, i.e.

$$\mathcal{H}_{n-1}(y)=\text{span}\ \{y_i(n-l)/i=1,\cdots,M;\ l\ge 1\}.\tag{41}$$

Here, span $\{x_i\in I\}$ stands for the Hilbert subspace generated by $\{x_i\in I\}$. Accordingly, denote by $\mathcal{H}_{n-1,P}(y)$ the span of the finite past of $\mathbf{y}$, i.e.,

$$\begin{aligned}&\mathcal{H}_{n-1,P}(y)\\ &\quad=\text{span}\ \{y_i(n-l)/i=1,\cdots,M;\ 1\le l\le P\}.\end{aligned}\tag{42}$$

The innovation process [35] $[\mathbf{i}(n)]_{n\in\mathbb{Z}}$ of $\mathbf{y}$ is the M-variate white noise sequence defined by

$$\mathbf{i}(n)=\mathbf{y}(n)-\mathbf{y}(n)|_{\mathcal{H}_{n-1}(y)}\tag{43}$$

where $|$ stands for the orthogonal projection onto $\mathcal{H}$. The term $\mathbf{y}(n)|_{\mathcal{H}_{n-1}(y)}$ is, in fact, the LP of $y(n)$. The process $\mathbf{y}$ will be said to be AR of order P if $\mathbf{i}(n)$ coincides with the finite order innovation sequence $\mathbf{i}_P(n)=\mathbf{y}(n)-\mathbf{y}(n)|_{\mathcal{H}_{n-1,P}(y)}$. We have the following theorem.

[3] It is equivalent to assume that $\mathbf{S}_{W+L+1}$ is full row rank; see [19] for more details.

Theorem 5 [36], [84]: $\mathbf{y}(n)$ is an order L AR process. Its innovation process is given by $\mathbf{i}(n) = \mathbf{h}(0)s(n)$, $\mathbf{h}(0) \stackrel{\text{def}}{=} [h_1(0), \cdots, h_M(0)]^T$.

The innovation $\mathbf{i}(n)$ is computed by projecting $\mathbf{y}(n)$ onto the space generated by the random variables $\{y_i(n-l)/i = 1, M, l = 1, \cdots, P\}$ where $P \geq L$. Let $[\mathbf{A}(1), \cdots, \mathbf{A}(P)]$ be $M \times M$ LP coefficient matrixes such that $\mathbf{y}(n) + \sum_{k=1}^{P} \mathbf{A}(k)\mathbf{y}(n-k) = \mathbf{i}(n)$. Then it is easy to show that $\mathbf{A}(1), \cdots, \mathbf{A}(P)$ satisfy the following:

$$[\mathbf{A}(1), \cdots, \mathbf{A}(P)]\mathbf{R} = -[\mathbf{r}(1), \cdots, \mathbf{r}(P)] \quad \text{(Yule–Walker equation)} \tag{44}$$

where $\mathbf{R}$ is the $MP \times MP$ matrix defined by

$$\mathbf{R} = \begin{bmatrix} \mathbf{r}(0) & \mathbf{r}(1) & \cdots & \mathbf{r}(P-1) \\ \mathbf{r}(-1) & \mathbf{r}(0) & \cdots & \mathbf{r}(P-2) \\ \vdots & \vdots & & \vdots \\ \mathbf{r}(1-P) & \mathbf{r}(2-P) & \cdots & \mathbf{r}(0) \end{bmatrix}$$

with $\mathbf{r}(k) \stackrel{\text{def}}{=} E\{\mathbf{y}(n+k)\mathbf{y}^H(n)\}$. Since $\mathbf{R}$ is not full rank [36], [84], the LP coefficients satisfying (44) are not unique. An LS solution to (44) is given by

$$[\mathbf{A}(1), \cdots, \mathbf{A}(P)] = -[\mathbf{r}(1), \cdots, \mathbf{r}(P)]\mathbf{R}^{\#}. \tag{45}$$

Due to Theorem 5, the covariance of the innovation $\mathbf{D} \stackrel{\text{def}}{=} E\{\mathbf{i}(n)\mathbf{i}^H(n)\}$ is given by $\mathbf{D} = \sigma_s^2\mathbf{h}(0)\mathbf{h}^H(0)$, where σ_s^2 is the source signal power. The nonzero eigenvalue of $\mathbf{D}$ is equal to $\lambda_d = \sigma_s^2\|\mathbf{h}(0)\|^2$ and the associated eigenvector $\mathbf{d}$ is $\mathbf{h}(0)/\|\mathbf{h}(0)\|$. Let $\mathbf{l} = \mathbf{d}/\sqrt{\lambda_d}$. We now have the following theorem.

Theorem 6 [36], [84]: $\mathrm{G}(z) = \mathbf{l}^H[\mathbf{I} + \sum_{k=1}^{P} \mathbf{A}(k)z^{-k}]$ is a $1 \times M$ polynomial vector such that $[\mathrm{G}(z)]\mathbf{y}(n) = s(n)$.

The LP method can now be summarized as follows.

- Estimate the covariance matrixes

$$\mathbf{r}(k) = \frac{1}{N-k}\sum_{i=0}^{N-k-1} \mathbf{y}(n+k)\mathbf{y}^H(n), \quad k = 0, \cdots, P \geq L$$

- Solve the Yule–Walker equation (44).
- Estimate the covariance of the innovation by $\mathbf{D} = \mathbf{r}(0) + \sum_{k=1}^{P} \mathbf{A}(k)\mathbf{r}^H(k)$ and estimate vector $\mathbf{l}$ from its eigendecomposition.
- Use the prediction filter $\mathrm{G}(z) = \mathbf{l}^H[\mathbf{I} + \sum_{k=1}^{P} \mathbf{A}(k)z^{-k}]$ to estimate the input $s(n)$ by $[\mathrm{G}(z)]\mathbf{y}(n)$.

In the presence of additive white noise, it suffices asymptotically to replace the covariance matrix $\mathbf{r}(0)$ by the noiseless autocovariance matrix $\mathbf{r}(0) - \sigma^2\mathbf{I}$, where σ^2 is the noise power.

IV. Other Existing Concepts and Techniques

BSI or blind deconvolution has been a well-established research topic for more than 20 years. The concepts and techniques presented in the previous section are only a sample of more recent developments in this area. All blind deconvolution techniques, however, can be grouped into two categories: one consisting of the SOS-based methods and the other consisting of the HOS-based methods. The SOS-based methods require a system model for which there are more than one output signals (unless certain constraint on the input is applied). It is on this model that the previous section was focused. In fact, all techniques shown in Section III are based on SOS.

For some applications, there is only a single output signal available. For the stationary single output system model, SOS's are well known to be insufficient to reveal all information (the phase function in particular) of the system unless source "training" samples are applied at the input or an additional property of the input is known. Therefore, higher order statistics are often necessary for the single output systems.

A. HOS-Based Methods

There are three subgroups in the category of HOS-based methods: hidden Markov model (HMM)-based methods, polyspectra methods, and Bussgang methods.

1) HMM Methods: The HMM methods are useful for situations where each input sample takes values from a finite alphabet. Assuming that the system transfer function is causal and of finite duration, it can be shown [38] that an ML solution for joint-channel-parameters estimation and input-symbol detection can be derived.

The Markovian properties of the channel state sequence enable us to maximize the required likelihood using the expectation-maximization (EM) algorithm [24], [39]. The main drawback of this method is the large memory requirement and large computational complexity inherent to the state Markov processes. Reference [40] proposes an on-line algorithm that significantly reduces the memory requirement but still is computationally expensive. Suboptimal methods that reduce the computational cost are proposed in [41] and [42]. Another weakness of the HMM methods is the possible convergence to local minima [43].

2) Polyspectra Methods: The polyspectra methods [20], [44]–[50] explicitly exploit the high-order spectra to first estimate the channel transfer function $H(z)$ and then estimate $\{s(n)\}$. Assuming that $s(n)$ has nonzero skewness, i.e., $E[s(n)^3] \neq 0$, then the bispectrum (third-order spectral cumulant) of $\{s(n)\}$ contains the phase information of $H(z)$. It is given by

$$B_y(f_1, f_2) = |B_y(f_1, f_2)| \exp j\Psi_y(f_1, f_2)$$

where

$$\Psi_y(f_1, f_2) = \phi_H(f_1) + \phi_H(f_2) - \phi_H(f_1 + f_2)$$

with

$$H(f) = |H(f)| \exp j\phi_H(f).$$

For communication signals, the skewness generally is equal to zero due to the symmetry of quadratic amplitude modulation constellations. In this case, we replace the bispectrum by the trispectrum (fourth-order spectral cumulant), and

similar phase relations can be exploited to extract the phase of the channel $H(f)$ [45], [51].

Earlier HOS approaches have been suggested by Brillinger and Rosenblatt [52], Wiggins [53], Donoho [54], Lii and Rosenblatt [55], etc. Following these pioneering works, various identification methods have been proposed to identify the transfer function of linear autoregressive moving average (ARMA) systems exploiting the cumulants of input and output signals [52]. Detailed studies of polyspectra methods are available in the tutorial papers [44], [46], [56], and [57].

3) Bussgang Methods: These methods do not explicitly use HOS but try to equalize the channel by minimizing a cost function that contains implicitly the information of higher order moments of the output. The inverse channel or the channel equalizer is a linear filter $G(z)$ applied to the channel output to yield an estimate of the input. The desired $G(z)$ must satisfy

$$G(z)H(z) = cz^{-m}, \qquad c \neq 0 \tag{46}$$

where the constant time delay m is an inherent indetermination in blind equalization due to the fact that a nonminimum phase channel in general will have an inverse that is noncausal [58].

A simple example of Bussgang algorithms is the decision-directed (DD) equalization [65], where the cost function is

$$J_{DD}(G) = E\{|\text{dec}\,[x(k)] - x(k)|^2\}$$

where

$$x(k) = [G(z)]y(k)$$

and the corresponding nonlinearity is dec $[x(k)]$, which denotes the closest constellation symbol to $x(k)$.

More generally, the equalizer parameters are adjusted via the minimization of some mean cost function

$$J(G) = E\{\mathcal{F}[x(k)]\}$$

where $\mathcal{F}$ is a properly chosen nonlinearity that reflects the level of ISI's in the equalizer output.

Bussgang algorithms generally are adaptive algorithms that update the equalizer parameters using a stochastic gradient descent scheme [59]. The error function is given by the derivative of the cost function $\mathcal{F}$.

Various Bussgang algorithms have been proposed in the literature, differing mainly by their selections of the cost function $\mathcal{F}$. Chronologically, the first well-known Bussgang algorithms were successively presented by Sato [60], Benvenist *et al.* [61], who extended Sato's work, Godard [62], Treichler *et al.* [63], [64], who developed the well-known constant-modulus algorithm, and Picchi *et al.* [65], who developed the stop-and-go algorithm. Further studies on the convergence properties, performance analysis, and fast implementations have been proposed in [66]–[74].

4) General Remark: For all HMM-based methods, the main drawback is the high computational cost (even for the suboptimal methods proposed in [40] and [42]), which cannot be accommodated in most on-line applications. Polyspectra and Bussgang methods also suffer from many weaknesses, such as the possible convergence to local minima [71], [75] and the sensitivity to timing jitter [76]. The main limitation of all HOS-based methods, however, consists of their slow convergence rate due to the large estimation variance of HOS and thus the need of a large sample size for accurate time-average approximations of HOS. Consequently, HOS-based methods cannot be applied in applications where fast channel variations and rapid adaptivity are essential.

B. SOS-Based Methods

Identification methods using only output SOS appear to be a recent breakthrough in the attempts to achieve fast blind channel identification and equalization. Since the recent works by Tong *et al.* [13] and Gardner [14], a plethora of SOS-based methods have been proposed in the literature. These methods include the matrix pair method [16], [77], cross-relations method [19], [27], subspace method [15], [78]–[80], [83], LP method [36], [84], [85], MRE method [32], [34], outer-product method [86], ML method [17], [25], frequency-domain methods [87]–[90], and input-subspace methods [28], [29], [31].

All of these methods use a multichannel observation model that arises from the exploitation of time or spatial diversity of the received signal. A detailed review of some of the SOS-based methods was given in Section III.

C. Multiple-Input Case

Last, another important system model where there are multiple inputs deserves mention. The multiple inputs may represent communication signals from multiple users or speech signals from multiple speakers. For such a system, one often needs to identify each input signal from the system's output.

The concepts and techniques developed for the multi-input system have clustered into two major groups: the instantaneous mixture case and the convolutive mixture case.

1) Instantaneous Mixture Case: In this case, the transfer function $H(z)$ is simply a constant mixing matrix. Blind deconvolution of instantaneous mixtures is known under different names, such as blind separation of sources, blind array processing, signal copy, independent component analysis, and waveform preserving estimation. The techniques for this problem depend on the *a priori* information on the source signals. These techniques include HOS methods [91]–[100] for non-Gaussian source signals, SOS methods [101]–[105] for temporally correlated sources, cyclostationary-based methods [106] for cyclostationary emitters, the mutual information method for independent sources [109], and ML or subspace fitting methods [107], [108] for finite alphabet sources. A detailed review of the

instantaneous mixture estimation techniques can be found in [110].

2) Convolutive Mixture Case: In this case, the transfer function $H(z)$ is a function of z. This case is more general than the previous one and arises from more applications. HOS-based techniques for this case are reported in [112]–[114]. A detailed review is given in [111].

A particular situation arises when the number of inputs is strictly less than the number of system outputs, i.e., $H(z)$ is a $q \times p$ matrix with $p < q$. As in the single-input case, SOS equalization[4] is possible under the fundamental assumption that the transfer function is irreducible, i.e., $H(z)$ is full column rank for each z. Therefore, identification methods of such systems generally proceed in two steps: 1) second-order identification up to a constant matrix (or channel equalization by removing a convolutional contribution) and 2) higher order identification of a mixing matrix (or extracting source signals by separation of instantaneous mixtures). These methods include the LP method [84], outer-product method [86], signal subspace method [29], and subspace method [78], [115].

V. Conclusion

BSI is a fundamental signal processing technology useful for a wide range of applications. The purpose of this paper has been to introduce readers to the technology of BSI and to promote the theory and applications of this technology. In this paper, we have presented some basic concepts and techniques of BSI and highlighted some of the recently developed SOS methods. Other major contributions in this field have been briefly discussed. The field of BSI is still developing rapidly, especially in the context of wireless communications.

References

[1] T. G. Stockham, Jr., T. M. Cannon, and R. B. Ingebretsen, "Blind deconvolution through digital signal processing," *Proc. IEEE*, vol. 63, pp. 678–692, Apr. 1975.

[2] J. Makhoul, "Linear prediction: A tutorial review," *Proc. IEEE*, vol. 63, pp. 561–580, Apr. 1975.

[3] R. W. Schafer and L. R. Rabiner, "Digital representations of speech signals," *Proc. IEEE*, vol. 63, pp. 662–677, Apr. 1975.

[4] L. C. Wood and S. Treitel, "Seismic signal processing," *Proc. IEEE*, vol. 63, pp. 649–661, Apr. 1975.

[5] B. R. Hunt, "Digital image processing," *Proc. IEEE*, vol. 63, pp. 693–708, Apr. 1975.

[6] S. Haykin, Ed., *Array Signal Processing.* Englewood Cliffs, NJ: Prentice-Hall, 1985.

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

[8] A. Acero, *Acoustical and Environmental Robustness in Automatic Speech Recognition.* Boston, MA: Kluwer, 1993.

[9] B. Juang, R. J. Perdue, Jr., and D. L. Thomson, "Deployable automatic speech recognition systems: Advances and challenges," *AT&T Tech. J.*, pp. 45–55, Mar./Apr. 1995.

[10] M. Z. Nashed, "Aspects of generalized inverses in analysis and regularization," in *Generalized Inverses and Applications*, M. Z. Nashed, Ed. New York: Academic, 1976.

[11] D. Kundur and D. Hatzinakos, "Blind image deconvolution," *IEEE Signal Processing Mag.*, vol. 13, pp. 43–64, May 1996.

[4] Equalization here means cancellation of the ISI. A further step of source separation is required in a second stage to fully identify the transfer function.

[12] Y. You and M. Kaveh, "A regularization approach to joint blur identification and image restoration," *IEEE Trans. Image Processing*, vol. 5, pp. 416–428, Mar. 1996.

[13] L. Tong, G. Xu, and T. Kailath, "A new approach to blind identification and equalization of multipath channels," in *Proc. 25th Asilomar Conf.*, Pacific Grove, CA, 1991, pp. 856–860.

[14] W. Gardner, "A new method of channel identification," *IEEE Trans. Commun.*, vol. 39, pp. 813–817, Aug. 1991.

[15] E. Moulines, P. Duhamel, J. Cardoso, and S. Mayrargue, "Subspace methods for the blind identification of multichannel FIR filters," *IEEE Trans. Signal Processing*, vol. 43, pp. 516–525, Feb. 1995.

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

[17] Y. Hua, "Fast maximum likelihood for blind identification of multiple FIR channels," *IEEE Trans. Signal Processing*, vol. 44, pp. 661–672, Mar. 1996.

[18] K. Abed-Meraim and Y. Hua, "Faster TSML algorithms for multichannel blind identification," *IEEE Trans. Signal Processing*, submitted for publication.

[19] 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.

[20] G. Giannakis and J. Mendel, "Identification of nonminimum phase systems using higher-order statistics," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 360–377, 1989.

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

[22] Y. Hua and M. Wax, "Strict identifiability of multiple FIR channels driven by an unknown arbitrary sequence," *IEEE Trans. Signal Processing*, vol. 44, pp. 756–759, Mar. 1996.

[23] K. Abed-Meraim and Y. Hua, "Strict Identifiability of SIMO systems: Further results and developments," in *Proc. Int. Conf. Telecommunications*, Melbourne, Australia, vol. 3, 1997, pp. 1029–1032.

[24] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via EM algorithm," *J. Roy. Statist. Soc. SER.*, vol. 39, pp. 1–38, 1977.

[25] K. Abed-Meraim and E. Moulines, "A maximum likelihood solution to blind identification of multichannel FIR filters," in *Proc. EUSIPCO*, Edinburgh, Scotland, vol. 2, 1994, pp. 1011–1014.

[26] D. Starer and A. Nehorai, "Passive localization of near-field sources by path following," *IEEE Trans Signal Processing*, vol. 42, pp. 677–680, Mar. 1994.

[27] H. Liu, G. Xu, and L. Tong, "A deterministic approach to blind identification of multichannel FIR systems," in *Proc. ICASSP*, Adelaide, Australia, vol. 4, 1994, pp. 581–584.

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

[29] A. J. Van der Veen, S. Talwar, and A. Paulraj, "Blind identification of FIR channels carrying multiple finite alphabet signals," in *Proc. ICASSP*, Detroit, MI, 1995, pp. 1213–1216.

[30] H. Liu and G. Xu, "Closed-form blind symbol estimation in digital communications," *IEEE Trans. Signal Processing*, vol. 45, pp. 2714–2723, Nov. 1995.

[31] A. J. 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. 47, pp. 173–190, Jan. 1997.

[32] D. Gesbert, P. Duhamel, and S. Mayrargue, "Subspace-based adaptive algorithms for the blind equalization of multichannel FIR filters," in *Proc. EUSIPCO*, Edinburgh, Scotland, 1994, pp. 712–715.

[33] G. B. Giannakis and C. Tepedelenlioglu, "Direct blind equalizers of multiple FIR channels: A deterministic approach," in *Proc. 30th Asilomar Conf.*, Pacific Grove, CA, 1996, pp. 290–294.

[34] D. Gesbert and P. Duhamel, "Unimodal blind adaptive channel equalization: An RLS implementation of the mutually referenced equalizers," in *Proc. SPAWC*, Paris, France, 1997, pp. 29–32.

[35] S. L. Marple, Jr., *Digital Spectral Analysis With Applications.* Englewood Cliffs, NJ: Prentice-Hall, 1987.

[36] K. Abed-Meraim, P. Duhamel, D. Gesbert, P. Loubaton, S.

Mayrargue, E. Moulines, and D. Slock, "Prediction error methods for time-domain blind identification of multichannel FIR filters," in *Proc. ICASSP*, Detroit, MI, vol. 3, 1995, pp. 1968–1971.
[37] J. Proakis, *Digital Communications.* New York: McGraw-Hill, 1989.
[38] M. Ghosh and C. L. Weber, "Maximum likelihood blind equalization," presented at the 1991 SPIE Conf., San Diego, CA, July 1991.
[39] D. M. Titterington, "Recursive parameter estimation using incomplete data," *J. Roy. Stat. Soc. B*, vol. 46, no. 2, pp. 257–267, 1984.
[40] V. Krishnamurthy and J. B. Moore, "On-line estimation of hidden Markov model parameters based on the Kullback–Leibler information measure," *IEEE Trans. Signal Processing*, vol. 41, pp. 1557–1572, Aug. 1993.
[41] M. Shao and C. Nikias, "An ML/MMSE estimation approach to blind equalization," in *Proc. ICASSP*, Adelaide, Australia, vol. 4, 1994, pp. 569–572.
[42] L. B. White, S. Perreau, and P. Duhamel, "Reduced computation blind equalization for FIR channel input Markov models," presented at the ICC, Seattle, WA, June 1995.
[43] C. F. J. Wu, "On the convergence properties of the EM algorithm," *Ann. Statist.*, vol. 11, no. 1, pp. 95–103, 1983.
[44] J. A. Cadzow, "Blind deconvolution via cumulant extrema," *Signal Processing Mag.*, vol. 13, no. 3, pp. 24–42, 1996.
[45] D. Hatzinakos and C. Nikias, "Blind equalization using a trispectrum based algorithm," *IEEE Trans. Commun.*, vol. 39, pp. 669–682, May 1991.
[46] J. 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.
[47] C. Nikias and J. Mendel, "Signal processing with higher-order spectra," *IEEE Signal Processing Mag.*, vol. 10, pp. 10–37, July 1993.
[48] O. Shalvi and E. Weinstein, "New criteria for blind deconvolution of nonminimum phase systems (channels)," *IEEE Trans. Inform. Theory*, vol. 36, pp. 312–321, Mar. 1990.
[49] J. K. Tugnait, "Identification of nonminimum phase linear stochastic systems," *Automatica*, vol. 22, pp. 454–464, 1986.
[50] J. A. Cadzow and X. Li, "Blind deconvolution," *Digital Signal Processing*, vol. 5, pp. 3–20, 1995.
[51] C. Nikias and M. Raghuveer, "Bispectrum estimation: A digital signal processing framework," *Proc. IEEE*, vol. 75, pp. 869–891, July 1988.
[52] D. R. Brillinger and M. Rosenblatt, "Computation and interpretation of k-th order spectra," in *Spectral Analysis of Time Series*, B. Harris, Ed. New York: Wiley, 1967.
[53] R. A. Wiggins, "Minimum entropy deconvolution," *Geoexploration*, vol. 16, pp. 21–35, 1978.
[54] D. Donoho, "On minimum entropy deconvolution," in *Applied Time-Series Analysis II.* New York: Academic, 1981, pp. 565–609.
[55] K. S. Lii and M. Rosenblatt, "Deconvolution and estimation of transfer function phase and coefficients for non-Gaussian linear processes," *Ann. Statistics*, vol. 10, pp. 1195–1208, 1982.
[56] G. B. Giannakis and A. Swami, "On estimating noncausal nonminimum phase ARMA models of non-Gaussian processes," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, pp. 478–495, 1990.
[57] J. L. Lacoume, Ed., *Proc. 2nd Int. Conf. Higher-Order Statistics*. The Netherlands: Elsevier, 1992.
[58] A. Benveniste and M. Goursat, "Blind equalizers," *IEEE Trans. Commun.*, vol. 32, pp. 871–883, Aug. 1983.
[59] S. Haykin, *Adaptive Filter Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1986.
[60] Y. Sato, "A method of self-recovering equalization for multilevel amplitude modulation," *IEEE Trans. Commun.*, vol. 6, pp. 679–682, 1975.
[61] A. Benveniste, M. Goursat, and G. Ruget, "Robust identification of a nonminimum phase system: Blind adjustment of a linear equalizer in data communication," *IEEE Trans. Automat. Contr.*, vol. 25, pp. 385–399, Mar. 1980.
[62] D. Godard, "Self-recovering equalization and carrier-tracking in two-dimensional data communication systems," *IEEE Trans. Commun.*, vol. 28, pp. 1867–1875, Nov. 1980.
[63] J. R. Treichler and B. G. Agee, "A new approach to multipath correction of constant modulus signals," *IEEE Trans. Signal Processing*, vol. 31, pp. 349–372, Apr. 1983.
[64] J. R. Treichler and M. G. Larimore, "New processing techniques based on the constant modulus adaptive algorithm," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 33, pp. 420–431, Apr. 1985.
[65] G. Picchi and G. Prati, "Blind equalization and carrier recovery using a 'stop-and-go' decision-directed algorithm," *IEEE Trans. Commun*, vol. 35, pp. 877–887, Sept. 1987.
[66] O. Macchi and E. Eweda, "Second order convergence analysis of stochastic adaptive linear filtering," *IEEE Trans. Automat. Contr.*, vol. 28, pp. 76–85, Jan. 1983.
[67] ——, "Convergence analysis of self-adaptive equalizers," *IEEE Trans. Inform. Theory*, vol. 30, pp. 161–176, Mar. 1984.
[68] R. R. Bitmead, "Convergence properties of LMS adaptive estimators with unbounded dependent inputs," *IEEE Trans. Automat. Contr.*, vol. 29, pp. 477–479, May 1984.
[69] S. Bellini, "Blind equalization," *Alta Frequenza*, vol. LVII-N.7, pp. 445–450, Sept. 1988.
[70] R. A. Kennedy, B. D. O. Anderson, Z. Ding, and C. R. Johnson, Jr., "On the failure of proposed blind recursive identification schemes based on the Sato algorithm and its generalizations," in *Proc. Int. Symp. Signal Processing and its Applications*, 1990, pp. 267–270.
[71] Z. Ding, R. Kennedy, B. Anderson, and C. Johnson, "Ill-convergence of Godard blind equalizers in data communication systems," *IEEE Trans. Commun.*, vol. 39, pp. 1313–1327, Sept. 1991.
[72] Z. Ding, C. R. Johnson, and R. A. Kennedy, "On the admissibility of blind adaptive equalizers," in *Proc. ICASSP*, Albuquerque, NM, 1990, pp. 1707–1710.
[73] D. T. M. Slock, "On the convergence behavior of the LMS and the normalized LMS algorithms," *IEEE Trans. Signal Processing*, vol. 41, pp. 2811–2825, Sept. 1993.
[74] J. Benesty and P. Duhamel, "A fast exact LMS adaptive algorithm," *IEEE Trans. Signal Processing*, vol. 40, pp. 2904–2920, Dec. 1992.
[75] S. Haykin, *Blind Deconvolution.* Englewood Cliffs, NJ: Prentice-Hall, 1994.
[76] S. U. H. Qureshi, "Adaptive equalization," *Proc. IEEE*, vol. 73, pp. 1349–1387, Sept. 1985.
[77] Y. Hua, H. Yang, and W. Qiu, "Source correlation compensation for blind channel identification based on second-order statistics," *IEEE Signal Processing Lett.*, vol. 1, pp. 119–120, Aug. 1994.
[78] 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.
[79] K. Abed-Meraim and Y. Hua, "Blind equalization of multichannel IIR system using second-order statistics," in *Proc. IEEE Region Ten Conference (TENCON)*, Perth, Australia, vol. 1, 1996, pp. 446–451.
[80] ——, "Blind identification of multi-input multi-output system using minimum noise subspace," *IEEE Trans. Signal Processing*, vol. 45, pp. 254–257, Jan. 1997.
[81] W. Qiu and Y. Hua, "Performance analysis of the subspace method for blind channel identification," *Signal Process.*, vol. 50, pp. 71–82, Apr. 1996.
[82] K. Abed-Meraim, J. F. 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.
[83] Y. Hua, K. Abed-Meraim, and M. Wax, "Blind system identification using minimum noise subspace," in *Proc. 8th IEEE Workshop on ISSAP*, Corfu, Greece, June 1996, pp. 308–311.
[84] K. Abed-Meraim, E. Moulines, and P. Loubaton, "Prediction error method for second-order blind identification: Algorithms and statistical performance," *IEEE Trans. Signal Processing*, vol. 45, pp. 694–705, Mar. 1997.
[85] D. Slock, "Blind fractionally-spaced equalization, perfect-reconstruction filter-banks and multichannel linear prediction," in *Proc. ICASSP*, Adelaide, Australia, vol. 4, 1994, pp. 585–588.
[86] Z. Ding, "An outer product decomposition algorithm for multichannel blind identification," in *Proc. 8th IEEE Workshop on ISSAP*, Corfu, Greece, June 1996, pp. 132–135.
[87] Y. Li and Z. Ding, "Blind channel identification based on second-order cyclostationary statistics," in *Proc. ICASSP*, vol. 4, 1993, pp. 81–84.
[88] G. Giannakis, "A linear cyclic correlation approach for blind

identification of FIR channels," in *Proc. 28th Asilomar Conf.*, Pacific Grove, CA, 1994, pp. 420–424.
[89] M. Giannakis and S. Halford, "Performance analysis of blind equalizers based on cyclostationary statistics," in *Proc. Conf. Inform. Sci. and Syst.*, 1994, pp. 873–876.
[90] K. Abed-Meraim and Y. Hua, "Combining cyclostationarity and spatial diversity for blind system identification," in *Proc. 30th Asilomar Conf.*, Pacific Grove, CA, 1996, pp. 325–328.
[91] P. Comon, "Independent component analysis: A new concept," *Signal Process.*, vol. 36, pp. 287–314, 1994.
[92] X. R. Cao and R. Liu, "General approach to blind source separation," *IEEE Trans. Signal Processing*, vol. 44, Mar. 1996.
[93] C. Jutten and J. Hérault, "Blind separation of sources: An adaptive algorithm based on neuromimetic architecture," *Signal Process.*, vol. 24, pp. 1–10, 1991.
[94] D. T. Pham, P. Garrat, and C. Jutten, "Separation of a mixture of independent sources through a maximum likelihood approach," in *Proc. EUSIPCO*, Brussels, Belgium, 1992, pp. 771–774.
[95] J.-L. Lacoume and P. Ruiz, "Sources identification: A solution based on cumulants," in *Proc. IEEE ACSP Workshop V*, Minneapolis, MN, 1988.
[96] J. Cardoso, "Source separation using higher-order moments," in *Proc. ICASSP*, Glasgow, Scotland, 1989, pp. 2109–2112.
[97] ——, "Eigen-structure of the fourth-order cumulant tensor with application to the blind source separation problem," in *Proc. ICASSP*, Albuquerque, NM, 1990, pp. 2655–2658.
[98] J.-F. Cardoso and A. Souloumiac, "Blind beamforming for non Gaussian signals," *Proc. Inst. Elect. Eng.*, vol. 140, pt. F, no. 6, pp. 362–370, 1993.
[99] J.-F. Cardoso and B. Laheld, "Equivalent adaptive source separation," *IEEE Trans. Signal Processing*, vol. 44, pp. 3017–3030, Dec. 1996.
[100] A. Belouchrani and K. Abed-Meraim, "Constant modulus blind source separation technique: A new approach," in *Proc. Int. Symp. Signal Processing and Its Applications*, Gold Coast, Australia, vol. 1, 1996, pp. 232–235.
[101] A. Belouchrani, K. Abed-Meraim, J. F. Cardoso, and E. Moulines, "Second-order blind separation of correlated sources," in *Proc. Int. Conf. Digital Signal Processing*, Cyprus, 1993, pp. 346–351.
[102] K. Abed-Meraim, A. Belouchrani, J.-F. Cardoso, and E. Moulines, "Asymptotic performance of second-order blind separation," in *Proc. ICASSP*, Adelaide, Australia, vol. 4, 1994, pp. 277–280.
[103] A. Belouchrani, K. Abed-Meraim, J. F. Cardoso, and E. Moulines, "Blind source separation using second-order statistics," *IEEE Trans. Signal Processing*, submitted for publication.
[104] L. Tong and R. Liu, "Blind estimation of correlated source signals," in *Proc. Asilomar Conf.*, Pacific Grove, CA, Nov. 1990.
[105] L. Tong, R. Liu, V. C. Soon, and Y. F. Huang, "Indeterminacy and identifiability of blind identification," *IEEE Trans. Circuits Syst.*, vol. 38, pp. 499–509, May 1991.
[106] B. G. Agee, S. V. Schell, and W. A. Gardner, "Spectral self-coherence restoral: A new approach to blind adaptive signal extraction using antenna arrays," *Proc. IEEE*, vol. 78, pp. 753–766, Apr. 1990.
[107] A. Belouchrani and J.-F. Cardoso, "Maximum likelihood source separation for discrete sources," in *Proc. EUSIPCO*, vol. 2, Edinburgh, Scotland, 1994, pp. 768–771.
[108] 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.
[109] S. Amari, A. Cichocki, and H. Yang, "Recurrent neural networks for blind separation of sources," in *Proc. Int. Symp. Nonlinear Theory and Applications*, Dec. 1995, pp. 37–42.
[110] K. J. Pope and R. E. Bogner, "Blind signal separation I–Linear, instantaneous combinations," *Digital Signal Processing*, vol. 6, no. 1, pp. 5–16, 1996.
[111] K. J. Pope, "Blind signal separation II–Linear, convolutive combinations," *Digital Signal Processing*, vol. 6, no. 1, pp. 17–28, 1996.
[112] G. G. A. Swami and S. Shamsunder, "Multichannel ARMA processes," *IEEE Trans. Signal Processing*, vol. 42, pp. 898–913, 1994.
[113] G. Giannakis, Y. Inouye, and J. Mendel, "Cumulant-based identification of multichannel moving average models," *IEEE Trans. Automat. Contr.*, vol. 34, pp. 783–787, 1989.
[114] L. Tong, Y. Inouye, and R. Liu, "A finite-step global convergence algorithm for the cumulant-based parameter estimation of multichannel moving average processes," in *Proc. ICASSP*, Toronto, Ont., Canada, 1991, pp. 3445–3448.
[115] A. Gorokhov and P. Loubaton, "Subspace based techniques for second-order blind separation of convolutive mixtures with temporally correlated sources," *IEEE Trans. Circuits Syst.*, submitted for publication.
[116] G. Harikumar and Y. Bresler, "Blind restoration of images blurred by multiple filters: Theory and efficient algorithms," *IEEE Trans. Image Processing*, submitted for publication.
[117] ——, "Exact image deconvolution from multiple FIR blurs," *IEEE Trans. Image Processing*, submitted for publication.

Karim Abed-Meraim was born in Algiers, Algeria, in 1967. He received the M.S. degree from Ecole Polytechnique in 1990 and the Ph.D. degree in signal processing from Telecom-Paris in 1995.

He currently is a Research Fellow with the Electrical Engineering Department, Melbourne University, Australia. His research interests are in statistical signal processing and include system identification, (blind) array processing, and performance analysis.

Wanzhi Qiu (Member, IEEE) was born in Sichuan, China, in 1963. He received the B.E. and M.E. degrees in electronic engineering from the University of Electronic Science and Technology of China (former Chengdu Institute of Radio Engineering), Chengdu, Sichuan, China, in 1983 and 1989, respectively, and the Ph.D. degree in electronic engineering from the University of Melbourne, Australia, in 1996.

From 1983 to 1986, he was an Engineer at China Neihua Electronic Technology Research Institute, Sichuan, China. During 1989–1993, he was a Lecturer at Chengdu Institute of Meteorology, Chengdu, China. Since October 1996, he has been a Researcher at Motorola Australian Research Center. His research interests are in blind equalization, sensor array processing, and error control decoding.

Yingbo Hua (Senior Member, IEEE) was born in China in 1960. He received the B.S. degree from the Nanjing Institute of Technology (currently Southeast University), Nanjing, China, in February 1982 and the M.S. and Ph.D. degrees from Syracuse University, NY, in 1983 and 1988, respectively.

Since February 1990, he has been with the University of Melbourne, Australia. From 1990 to 1992, he was a Lecturer. From 1993 to 1995, he was a Senior Lecturer. Since 1996, he has been an Associate Professor and Reader. He has published more than 50 journal papers and 80 conference papers in the areas of spectral estimation, array processing, radar and NMR imaging, system identification, neural networks, and wireless communications. Since 1994, he has been an Associate Editor of IEEE TRANSACTIONS ON SIGNAL PROCESSING.

Dr. Hua is a member of a number of international technical committees.