ELSEVIER

Signal Processing 50 (1996) 83–99

SIGNAL PROCESSING

Recent developments in blind channel equalization: From cyclostationarity to subspaces

Hui Liu,[a,*] Guanghan Xu[b], Lang Tong[c], Thomas Kailath[d]

[a]*Department of Electrical Engineering, University of Virginia, Charlottesville, VA 22903-2442, USA*
[b]*Department of Electrical & Computer Engineering, The University of Texas at Austin, Austin, TX 78712-108, USA*
[c]*Department of Electrical & Systems Engineering, University of Connecticut, Storrs, CT 06269-3157, USA*
[d]*Information Systems Lab., Stanford University, Stanford, CA 94305-4055, USA*

Received 24 April 1995; revised 15 December 1995

Abstract

Since Tong et al. (1991) demonstrated the feasibility of identifying possibly nonminimum phase channels using second-order statistics, considerable research activity, both in algorithm development and fundamental analysis, has been seen in the area of blind identification of multiple FIR channels. Many of the recently developed approaches invoke, either explicitly or implicitly, the algebraic structure of the data model, while some others resort to the use of cyclic correlation/spectral fitting techniques. The objective of this paper is to establish insightful connections among these studies and present recent developments of blind channel equalization. We also unify various representative algorithms into a common theoretical framework.

Zusammenfassung

Seitdem Tong, Xu und Kailath (1991) die Machbarkeit der Identifizierung möglicherweise nichtminimalphasiger Kanäle unter Nutzung von "Second-Order Statistics" demonstrierten, waren erhebliche Forschungsaktivitäten sowohl in der Entwicklung von Algorithmen, als auch einer grundlegenden Analyse, auf dem Gebiet der "Blind-Identification" von multiplen FIR-Kanälen zu beobachten. Viele der kürzlich entwickelten Näherungen nehmen entweder explizit oder implizit die algebraische Struktur der Daten zu Hilfe, während einige andere Zuflucht bei der Anwendung von zyklischen "Correlation/Spectral Fitting"-Techniken suchen. Die Intention dieses Beitrages ist es, einsichtige Verbindungen zwischen diesen Untersuchungen aufzuzeigen und neuere Entwicklungen der "Blind-Channel"-Entzerrung zu präsentieren. Wir vereinheitlichen ebenfalls verschiedene repräsentative Algorithmen in einem gemeinsamen theoretischen Rahmen.

Résumé

Depuis que Tong, Xu et Kailath (1991) ont démontré la faisabilité de l'identification de canaux à phase non minimum à l'aide de statistiques de second ordre, une activité de recherche considérable, à la fois pour le développement d'algorithmes et l'analyse fondamentale, a vu le jour dans le domaine de l'identification aveugle de canaux FIR multiples.

*Corresponding author.

0165-1684/96/$15.00 © 1996 Elsevier Science B.V. All rights reserved
PII S0165-1684(96)00013-8

Beaucoup des approches récemment développées impliquent, soit explicitement soit implicitement, la structure algébrique du modèle des données, alors que d'autres font appel aux techniques de corrélation cyclique/ajustement spectral. L'objectif de cet article est d'établir des connexions pénétrantes parmi ces études et de présenter des développements récents en égalisation aveugle de canal. Nous unifions également les algorithmes représentatifs divers dans un cadre théorique commun.

Keywords: Blind equalization; Nonminimum phase channels; Fractionally spaced equalizers; Cyclic correlation

1. Introduction

Blind channel equalization (BCE) [38], i.e., determining and equalizing the channel response based solely on the channel output without the use of a training sequence, has received considerable attention recently in communications and signal processing. Earlier approaches to blind identification exploit the higher-order statistics of the output; see [5, 10, 52, 41, 20, 14] and the references therein. These methods, although reliable and robust in some scenarios, require a large number of data samples and a large amount of computation. In fast changing environments, such as in cellular communications, their applications may be limited. These problems are alleviated by a method proposed by Tong et al. [49, 51] which explored the cyclostational properties of an *oversampled* communication signal to allow the blind channel estimation to be accomplished based on second-order statistics of the channel output. This result is believed to have inspired all the subsequent development in identifying a single-input multiple-output FIR system without using higher-order statistics.

Since it is well-known that the second-order statistics of *scalar* system output do not contain enough information to identify a possibly nonminimum phase system, and since the original approach [49] employs the temporal oversampling technique which converts a stationary communication sequence into a cyclostationary process,[1] it was, for a while, believed that cyclostationarity was the *only* key to the surprising success of the original algorithm. Many early algorithms employ the cyclic correlation/spectral fitting techniques. Although it was later shown that blind identification can be accomplished without the use of any statistics, studies in that period brought to light many important aspects of the problem, e.g., the channel identifiability conditions [33, 11, 26].

Some of the early statistics-based methods, including the original approach, suffer from the performance degradation caused by the model mismatch when only a limited number of observations is available. The desire for more data-efficient algorithm led to the development of a class of subspace-based blind identification algorithms. The exploitation of a certain inherent subspace structure arising from the combination of a block-Toeplitz FIR channel matrix and a Hankel input matrix in the absence of noise, and allows these approaches to provide exact channel estimation using only a *finite* number of observations [32, 43, 30 14]. These techniques significantly outperform several previously developed statistics-based methods, especially for short data sequences. The success of these algorithms revealed an important fact: it is the intrinsic single-input multiple-output structure that is essential to blind identification.

Despite many promising features, the performance of these subspace-based algorithms, may be fundamentally limited by the nature of the channel. For example, singularity of the channel matrix can cause divergence of the subspace, and result in failure of the subspace approaches. This fact triggered some of the latest studies on algorithm performance, and saw a reemergence of correlation/spectral fitting techniques. Other currently active topics include performance analysis, channel order detection, fast algorithm development and implementation, and algorithm generalization to multiple inputs systems, etc.

[1] Although Gardner showed that the second-order statistics of cyclostationary signals do contain phase information which can be used in nonminimum phase system identification, the algorithm proposed in [13] requires the use of training sequences.

The purpose of this paper is to survey various blind equalization approaches, explore their connections, and highlight several important and somewhat surprising results. With the mushrooming of studies and the proliferation of algorithms, our main objective is not to give an encyclopedic coverage of all the blind estimation techniques, but rather to provide a systematic summary of recent development in the area of blind channel equalization.

The rest of this paper is organized as follows. Following the problem formulation in Section 2, we first review in Section 3 the original method [19] that marked the beginning of a brand new direction in blind channel identification, and then briefly summarize several other cyclostationarity-based algorithms. In Section 4, we highlight the channel identifiability conditions and discuss several important results concerning system identifiability. In Section 5, we present two subspace channel identification approaches, namely, the least-squares (LS) approach and the channel subspace (CS) approach, both of which exploit the system structure and offer superior estimation performance; a signal subspace (SS) approach and a deterministic maximum likelihood framework are then described. Finally in Section 6, we outline some of the latest results and trends in blind identification.

2. Data formulation

Fig. 1 depicts a discrete multichannel system with FIR impluse responses $\{h_i(\cdot)\}$ driven by the same input $s(\cdot)$. The system output $x_i(\cdot)$ is related to $s(\cdot)$ and $h_i(\cdot)$ by

$$x_i(k) = \sum_{j=0}^{L} h_i(j)\, s(k-j), \quad i = 1, \ldots, M, \tag{1}$$

where M is the number of channels, and $L+1$ is the maximum order of the M channels ($M \geqslant 2$ and $L \geqslant 1$). The principal interest herein is estimation of the channels $\{h_i(\cdot)\}$ and the input $s(\cdot)$ from the noise-corrupted outputs, without using higher-order statistics.

Such a problem arises in a variety of applications such as mobile communications, seismic signal analysis, and image restoration. A paradigmatic example concerns a wireless communication system with multiple receivers (antenna array); the propagation channel from the transmitter to each individual receiver can be practically modeled as an FIR filter.

In addition to such applications where *physical* receivers are employed, a multichannel system can also be formulated using *virtual* receivers. Fig. 2 illustrates a single receiver digital communication system where the channel lasts for 3 adjacent bauds. By temporally oversampling the channel

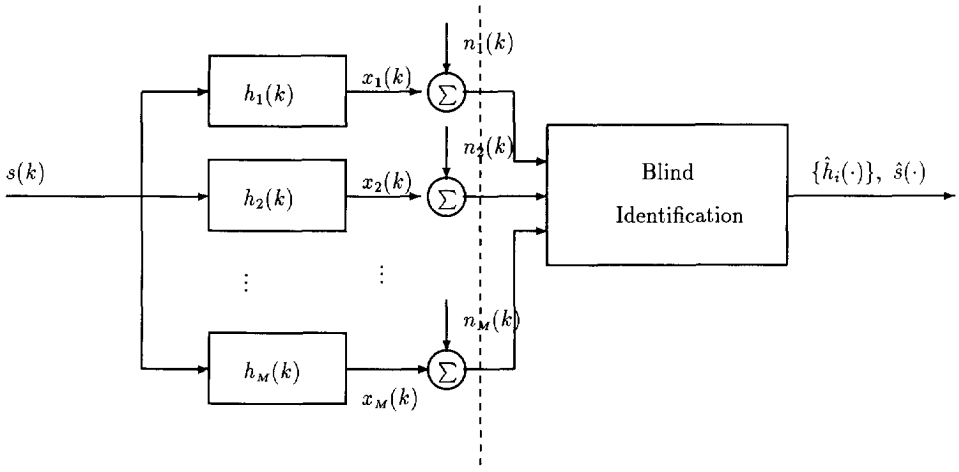

Fig. 1. Blind estimation for multiple channels.

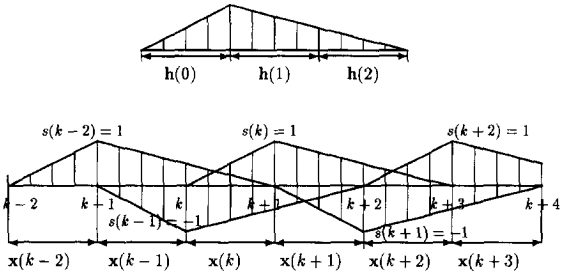

Fig. 2. A special channel with $L = 2$ and $s_k = \pm 1$.

outputs at 4 times the baud rate, and collecting the data samples within each baud period as $\boldsymbol{x}(k) = [x_0(k), x_1(k), x_3(k), x_4(k)]^{\mathrm{T}}$, we obtain

$$x_i(k) = h_i(2)s(k-2) + h_i(1)s(k-1) + h_i(0)s(k), \quad (2)$$

where $h_i(k)$ is the ith element of $\boldsymbol{h}(k)$, $k = 0, 1, 2$ and $i = 0, 1, 2, 3$. Clearly, (2) has the same form as (1) with $L = 2$.

Unless otherwise stated, we make the following assumptions regarding the channels and noise throughout this paper.

(A1) The noise is complex normal and white with spectral density σ_n^2.

(A2) All channels are linear time-invariant (LT1) and are of finite duration.

(A3) There is at least one channel of length L.

As a general notational convention, matrices (in capital letters) and vectors will be in boldface. The symbols $(\cdot)^{\mathrm{H}}$, $(\cdot)^{\mathrm{T}}$, $\odot$ and $\otimes$ stand for Hermitian, transpose, convolution and Kronecker product, respectively; $a(z) = a(0) + a(1)z + \cdots + a(p)z^p$ denotes a polynomial whose coefficients are the elements of a vector $\boldsymbol{a}$. $\mathrm{Re}(\theta)$, $\mathrm{Im}(\theta)$ and $\hat{\theta}$ are the real part, the imaginary part and the estimate of the quantity θ. The symbol $\boldsymbol{I}(\mathbf{0})$ stands for the identity (zero) matrix or vector with a proper dimension.

Using the above notation, the channel and its outputs can be represented in vector form: $\boldsymbol{x}(k) = [x_1(k) \ \ldots \ x_M(k)]^{\mathrm{T}}$, $\boldsymbol{h}_j = [h_1(j) \ \ldots \ h_M(j)]^{\mathrm{T}}$. The input–output relation in (1) can consequently be compactly rewritten as (see [18])

$$\boldsymbol{x}(k) = \sum_{j=0}^{L} \boldsymbol{h}_j s(k-j) = \boldsymbol{H}\boldsymbol{s}(k), \quad (3)$$

where $\boldsymbol{H} = [\boldsymbol{h}_L, \ldots, \boldsymbol{h}_0]$ and $\boldsymbol{s}(k) = [s(k-L), \ldots, s(k)]^{\mathrm{T}}$. We assume only a finite number of observations and index them from $L+1$ to N: $\boldsymbol{x}(L+1)$ to $\boldsymbol{x}(N)$; $L+1$ is used here as the initial index so that the input signals to be unraveled are $s(1), \ldots, s(N)$.

For later reference, we define a *Toeplitz block transform* (TBT), $\mathscr{T}(\cdot)$ and *Hankel block transform* (HBT), $\mathscr{H}(\cdot)$ on a given column vector sequence $z_1, z_2, \ldots, z_N$ as follows:

$$\mathscr{T}_K(z_1 \ \ldots \ z_N) = \underbrace{\begin{bmatrix} z_1 & z_2 & \cdots & z_N & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & z_1 & z_2 & \ldots & z_N & \ldots & \mathbf{0} \\ \vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \mathbf{0} & \ldots & \mathbf{0} & z_1 & z_2 & \cdots & z_N \end{bmatrix}}_{K+N-1 \text{ blocks}}; \quad (4)$$

$$\mathscr{H}_k(z_1 \ \ldots \ z_N) = \begin{bmatrix} z_1 & z_2 & \cdots & z_{N-K+1} \\ z_2 & z_3 & \cdots & z_{N-K+2} \\ \vdots & \vdots & \cdots & \vdots \\ z_K & z_{K+1} & \cdots & z_N \end{bmatrix}.$$

With the notation

$$\begin{aligned} &\boldsymbol{X}_i(K) = \mathscr{T}_K(x_i(L+1) \ \ldots \ x_i(N)); \\ &\boldsymbol{X}(K) = \mathscr{T}_K(x(L+1) \ \ldots \ x_N)); \\ &\boldsymbol{H}_i(K) = \mathscr{T}_K(h_i(L) \ \ldots \ \boldsymbol{h}_i(0)); \quad (5) \\ &\boldsymbol{H}(K) = \mathscr{T}_K(\boldsymbol{h}_L \ \ldots \ \boldsymbol{h}_0); \\ &\boldsymbol{S}(L+K) = \mathscr{H}_{L+K}(s_1 \ \ldots \ s_N), \end{aligned}$$

it can be easily verified that

$$\boldsymbol{X}(K) = \boldsymbol{H}(K)\boldsymbol{S}(L+K). \quad (6)$$

K here is defined as the *smoothing factor*. The explicit dependence of $\boldsymbol{X}(K)$, $\boldsymbol{H}(K)$ and $\boldsymbol{S}(L+K)$ on K may sometimes be dropped when there is no possibility of confusion. Also denote by $\boldsymbol{x}(k)$ and $\boldsymbol{s}(k)$ the kth column vectors of $\boldsymbol{X}(K)$ and $\boldsymbol{S}(K)$, respectively. When $K = 1$, they reduce to the output and input vectors in (3).

3. Blind identification using oversampling techniques

Classical solutions [38, 18, 4, 20] to blind identification in digital communication systems are based upon data sampled at the baud rate, although it has been known for some time [54, 17]

that fractionally spaced equalizers are more robust under timing uncertainties. Since communication channels are, in general, nonminimum phase, the second-order statistics of baud-rate-sampled stationary signals are inadequate for channel identification. Gardner [13] was perhaps the first to recognize that the phase information is available in the cyclostationary sequence; early work on blind deconvolution of cyclostationary signals with restrictive channels can be found in [7].

3.1. The original approach

The first *blind* channel estimation algorithm that *only* utilizes the second-order statistics was introduced by Tong et al. [49]. By employing the temporal oversampling technique and assuming that the system inputs $s(\cdot)$ are i.i.d., the autocorrelation matrix of the system outputs is given by

$$\boldsymbol{R}_x(0) = E\{x(k)x^{\mathrm{H}}(k)\} = \boldsymbol{H}\underbrace{E\{s(k)s^{\mathrm{H}}(k)\}}_{=\boldsymbol{I}}\boldsymbol{H}^{\mathrm{H}} = \boldsymbol{H}\boldsymbol{H}^{\mathrm{H}}. \tag{7}$$

Here, we assume that a proper smoothing factor K is selected such that $\boldsymbol{H}$ has more rows than columns. Although from the subspace generated by the columns of $\boldsymbol{R}_x(0)$, we can only obtain the $\boldsymbol{H}$ up to a unitary rotation, i.e., $\boldsymbol{F} = \boldsymbol{H}\boldsymbol{T}$, where $\boldsymbol{T}$ is a unitary matrix, a means for eliminating such an ambiguity or to find $\boldsymbol{T}$ is to introduce $\boldsymbol{R}_x(1)$ (see [49]):

$$\begin{aligned}\boldsymbol{R}_x(1) &= E\{x(k)x^{\mathrm{H}}(k-1)\}\\ &= \boldsymbol{H}\underbrace{E\{s(k)s^{\mathrm{H}}(k-1)\}}_{=\boldsymbol{J}}\boldsymbol{H}^{\mathrm{H}} = \boldsymbol{H}\boldsymbol{J}\boldsymbol{H}^{\mathrm{H}},\end{aligned} \tag{8}$$

where $\boldsymbol{J}$ is the *shifting* matrix with 0's everywhere except its second diagonal elements which are 1's. Since

$$\boldsymbol{F}^{\dagger} \triangleq (\boldsymbol{F}^{H}\boldsymbol{F})^{-1}\boldsymbol{F}^{\mathrm{H}} = \boldsymbol{T}^{\mathrm{H}}(\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H})^{-1}\boldsymbol{H}^{\mathrm{H}} = \boldsymbol{T}^{\mathrm{H}}\boldsymbol{H}^{\dagger},$$

where $\boldsymbol{F}^{\dagger}$ denotes the pseudo-inverse of $\boldsymbol{F}$, then

$$\begin{aligned}\boldsymbol{R} \triangleq \boldsymbol{F}^{\dagger}\boldsymbol{R}_x(1)(\boldsymbol{F}^{\dagger})^{\mathrm{H}} &= \boldsymbol{F}^{\dagger}\boldsymbol{H}\boldsymbol{J}\boldsymbol{H}^{\mathrm{H}}(\boldsymbol{F}^{\dagger})^{\mathrm{H}}\\ &= \boldsymbol{T}^{\mathrm{H}}\underbrace{\boldsymbol{H}^{\#}\boldsymbol{H}}_{=\boldsymbol{I}}\boldsymbol{J}\underbrace{(\boldsymbol{H}^{\dagger}\boldsymbol{H})^{\mathrm{H}}}_{=\boldsymbol{I}}\boldsymbol{T}\\ &= \boldsymbol{T}^{\mathrm{H}}\boldsymbol{J}\boldsymbol{T}.\end{aligned} \tag{9}$$

Since $\boldsymbol{T}$ is unitary, $\boldsymbol{T}\boldsymbol{R} = \boldsymbol{T}\boldsymbol{T}^{\mathrm{H}}\boldsymbol{J}\boldsymbol{T} = \boldsymbol{J}\boldsymbol{T}$. Letting $\boldsymbol{T} = [\boldsymbol{t}_1, \ldots, \boldsymbol{t}_d]^{\mathrm{T}}$, we see that (9) yields $\boldsymbol{R}\boldsymbol{t}_1 = \boldsymbol{t}_2$, $\boldsymbol{R}\boldsymbol{t}_{d-1} = \boldsymbol{t}_d$ and $\boldsymbol{R}\boldsymbol{t}_d = \boldsymbol{0}$. From such a recursion, we can *uniquely* determine $\boldsymbol{T}$ up to a scalar phase ambiguity $\mathrm{e}^{\mathrm{j}\phi}$. For the detailed algorithm and the technique of handling the additive noise, please refer to [49].

The algorithm enforces a Jordan structure on the input correlation matrix and essentially reconstructs the channel vectors in $\boldsymbol{H}$ based on the information provided by the change of rank of $\boldsymbol{R}_x(k)$. It, however, does not fully exploit the Toeplitz structure of $\boldsymbol{H}$. Although this algorithm is preliminary and suboptimal in retrospect, it reformulated a traditional spectral channel identification problem into a parameter estimation framework and eliminated the issue of the lack of phase information in the second-order statistics [8]. More importantly, it opened up a new research direction and led to the development of a class of data-efficient and fast converging blind equalization techniques.

3.2. Frequency domain approaches

If we look at the above time-domain approach more carefully, it is not difficult to see that it works because $\boldsymbol{H}$ has more rows than columns. Since $\boldsymbol{H}$ has MK rows and $L+K$ columns, $MK \geqslant L+K$ or $K \geqslant L/(M-1)$ clearly indicates the magic of oversampling. If there is no oversampling, i.e., $M=1$, then $K \to \infty$. Otherwise, for $M \geqslant 2$, we can always find a large enough K such that $\boldsymbol{H}$ has more rows than columns. Therefore, oversampling is a key to the aforementioned success. This explanation in the time domain is superficial and it turns out that much more insight can be gained by examining the problem in a frequency domain formulation.

For a single-input and single-output system, the output $x(k) = \sum_{j=0}^{L} h_j s(k-j)$. *If* $s(\cdot)$ is an unknown *stationary* random process, the *only* available second-order statistic of $x(\cdot)$ is the correlation function $R_x(\cdot)$ in the time domain or the spectral density function $S_x(\cdot)$ in the frequency domain. If we assume that $s(\cdot)$ is white with variance σ_s^2, then

$$S_x(\mathrm{j}\omega) = |H(\mathrm{j}\omega)|^2 S_s(\mathrm{j}\omega) = \sigma_s^2 |H(\mathrm{j}\omega)|^2. \tag{10}$$

Clearly, there is no way to determine $H(j\omega)$ or $H(z)$ from $S_x(j\omega)$ since the phase information of $H(j\omega)$ is missing in (10), unless we assume that $H(z)$ is a minimum-phase system. That is also why most conventional blind equalization techniques rely on higher-order statistics.

Nevertheless, if $s(\cdot)$ is a *cyclostationary* process, so is $x(\cdot)$. Then, the power spectral density is no longer the only second-order statistic available; there exist the so-called *cyclic spectra* of the obserations [48],

$$\Gamma^k(j\omega) = H(j\omega)H^*\left(j\omega - jk\frac{2\pi}{M}\right) + \sigma^2\delta(k),$$

$$k = 0, 1, \ldots, M-1. \tag{11}$$

When $k = 0$, the cyclic spectrum becomes the conventional power spectral density function. However, $\Gamma^k(j\omega)$ for $k \neq 0$ provides phase information on $H(j\omega)$, which can be exploited to identify the $H(j\omega)$ as shown below.

In the z-domain, (11) leads to the equation

$$\Gamma^k(z) = H(z)H^*\left(e^{jk(2\pi/M)}\frac{1}{z^*}\right), \quad k = 0, \ldots, M-1. \tag{12}$$

The problem of channel identification is then equivalent to identifying $H(z)$ using $\Gamma^k(z)$, which can be approached by identifying the zeros of $H(z)$ from those of $\Gamma^k(z)$. The following theorem was given in [50].

Theorem 1. *The channel transfer function $H(z)$ is uniquely determined (identified) by $\{\Gamma^k(z))\}$ up to a multiplicative constant if and only if $H(z)$ does not have zeros uniformly spaced around a circle with separation of $2\pi/M$ radians. Moreover, if the channel is identifible,*

$$\mathscr{Z}(H(z)) = \bigcap_k \mathscr{Z}(\Gamma^k(z)), \tag{13}$$

where $\mathscr{Z}(H(z))$ denotes for the set of zeros of $H(z)$.

Several frequency domain approaches have been proposed that estimate the channel parameters from the cyclic spectra.

3.2.1. Identification by poles and zeros.

In [9], Ding proposed a two-step approach for the identification of $H(z)$. The cyclic spectra $\{\Gamma^{(k)}(e^{j\omega})\}$ are estimated first by FFT-based spectrum estimators. Next, the poles and zeros of the channel transfer function are determined from the estimated cyclic spectra.

3.2.2. Eigenstructure-based identification

A closed-form eigenstructure-based approach was invented by Hassibi [48] and later independently by Giannakis [14]. From (11), it is clear that the problem of identifying $H(z)$ from $\{\Gamma^k(z)\}$ can be formulated as one of finding the greatest common divisor (GCD) from a set of polynomials. Although the Euclid algorithm, often considered as one of the oldest nontrivial algorithms, can be applied to find the GCD, it is sensitive to the perturbation caused of the estimation. To obtain a more robust algorithm, Hassibi formulated the following optimization approach:

$$\hat{H}(z) = \arg\min_{H(z)} \left\| \hat{\Gamma}^{(k_1)}(z)H^*\left(e^{jk_2(2\pi/M)}\frac{1}{z^*}\right) - \hat{\Gamma}^{(k_2)}(z)H^*\left(e^{jk_1(2\pi/M)}\frac{1}{z^*}\right)\right\|^2, \tag{14}$$

where $\hat{\Gamma}^{(k_1)}(z)$ and $\hat{\Gamma}^{(k_2)}(z)$ are estimated output spectra. The above optimization can be translated to a quadratic optimization

$$\hat{\boldsymbol{h}} = \arg\min_{\|\boldsymbol{h}\|=1} \boldsymbol{h}^{\mathrm{H}}\boldsymbol{Q}\boldsymbol{h}, \tag{15}$$

whose solution can be obtained performing eigendecomposition of Q. Details can be found in [48].

3.2.3. New stochastic method [57]

Based on (3), define the autocorrelation matrix $\boldsymbol{R}_x(i) = E\{\boldsymbol{x}(k)\boldsymbol{x}^{\mathrm{H}}(k-i)\}$ and correspondingly, the power spectral matrix $\boldsymbol{Q}(e^{j\omega}) = \sum_{i=-L+1}^{L-1}\boldsymbol{R}_x(i)e^{j\omega i}$. Then Xu et al. [57] show that

$$\boldsymbol{Q}(e^{j\omega}) = \left(\sum_{k=1}^{L}\boldsymbol{h}_k e^{j\omega(k-1)}\right)\left(\sum_{k=1}^{L}\boldsymbol{h}_k e^{j\omega(k-1)}\right)^* + \sigma_n^2\boldsymbol{I}. \tag{16}$$

In other words, $\boldsymbol{Q}(e^{j\omega})$ is a rank-one $M \times M$ Hermitian matrix whose principal eigenvector is $\boldsymbol{c}(\omega) = e^{j\alpha(\omega)}\sum_{k=1}^{L}\boldsymbol{h}_k e^{j\omega(k-1)}$, where $e^{j\alpha(\omega)}$ is an

unknown phase. By evaluating $\boldsymbol{c}(\omega_k)$ at different frequency points $\{\omega_k\}$ based on the sample power spectrum $\hat{\boldsymbol{Q}}(e^{j\omega_k})$, the channel vectors can be determined by the following estimator:

$$\{\hat{\boldsymbol{h}}_i\} = \arg\min_{\{\boldsymbol{h}_i\}} \sum_{\omega_k} \|\hat{\boldsymbol{c}}(\omega_k) - \hat{\boldsymbol{c}}(\omega_k, \boldsymbol{h}_i)\|^2.$$

It turns out that the above optimization problem can also be nicely solved by using SVDs [48].

4. Channel identifiability

Despite the early success of the aforementioned statistics-based blind identification algorithms, it was still not clear why oversampling can overcome the lack of phase information, other than the explanation that oversampling generates a cyclostationary process whose frequency correlation (cyclic spectra) reveals the phase information of the channel. Then a legitimate question to ask is what kind of channels can be identified using second-order statistics only? This fundamental question attracted the attention of many researchers, whose investigations gradually resolved this enigma.

To clarify the discussion, one needs to accept the following.

Definition 1. An FIR channel is identifiable if it can be determined up to a multiplicative constant.

4.1. Identifiability condition

There have been a number of studies based on Definition 1. One of the earliest and most referenced results in the following [51].

Theorem 2. *The channels are identifiable if and only if the matrix $\boldsymbol{H}(L+1)$ has full column rank.*

Recall that

$$\boldsymbol{H}(L) = \underbrace{\begin{bmatrix} \boldsymbol{h}_L & \boldsymbol{h}_{L-1} & \cdots & \boldsymbol{h}_0 & 0 & \cdots & 0 \\ 0 & \boldsymbol{h}_L & \boldsymbol{h}_{L-1} & \cdots & \boldsymbol{h}_0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \boldsymbol{h}_L & \boldsymbol{h}_{L-1} & \cdots & \boldsymbol{h}_0 \end{bmatrix}}_{2L+1 \text{ blocks}}. \tag{17}$$

The rank condition on $\boldsymbol{H}(L+1)$, which is a so-called Sylvester resultant matrix, was studied by Kung et al. as early as 1976 [25]. The Toeplitz-block structure exhibits many interesting properties which can be used to interpret the identifiability condition. To provide further insight, Xu et al. [58] presented an alternative representation of the multichannel system by transforming the blind identification problem into a standard system identification framework.

Denote $\boldsymbol{s}(k) = [s(k-1)\ s(k-2)\ \ldots\ s(k-L)]^{\mathrm{T}}$, $u(k) = x_1(k)$ and $\boldsymbol{y}(k) = [x_2(k)\ \ldots\ x_M(k)]^{\mathrm{T}}$. The multichannel FIR system can be described by an SIMO (single input, multiple output) state space model:

$$\boldsymbol{s}(k+1) = \underbrace{\begin{bmatrix} -h_1(1) & \cdots & -h_1(L-1) & -h_1(L) \\ 1 & & & \\ & \ddots & & \\ & & 1 & 0 \end{bmatrix}}_{\boldsymbol{A}} \boldsymbol{s}(k) + \underbrace{\begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}}_{\boldsymbol{b}} u(k), \tag{18}$$

$$\boldsymbol{y}(k) = \underbrace{\begin{bmatrix} \boldsymbol{c}_2 \\ \boldsymbol{c}_3 \\ \vdots \\ \boldsymbol{c}_M \end{bmatrix}}_{\boldsymbol{C}} \boldsymbol{s}(k) + \underbrace{\begin{bmatrix} h_2(0) \\ h_3(0) \\ \vdots \\ h_M(0) \end{bmatrix}}_{\boldsymbol{d}} u(k), \tag{19}$$

where

$$\boldsymbol{c}_i = (h_1(i) - h_i(0)h_1(1)\ \ldots\ h_i(L) - h_i(0)h_1(L)).$$

The transfer matrix of the above model is

$$\boldsymbol{h}(z) = \boldsymbol{C}(z\boldsymbol{I} - \boldsymbol{A})^{-1}\boldsymbol{b} + \boldsymbol{d} = \frac{[h_2(z)\ \ldots\ h_M(z)]^{\mathrm{T}}}{h_1(z)}. \tag{20}$$

Using standard system identification results [23, 31], it is now clear that Theorem 2 can be restated in the following more informative format [51, 33, 11, 26],

Theorem 3. *The channels are identifiable if and only if* $\{\sum_{k=0}^{L} h_i(k)z^{-k}\}$ *are coprime, i.e., they do not share any common roots.*

Note that in the above theorems, no requirement regarding the inputs is specified. For most statistics-based approaches, the implicit assumption is that the inputs be white noise, or at least that the input covariance matrix be of full rank. Such conditions are, of course, plausible in most communication scenarios with sufficient data samples. However, when the data sequence is considerably short, statistical characterization of the unknown input is hardly useful, since the statistical estimates based on finite samples are very poor. In these cases, it appears natural to model the inputs as *deterministic* unknown signals. This assumption means that the input characterstics also have to be accounted for in the identifiability studies.

For the finite-sample model in Section 2, the following was shown in [58].

Theorem 4. *The blind identification problem has a unique solution if*

- $\{\sum_{k=0}^{L} h_i(k)z^{-k}\}$ *are coprime, i.e., they do not share any common roots*
- $s(1), \ldots, s(N)$ *contain no fewer than* $2L + 1$ *modes*

The number of *modes*, often referred to as the *linear complexity*, is a measurement of the diversity in a *finite* sequence. It can be analogous to the number of frequency components in an infinite data sequence [23, 31]. Interestingly, the above conditions are the same as those for identifying a rational function with denominator and numerator both of order L [31]. It is true that the system under consideration is only FIR, and can be sufficiently identified with a *known* input $s(\cdot)$ that has $L + 1$ modes (sometimes referred to as persistent excitation of order $L + 1$ [3]). For blind identification, the input is *unknown* and hence more than $2L + 1$ modes are required in the input signal.

4.2. Related studies

An interesting and insightful observation was made by Hua [32] on the relation between the identifiability and the rank condition of Fisher Information (FI) matrices. Previous results on identifiability were usually algorithm-dependent, which made a unified comparison difficult.

With $s = [s(1), \ldots, s(N)]^T$, $\bar{\boldsymbol{h}}_i = [h_i(0), \ldots, h_i(L)]^T$, and $\bar{\boldsymbol{h}} = [\bar{\boldsymbol{h}}_1^T, \ldots, \bar{\boldsymbol{h}}_M^T]^T$, some of the major results in [32] can be summarized as follows. Let

$$\boldsymbol{a} = \begin{bmatrix} \mathrm{Re}(\bar{\boldsymbol{h}}) \\ \mathrm{Re}(\boldsymbol{s}) \\ \mathrm{Im}(\bar{\boldsymbol{h}}) \\ \mathrm{Im}(\boldsymbol{s}) \end{bmatrix}.$$

The Fisher information matrix for the parameters in $\boldsymbol{a}$ is

$$\boldsymbol{F} = \frac{2}{\sigma_n^2} \begin{bmatrix} \mathrm{Re}(\boldsymbol{F}_c) & -\mathrm{Im}(\boldsymbol{F}_c) \\ \mathrm{Im}(\boldsymbol{F}_c) & \mathrm{Re}(\boldsymbol{F}_c) \end{bmatrix},$$

where $\boldsymbol{F}_c$ is a *complex* FI matrix with $\boldsymbol{h}$ and $\boldsymbol{s}$ as its elements. Readers are referred to the original paper for details.

Many profound relations between nullity $(\boldsymbol{F}_c)$ and the system characteristics are revealed by the following theorem.

Theorem 5. (a) *nullity* $(\boldsymbol{F}_c) \geqslant 1$.

(b) *nullity* $(\boldsymbol{F}_c) > 1$ *if* $\{s(k), k = 1, \ldots, N\}$ *has less than* $L + 2$ *modes, or the* M *channels share a common zeros, or* $N \leqslant 2L + 1$.

(c) *nullity* $(\boldsymbol{F}_c) = 1$ *if* $\{s(k), k = 1, \ldots, N\}$ *has* $2L + 1$ *or more modes, the* M *channels do not share a common zero, or* $N \geqslant 4L + 1$.

(d) *if nullity* $(\boldsymbol{F}_c) = d$, *let* $\boldsymbol{F}_c'$ *be* $\boldsymbol{F}_c$ *without* d *rows and columns of* $\boldsymbol{F}_c$ *corresponding to* d *nonzero parameters in* $\boldsymbol{a}$, *then nullity* $(\boldsymbol{F}_c) = 0$.

Part (a) of Theorem 5 implies that there is always a complex degree of uncertainty in the M-channel system, which is expected. Hence, one can define that the M-channel system is identifiable if nullity $(\boldsymbol{F}_c) = 1$. With this definition, Part (b) provides a necessary channel identifiability condition, and

Part (c) gives a sufficient ID condition. Beside the coprimeness restriction on the channels, which has been studied extensively in previous studies, we see that the nature of the input sequence also plays a decisive role in blind identification.

4.3. Characteristics of unidentifiable channels

Although the identifiability conditions, and in particular the common zero requirement on the channels has been investigated extensively, understanding the connections among these conditions usually remains at a conceptual level. Clearly, it is very important for practitioners to understand physically what kinds of channel can be blindly identified. To provide an intuitive interpretation of these conditions, Tugnait [33] considered multipath channels in wireless systems, where $h_i(t)$ can be expressed as

$$h_i(t) = \sum_{j=1}^{d} \alpha_j g(t - \Delta t_j),$$

where $g(\cdot)$ is the pulse-shape function, $\{\Delta t_j\}$ are the multipath delays, $\{\alpha_j\}$ are the relative complex gains of eatch path and d is the total number of paths. Based on the above formulation, the following classes of multipath channels are shown to have common zeros:

1. channels with delays that are integer multiples of T;
2. channels with delays that are integer multiples of $T/2$ for even M.

If one of the above situation occurs, the channels become unidentifiable, regardless of the sampling rate. These results help practitoners to understand the limitations of the blind channel identification approaches discussed in this paper, and to decide what types of channel equalization methods to use.

4.4 Channel order selection

Although selection of the channel order L is not an identifiability issue, it is undoubtedly crucial in modeling the multichannel FIR system. Knowledge of the model order is as critical to the blind identification algorithms as it is to most parametric estimation methods. Most existing approaches assume the length of the channels to be known a priori. Among the limited studies on order selection, a majority part resorts to some type of ad hoc eigen-based detection schemes. More objective criteria such as MDL [37], AIC [1] are yet to be derived.

In [30] it was shown that $\boldsymbol{D}_M$ in (24) will have more than one null vector if the channel is overdetermined, i.e., $L > L_{\text{true}}$; there is also a one-to-one relation between $(L - L_{\text{true}})$ and the number of null vectors in $\boldsymbol{D}_M$. Therefore, one can in principle determine the true order by checking the rank of $\boldsymbol{D}_M$. Naturally, in the presence of noise, we have to count the number of smaller singular (eigen) values that are close to one another.

A similar detection technique was proposed in [57] based on the use of data covariance matrix $\hat{\boldsymbol{R}}_x = (1/N)\boldsymbol{X}(K)\boldsymbol{X}^{\mathrm{H}}(K)$. Ideally, $\boldsymbol{E}[(1/N)\boldsymbol{X}(K)\boldsymbol{X}^{\mathrm{H}}(K)]$ is of rank $L + K$ and thus has $L + K$ nonzero eigenvalues. Hence, by observing the distribution of the eigenvalues of the $\hat{\boldsymbol{R}}_X$, it is possible to determine the channel order L. This problem seems very similar to that of estimating the number of sources in direction finding [62, 56, 55]. However, the problems are not exactly the same since all the data vectors in sensor array processing are mutually uncorrelated. In the channel identification case, neighboring data vectors have overlapping elements and hence are correlated. This correlation can complicate the development of a statistical direction method.

5. Subspace approaches

In the early studies, it was commonly believed that knowledge of the input statistics (not the actual input) must be available for channel identification. However, in some typical applications, such as mobile communications, the channel may vary rapidly and we can obtain only a short data sequence associated with the 'same' channel. In this case, the statistics estimated from the finite channel output may not be accurate, causing model mismatch and leading to performance degradation of the statistics-based approaches. In the search for more

robust techniques to handle short data sequences, Liu et al. [30, 58] developed a least squares (LS) method that accomplishes channel identification without requiring the input statistics. Similar techniques were also proposed independently in [19, 2].

5.1. Least-squares approach

The algorithm assumes that the noise is a stationary Gaussian white random process, whereas the signals are *deterministic* unknowns. It was originally termed as the deterministic approach to emphasize its deterministic nature. The basic observation here is that for any pair of two noise-free outputs $x_i(k)$ and $x_j(k)$,

$$h_j(k) \odot x_i(k) = h_j(k) \odot (h_i(k) \odot s(k)) = h_i(k) \odot \underbrace{(h_j(k) \odot s(k))}_{=x_j(k)} = h_i(k) \odot x_j(k), \tag{21}$$

or in matrix form,

$$[\boldsymbol{X}_i(L) \quad -\boldsymbol{X}_j(L)] \begin{bmatrix} \bar{\boldsymbol{h}}_j \\ \bar{\boldsymbol{h}}_i \end{bmatrix} = 0. \tag{22}$$

To exploit all the possible cross relations simultaneously, the LS method employs a transform which was later defined as the *data selection transform* (DST), $\mathscr{D}$ [61]. Given the data matrices from all channels $\boldsymbol{X}_1(L), \ldots, \boldsymbol{X}_M(L)$,

$$\mathscr{D}_2(\boldsymbol{X}_{(\cdot)}(L)) = [\boldsymbol{X}_2(L) - \boldsymbol{X}_1(L)],$$

$$\mathscr{D}_M(\boldsymbol{X}_{(\cdot)}(L)) = \left[\begin{array}{ccc|c} & \mathscr{D}_{M-1}(\boldsymbol{X}_{(\cdot)}(L)) & & \boldsymbol{0} \\ \hline \boldsymbol{X}_M(L) & & & -\boldsymbol{X}_1(L) \\ & \boldsymbol{X}_M(L) & & -\boldsymbol{X}_2(L) \\ & & \ddots & \vdots \\ & & \boldsymbol{X}_M(L) & -\boldsymbol{X}_{M-1}(L) \end{array}\right]. \tag{23}$$

With $\boldsymbol{D}_M \overset{\text{def}}{=} \mathscr{D}_M(\boldsymbol{X}_{(\cdot)}(L))$, the channel estimate problem is cast into a least-squares form:

$$\hat{\bar{\boldsymbol{h}}} = \arg \min_{\|\bar{\boldsymbol{h}}\|=1} \bar{\boldsymbol{h}}^{\mathrm{H}} (\boldsymbol{D}_M^H \boldsymbol{D}_M) \bar{\boldsymbol{h}}, \tag{24}$$

where $\bar{\boldsymbol{h}} = [\bar{\boldsymbol{h}}_1^{\mathrm{T}} \ \ldots \ \boldsymbol{h}_M^{\mathrm{T}}]^{\mathrm{T}}$.

The LS method turned out to be highly data efficient when the SNR is relatively high. Indeed, from (22), it is clear that in the absence of noise,

$$\boldsymbol{D}_M \bar{\boldsymbol{h}} = \boldsymbol{0}.$$

Therefore, if $\boldsymbol{D}_M$ has only one null vector, which is true in most scenarios [30], the LS method provides *exact* channel estimates. This result substantiates the superior performance promised by the parametric methods. As an aside, it also implies that the structure of the output alone contains *sufficient* information to identify a single-input multiple-output system.

5.2. Channel subspace method

The key role of channel structure in single-input, multiple-output blind estimation was possibly disclosed most lucidly by the channel subspace (CS) method proposed independently by Moulines et al. [32] and Slock [43].

Built upon the concept of signal subspace used by many eigen-based algorithms in the context of array signal processing [35], especially the MUSIC approach [40], the CS method defines the column span of $\boldsymbol{H}(K)$, and its corresponding orthogonal complement, as the *channel signal subspace* $\boldsymbol{U}_s$ and the *channel orthogonal subspace* $\boldsymbol{U}_o$. Evidently, both of them can be calculated from the data matrix (defined in (6))

$$\boldsymbol{X}(K) = \boldsymbol{H}(K) \boldsymbol{S}(L + K), \tag{25}$$

provided that the input matrix $\boldsymbol{S}(L + K)$ is of full row rank. The algorithm relies on the following observation to determine the channel vector.

Theorem 6. *Assume that (i) $K > L$ and (ii) the matrix $\boldsymbol{H}(K-1)$ is of full-column rank. Let $\boldsymbol{H}'(K)$ be a filtering matrix with the same dimensions as $\boldsymbol{H}(K)$. The column space of $\boldsymbol{H}'(K)$ is included in the column space of $\boldsymbol{H}(K)$ if and only if the corresponding vectors $\boldsymbol{h}$ and $\boldsymbol{h}'$ are proportional.*

The above theorem asserts that in the absence of noise, $\boldsymbol{H}(K)$ constructed from the true channel

vector $\boldsymbol{h} = [\boldsymbol{h}_L^{\mathrm{T}}, \ldots, \boldsymbol{h}_0^{\mathrm{T}}]^{\mathrm{T}}$ is the only *structured* matrix that satisfies

$$\boldsymbol{U}_{\mathrm{o}}^{\mathrm{H}}(K)\boldsymbol{H}(K) = \boldsymbol{0}.$$

This indicates that the channel can be determined *uniquely* from the signal subspace of the data matrix $\boldsymbol{X}(K)$ or its corresponding orthogonal subspace $\boldsymbol{U}_{\mathrm{o}}$. Denoting by $\{\boldsymbol{u}_{\mathrm{o}}^i\}$ the orthogonal vectors in $\boldsymbol{U}_{\mathrm{o}}$, with some straightforward matrix and vector manipulation, the channel estimate can be formulated as

$$\hat{\boldsymbol{h}} = \arg \min_{\|\boldsymbol{h}\| = 1} \left\| \boldsymbol{h}^{\mathrm{H}} \sum_i \boldsymbol{G}_i \boldsymbol{G}_i^{\mathrm{H}} \boldsymbol{h} \right\|^2, \tag{26}$$

where $\boldsymbol{G}_i = \mathscr{T}_L(\boldsymbol{u}_{\mathrm{o}}^i(1), \ldots, \boldsymbol{u}_{\mathrm{o}}^i(K))$.

In practice, or course, $\boldsymbol{U}_{\mathrm{o}}$ has to be estimated from the data covariance matrix using eigen-decomposition techniques. Simulation studies show that the CS method performs similarly to the LS method given a short data sequence [61] and sometimes slightly better. A close investigation of the connection of the LS and CS approaches was conducted by Zeng and Tong [61]. They showed, somewhat expectedly, that the LS and CS approaches provide identical channel estimates with *probability one* for a system with $M = 2$. Although further analysis for $M > 2$ is yet to be performed, it will not be surprising if similar results hold.

5.3. Signal subspace method

In all the aforementioned algorithms, the focus has been on channel estimation. In practice, the ultimate goal is usually to recover the information bearing inputs rather than the channels. Theorem 5 on system identifiability indicates that the inputs are as the channels, if not more. Given a short data sequence where the number of inputs is comparable to the number of channel parameters, it may be practically preferable to determine the inputs directly from the system outputs, rather than via the two-step approach of channel identification and channel equalization. Such an idea turns out to be feasible by using the following observation of Liu and Xu [28].

Lemma 1. *The input vector* $\boldsymbol{s}$ *can be uniquely determined, up to a scalar multiplier, from the row span of* $\boldsymbol{S}(r)$ if $\boldsymbol{s}$ *contains more than* r *modes.*

It is seen that, similar to the assertion in Theorem 6, the row span of the Hankel input matrix $S(L + K)$ alone contains sufficient information to determine the input sequence. Denote by $\boldsymbol{V}_{\mathrm{s}}$, and correspondingly by $\boldsymbol{V}_{\mathrm{o}}$, the *input signal subspace* and the *input orthogonal subspace.*[2] Then $\boldsymbol{s}$, up to a scalar ambiguity, is the unique nontrivial solution of the following overdetermined system of linear equations,

$$\boldsymbol{V}\boldsymbol{s} = \boldsymbol{0}, \tag{27}$$

where

$$\boldsymbol{V}(r) = \underbrace{\begin{bmatrix} \boldsymbol{V}_{\mathrm{o}} & \cdots & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{V}_{\mathrm{o}} & \cdots & \vdots \\ \vdots & \cdots & \ddots & \boldsymbol{0} \\ \boldsymbol{0} & \cdots & \cdots & \boldsymbol{V}_{\mathrm{o}} \end{bmatrix}}_{r = L + K \text{ blocks}}, \quad \boldsymbol{0} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} \Bigg\} N - 2r + 1. \tag{28}$$

$\boldsymbol{V}(r)$ is characteristically interpreted as the *deconvolution matrix*. In effect, $\boldsymbol{V}(k)$, $k = 1, \ldots, r$, allow one to utilize the null space of $\boldsymbol{S}(r)$ to construct the null spaces of $\boldsymbol{S}(r - k + 1)$, $k = 1, \ldots, r$, thus eventually recover $\boldsymbol{S}(1) = \boldsymbol{s}$. The approach is similar in spirit to the channel subspace method.

5.4. Maximum likelihood approach

When the additive noise is white Gaussian, the maximum likelihood (ML) principle can be used for blind identification problem. Before one can formulate the likelihood function for parameter estimation, the input sequence needs to be characterized as well. In many applications where the channels need to be identified with a short data sequence, it appears natural to model the input as an unknown *deterministic* sequence as we implicitly did in the

[2] This time, however, we are concerned with row spans instead of column spans.

subspace methods. Similar assumptions can be found in array signal processing, where the *deterministic* model [33] is sometimes termed as *conditional* model [46].

To obtain the ML estimator, let $\boldsymbol{y} = [\boldsymbol{y}_1^{\mathrm{T}}, \ldots, \boldsymbol{y}_M^{\mathrm{T}}]^{\mathrm{T}}$, $\boldsymbol{y}_i = [x_i(L+1), \ldots, x_i(N)]^{\mathrm{T}}$ and $\bar{\boldsymbol{H}} = [\boldsymbol{H}_1^{\mathrm{T}}(N-L), \ldots, \boldsymbol{H}_M^{\mathrm{T}}(N-L)]^{\mathrm{T}}$, it is easy to show that the maximum likelihood estimation of $\bar{\boldsymbol{h}}$ from $\boldsymbol{y}$ is given by [32, 21, 43, 42]

$$\hat{\bar{\boldsymbol{h}}} = \arg \max_{\|\bar{\boldsymbol{h}}\| = 1} \boldsymbol{y}^{\mathrm{H}} \boldsymbol{P}_{\bar{\boldsymbol{H}}} \boldsymbol{y}, \tag{29}$$

where $\boldsymbol{P}_{\bar{\boldsymbol{H}}}$ is the projection matrix on the columns of $\bar{\boldsymbol{H}}$. Alternatively,

$$\hat{\bar{\boldsymbol{h}}} = \arg \min_{\|\bar{\boldsymbol{h}}\| = 1} \boldsymbol{y}^{\mathrm{H}} \boldsymbol{P}_{\bar{\boldsymbol{H}}}^{\perp} \boldsymbol{y}, \tag{30}$$

where $\boldsymbol{P}_{\bar{\boldsymbol{H}}}^{\perp}$ is the orthogonal complement of $\boldsymbol{P}_{\bar{\boldsymbol{H}}}$.

Conceptually, the above nonlinear minimization with respect to $\bar{\boldsymbol{h}}$ can be carried out by using the *iterative quadratic maximum likelihood* method [6, 24]. The difficulty, however, lies in the construction of $\boldsymbol{P}_{\bar{\boldsymbol{H}}}^{\perp}$ from each iteration. Hua [32] takes advantage of the unique structure of the system outputs and shows that $\boldsymbol{P}_{\bar{\boldsymbol{H}}}^{\perp}$ can be expressed as

$$\boldsymbol{P}_{\bar{\boldsymbol{H}}}^{\perp} = \boldsymbol{G}^{\dagger} \boldsymbol{G}^{\mathrm{H}},$$

where $\dagger$ denotes pseudo-inverse and $\boldsymbol{G} = \mathscr{D}(\boldsymbol{H}_1, \ldots, \boldsymbol{H}_M)$.

To isolate the channel vector $\boldsymbol{h}$ for update, note that

$$\boldsymbol{G}^{\mathrm{H}} \boldsymbol{y} = \boldsymbol{D}_M \bar{\boldsymbol{h}}.$$

Consequently, (30) becomes

$$\hat{\bar{\boldsymbol{h}}} = \arg \min_{\|\bar{\boldsymbol{h}}\| = 1} \bar{\boldsymbol{h}}^{\mathrm{H}} \boldsymbol{D}_M^{\mathrm{H}} (\boldsymbol{G}^{\mathrm{H}} \boldsymbol{G})^{\dagger} \boldsymbol{D}_M \bar{\boldsymbol{h}}.$$

This expression suggests the following iterative two-step estimation procedure:

1. Minimize $\bar{\boldsymbol{h}}^{\mathrm{H}} (\boldsymbol{D}_M^{\mathrm{H}} \boldsymbol{D}_M) \bar{\boldsymbol{h}}$ subject to $\|\bar{\boldsymbol{h}}\| = 1$ to yield $\hat{\bar{\boldsymbol{h}}}$.
2. Minimize $\bar{\boldsymbol{h}}^{\mathrm{H}} (\boldsymbol{D}_M^{\mathrm{H}} (\hat{\boldsymbol{G}}^{\mathrm{H}} \hat{\boldsymbol{G}})^{\dagger} \boldsymbol{D}_M) \bar{\boldsymbol{h}}$ subject to $\|\bar{\boldsymbol{h}}\| = 1$ to yield $\bar{\boldsymbol{h}}_{\mathrm{new}}$, where $\hat{\boldsymbol{G}}$ is constructed from $\hat{\bar{\boldsymbol{h}}}$.

One may have already noticed that step 1 coincides with the LS method. The relation between the LS method and the ML approach here resembles that between the Prony's method and the ML estimator in the context of exponential signal processing. Simulation studies show that the ML method outperforms the LS method when the channel condition is relatively poor; otherwise, it does not yield significant improvement.

It is worth pointing out that due to the complexity associated with the aforementioned algorithms, their practical usefulness still has to be proven. During the reviewing process, there emerges a promising direction in blind estimation that can be significant to practical applications. It is shown by Slock and Papadias [45], and also Giannakis and Halford [16], that instead of identifying the channels as a first step to estimating the equalizer, FIR blind equalizers can be constructed directly from the observations. Preliminary results indicate that the direct equalization schemes possess the robustness to the estimation of channel orders and their adaptive implementation may eventually enable blind equalization techniques to make an impact to practical systems.

6. Additional topics

6.1. Performance limitation

With the exception of the maximum-likelihood approach, many recent blind channel estimation algorithms utilize the second-order moments of the received signal. A natural question is: what is the best performance achievable by algorithms using the estimated second-order moments? One may expect that the performance of these algorithms degrades when the channel matrix is close to singular. This is indeed the case for all eigenstructure-based algorithms described above. Surprisingly, recent work by Zeng and Tong [61] showed that the singularity of the channel matrix is not necessarily the fundamental performance limitation on the blind channel estimation algorithms based on second-order statistics. At first glance, this seems to contradict the fact that the channel is not identifiable when the channel matrix is singular. In such a case, there is a finite number of solutions satisfying the constraints imposed by the second-order statistics of the observation. However, this does not imply that when the

parameter space is restricted to the neighborhood of the true channel, there may exist consistent estimators. Indeed, there are such estimators as shown in [61].

To elaborate the above argument, we highlight the results presented in [61]. For convenience, we consider the real case in the analysis below. Given the correlation function $r_{ij}(u) \triangleq E\{x_i(t)x_j(t+u)\}$, define

$$\boldsymbol{r}_{ij} = \begin{cases} [r_{ij}(0), \ldots, r_{ij}(L)]^{\mathrm{T}}, & i \geqslant j, \\ [r_{ij}(1), \ldots, r_{ij}(L)]^{\mathrm{T}}, & i < j, \end{cases} \tag{31}$$

$$\boldsymbol{r} = [\boldsymbol{r}_{11}^{\mathrm{T}}, \ldots, \boldsymbol{r}_{1M}^{\mathrm{T}}, \ldots, \boldsymbol{r}_{M1}^{\mathrm{T}}, \ldots, \boldsymbol{r}_{MM}^{\mathrm{T}}]^{\mathrm{T}}. \tag{32}$$

The Jacobian $\boldsymbol{s}(\boldsymbol{h})$ of the correlation with respect to the channel vector $\boldsymbol{h}$ is given by

$$\boldsymbol{s}(\boldsymbol{h}) = \frac{\partial \boldsymbol{r}}{\partial \boldsymbol{h}}. \tag{33}$$

The estimated correlation is given by

$$\hat{r}_{ij}(u) = \frac{1}{N_s - u} \sum_{t=1}^{N_s - u} x_i(t)x_j(t+u), \tag{34}$$

where $N_s = N - L + 1$ is the total number of output vectors. We consider the performance of all consistent estimators using $\{\hat{r}_{ij}(u)\}$. Assume, without loss of generality, that the channel vector is normalized, i.e., $\|\boldsymbol{h}\|_2 = 1$. The asymptotically normalized mean-square error (ANMSE) is defined by

$$\mathrm{ANMSE} = \lim_{N_s \to \infty} N_s E\{\|\hat{\boldsymbol{h}}(N_s) - \boldsymbol{h}\|^2\}, \tag{35}$$

where $\hat{\boldsymbol{h}}(N_s)$ is the estimated channel vector using all the output data. The signal-to-noise ratio (SNR) is defined by

$$\mathrm{SNR} = \frac{1}{M\sigma_n^2} E\left\{\sum_{i=1}^{M} |x_j(i)|^2\right\}. \tag{36}$$

The following theorem [61] gives the achievable ANMSE among all consistent estimators and the ANMSE of the LS/CS approach, when there are two subchannels.

Theorem 7. *Let* ANMSE* *and* $\mathrm{ANMSE}_{\mathrm{LS/CS}}$ *be the achievable* ANMSE *and the* ANMSE *of the* LS/CS *estimators, respectively. Under regularity conditions* [36],

$$\mathrm{ANMSE}_* = \mathrm{tr}\{(\boldsymbol{s}^{\mathrm{T}}(\boldsymbol{h})\Sigma^{-1}(\boldsymbol{h})\boldsymbol{s}(\boldsymbol{h}))^{-1}\} > \gamma_* \frac{\kappa^2(\boldsymbol{s}(\boldsymbol{h}))}{\mathrm{SNR}^2}, \tag{37}$$

$$\mathrm{ANMSE}_{\mathrm{LS/CS}} = \sum_{k=1}^{2L+1} \frac{\sigma^2(\lambda_k^2 + \sigma^2)}{\sigma_s^2(\lambda_k^2)^2} \geqslant \gamma_{\mathrm{LS/CS}} \frac{\kappa^2(\mathscr{H}_{L+1}(\boldsymbol{h}))}{\mathrm{SNR}}, \tag{38}$$

where $\Sigma(\boldsymbol{h})$ *is the normalized asymptotic covariance of the estimated correlation functions,* $\kappa(\boldsymbol{h})$ $(\kappa(\boldsymbol{s}(\boldsymbol{h})))$ *is the condition number of* $\boldsymbol{h}$ $(\boldsymbol{s}(\boldsymbol{h}))$, $\lambda_{\max} = \lambda_1 > \lambda_2 > \cdots > \lambda_{2L+1} = \lambda_{\min}$ *are the singular values of* $\boldsymbol{h}$, γ_* *and* $\gamma_{\mathrm{LS/CS}}$ *are constants independent of* SNR *and* $\boldsymbol{h}$.

One implication of the above theorem is that when $\boldsymbol{s}(\boldsymbol{h})$ is nonsingular, there exist consistent estimators that achieve the lowest possible ANMSE. It can be shown that the singularity of $\boldsymbol{s}(\boldsymbol{h})$ implies the singularity of $\boldsymbol{h}$. On the other hand, the singularity of $\boldsymbol{h}$ does not imply the singularity of $\boldsymbol{s}(\boldsymbol{h})$. In fact, $\boldsymbol{s}(\boldsymbol{h})$ is singular if and only if all the subchannels share the same *reciprocal zeros.*

6.2. Optimization in subspaces

A number of existing approaches can be unified by a new formulation given in [61]. The eigenstructure approaches to blind channel estimation optimize objective functions of the type

$$J_s(\boldsymbol{h}) = \boldsymbol{h}^{\mathrm{H}}\boldsymbol{Q}\boldsymbol{h}. \tag{39}$$

By choosing Q differently, one obtains the LS, CS algorithms as well as the two eigenstructure-based frequency domain approaches mentioned in Section 3.2. The advantage of optimizing $\boldsymbol{J}_s(\boldsymbol{h})$ is that it is quadratic, and a closed-form solution is readily available. Unfortunately, this approach fails when the channel is close to being unidentifiable. On the other hand, the algorithm that achieves the ANMSE_* minimizing

$$\boldsymbol{J}_*(\boldsymbol{h}) = \|\hat{\boldsymbol{R}} - \boldsymbol{R}(h)\|_W^2 \tag{40}$$

for some specific choice of weighting W, see [36, 15]. It is shown that [61] optimizing $\boldsymbol{J}_*(\boldsymbol{h})$ may lead to consistent estimators even when the

channel matrix is singular provided one can restrict the channel in the neighborhood of the true channel. Such information may be obtained from previous estimates. Unfortunately the optimization of $J_*(\boldsymbol{h})$ is highly nonlinear, and local minima exist. It is also not practical because of two factors. First, searching for the optimal channel vector $\boldsymbol{h}$ in high dimension can be costly. Secondly, the optimal weighting in (40) cannot be obtained easily from data. The issue is how to combine the two optimization criteria in a sensible way.

Zeng and Tong proposed the following joint optimization criterior that unifies a number of approaches:

$$\min_{\boldsymbol{h}\in\mathscr{S}} J_2(\boldsymbol{h}) \text{ subject to } J_1(\boldsymbol{h}) \leqslant \alpha, \tag{41}$$

where α is a given threshold, and $\mathscr{S}$ is the parameter space. By choosing α and $\mathscr{S}$ differently, the optimization of the above criteria leads to a number of algorithms including the LS, CS, eigenstructure-based frequency domain approaches, and the asymptotically best consistent estimators. In fact, the above criteria also apply to the ML approach by replacing $J_s(h)$ defined above by the likelihood function.

6.3. Incorporation of additional information

A general rule in the model-based parameter estimation problem is that the more information one incorporates into the algorithm, the better estimation performance one can achieve. The performance of the previously mentioned blind estimation methods can be enhanced by exploiting prior system information. For instance, in wireless communications, knowledge of the transmitter pulse, which is generally available in all commercial applications, can be utilized to substantially reduce the amount of data and/or input SNR necessary for channel identification, while simultaneously reducing the computational complexity. Schell et al. [39] extended the LS approach by including knowledge of the transmitter pulse; the resulting method outperforms most existing blind methods in computational complexity and data efficiency. The principal component structure of random fading channels is exploited in [61].

Beside exploiting knowledge of the transmitter pulse, an alternative way to enhance the estimation performance is to take advantage of the physical structure of the channels. For example, the FIR channels in narrow-band wireless communications consist of a sum of multipath rays arriving from various directions. Under this model, instead of FIR filter coefficients, the channel parameters are the directions of arrival (DOAs), complex amplitudes, and relative time delays of the multipath rays. Using such a channel representation, Yang and Swindlehurst [59] showed that channel parameters can be estimated by solving a least-squares minimization problem involving the array data in the frequency domain. Superior estimation performance is reported over several previously mentioned blind identification algorithms.

6.4. Extension to multiple sources

Most of the parametric blind identification research concerns only a single-input multiple-output (SIMO) system. In the following, we outline some preliminary results for multiple single-input multiple-output (MSIMO) system blind identification. We shall first distinguish an MSIMO system from a general MIMO system – the MSIMO system we consider here simply consists of several subsystems, each of which is an SIMO system. More specifically, different inputs are limited to their own subsystems. A typical example of a MSIMO system is an antenna array wireless system with co-channel users.

To put the blind identification problem of an MSIMO into perspective, consider the superposition of P SIMO systems, each of which can be described by (6). Mathematically, the system output matrix is given by

$$\boldsymbol{X} = \sum_{i=1}^{P} \boldsymbol{H}_i \boldsymbol{S}_i = [\boldsymbol{H}_1 \ \ldots \ \boldsymbol{H}_p]\,[\boldsymbol{S}_1^{\mathrm{T}} \ \ldots \ \boldsymbol{S}_P^{\mathrm{T}}]^{\mathrm{T}}.$$

The goal again is to identify the channel characteristics and the inputs from the system outputs without using higher-order statistics.

With sufficient data and proper smoothing, it is shown in [27] that the column span of $[\boldsymbol{H}_1\ \boldsymbol{H}_2\ \ldots\ \boldsymbol{H}_p]$ and the row span of $[\boldsymbol{S}_1^{\mathrm{T}}\ \ldots\ \boldsymbol{S}_P^{\mathrm{T}}]^{\mathrm{T}}$ can still be computed from $\boldsymbol{X}$. The problem, however, is whether or not these subspaces still contain sufficient information for identifying the channels and inputs.

Recent studies by Liu and Xu [27] and Slock [44] provide some encouraging answers to these questions. It is shown that both the channel vectors and the input vectors can be determined up to a $P \times P$ (full-rank) matrix transform. In other words, if $\boldsymbol{h}^i$, $\boldsymbol{s}^i$, $i = 1, \ldots, P$, are respectively the channel vector and input vector corresponding to the ith subsystem, one can determine

$$[\boldsymbol{h}^1\ \boldsymbol{h}^2\ \ldots\ \boldsymbol{h}^p]\boldsymbol{W} \text{ or } \boldsymbol{W}\begin{bmatrix}\boldsymbol{s}^1\\ \vdots\\ \boldsymbol{s}^p\end{bmatrix}$$

from the subspace structure of the system outputs. Here $\boldsymbol{W}$ represents the $P \times P$ transformation ambiguity. Although a fundamental identifiability analysis is yet to be conducted, it will not be surprising if the same results can be derived from the FI matrix.

Without extra information, it does not seem to be possible to remove the remaining ambiguity. In [44], it is suggested to apply higher-order statistics to isolate the channels. This approach, however, may be forbidden due to the lack of enough data samples. In comparison to the channels, the inputs are generally more restricted, especially in digital communications. In particular, most digital communication signals have a finite alphabet (e.g., BPSK, QPSK). It was proved in [60, 47] that such ambiguity can be easily removed given sufficient data samples. Therefore, one can use the subspace structure of the system to deconvolve the outputs, and then remove $\boldsymbol{W}$ exploiting the finite alphabet property. Preliminary studies, both theoretically and experimentally, have demonstrated the feasibility of such a scheme [29].

6.5. Direct equalizer estimation

Recently, there emerges a promising direction in blind channel equalization that may be more appropriate for practical implementation. It was shown by Slock and Papadias [45] and also by Giannakis and Halford [16] that an FIR equalizer can be estimated directly from the observations and the step of channel estimation can be skipped. Preliminary results indicated that this approach is not so sensitive to the channel order estimation as the two-step algorithms and it can be implemented adaptively, leading to much simpler realization than aforementioned techniques. Although further study is required to evaluate this new approach, it does have the potential to realize blind equalization in practical systems.

7. Conclusion

Channel equalization is traditionally perceived as a standard linear system identification (or black-box) problem with the training sequence as the *probing* input signal. In many applications, the probing signals may not be easy to inject or they may present an extra burden (e.g., requiring excess bandwidth in communication systems). The emergence of blind channel equalization techniques based on advanced signal processing algorithms eliminates the need for a probing signal and simplifies the requirements for channel equalization. In particular, recent developments in blind equalization research have led to a class of rapidly converging and data efficient algorithms that can effectively equalize the channel with a surprisingly small number of data points (e.g., 50–100 symbols). In this paper, we surveyed recent research efforts in blind equalization and showed that the cyclostationarity and subspace concepts are the key to the surprises we witnessed within the last 5 years.

References

[1] H. Akaike, "A new look at the statistical model identification", *IEEE Trans. Automat. Control*, Vol. 19, No. 6, December 1974, pp. 716–723.

[2] L.A. Baccala and S. Roy, "A new blind time-domain channel identification method based on cyclostationarity", *Proc. 26th Conf. Informations Sciences and Systems*, Princeton, NJ, March 1994.

[3] E.W. Bai and S.S. Sastry, "Persistency of excitation, sufficient richness and parameter convergence in discrete time adaptive control", *Systems Control Lett.*, Vol. 6, 1985, pp. 153–163.

[4] A. Benveniste and M. Goursat, "Blind equalizers", *IEEE Trans. Commun.*, Vol. 32, No. 8, August 1984, pp. 871–883.

[5] A. Benveniste, M. Goursat and G. Ruget, "Robust identification of a nonminimum phase system: Blind adjustment of a linear equalizer in data communications", *IEEE Trans. Automat. Control*, Vol. AC-25, June 1980, pp. 385–399.

[6] Y. Bresler, "Exact maximum likelihood parameter estimation of superimposed exponential signals in noise", *IEEE Trans. Acoust. Speech Signal Process.*, Vol. 5, 1986, pp. 1081–1089.

[7] L.R. Cerrato and B.A. Eisenstein, "Deconvolution of cyclostationary signals", *IEEE Trans. Acoust. Speech Signal Process.*, Vol. ASSP-25, December 1977, pp. 466–476.

[8] Y. Chen and C.L. Nikias, "New results with fractionally-spaced blind equalization algorithms", *Proc. IEEE ICC'544*, New Orleans, LA, 1994, pp. 554–588.

[9] Z. Ding, "Blind identification and equalization using spectral measures: Part I", in: W.A. Gardner, ed., *Cyclostationarity in Communications and Signal Processing*, IEEE Press, New York, 1993.

[10] Z. Ding, R.A. Kennedy, B.D.O. Anderson and C.R. Johnson, "Ill-convergence of Godard blind equalizers in data communication systems", *IEEE Trans. Commun.*, Vol. 39, September 1991, pp. 1313–1327.

[11] Z. Ding and Y. Li, "Channel identification using second order cyclic statistics", *Proc. Asilomar Conf. on Signal, Systems, and Computers*, Pacific Grove, CA, October 1992.

[12] W.A. Gardner, *Introduction to Random Processes with Application to Signals and Systems*, Macmillan, New York, 1985.

[13] W.A. Gardner, "A new method of channel identification", *IEEE Trans. Commun.*, Vol. 39, No. 6, June 1991, pp. 813–817.

[14] G.B. Giannakis, "Linear cyclic correlation approach for blind identification of FIR channels", *Proc. 28th Asilomar Conf. on Signals, Systems, and Comp.*, Pacific Grove, CA, November 1994, pp. 420–423.

[15] G.B. Giannakis and S.D. Halford, "Performance analysis in blind equalizers based on cyclostationary statistics", *Proc. 26th Conf. Information Sciences and Systems*, Princeton, NJ, March 1994, pp. 711–716.

[16] G.B. Giannakis and S.D. Halford, "Blind fractionally-spaced equalization of noisy FIR channels: Adaptive and optimal solutions", *Proc. Internat. Conf. Acoust. Speech Signal Process. '95*, Detroit, MI, May 1995.

[17] R.D. Gitlin and S.B. Weinstein, "Fractionally-spaced equalization: An improved digital transversal equalizer", *Bell Syst. Tech. J.*, Vol. 60, February 1981, pp. 275–296.

[18] D.N. Godard, "Self-recovering equalization and carrier tracking in two-dimensional data communication systems", *IEEE Trans. Commun.*, Vol. 28, No. 11, November 1980, pp. 1867–1875.

[19] M.İ. Gürelli and C.L. Nikias, "A new eigenvector-based algorithm for multichannel blind deconvolution of input colored signals", *Proc. Internat. Conf. Acoust. Speech Signal Process. '93*, Vol. IV, Minneapolis, MN, April 1993, pp. 448–451.

[20] D. Hatzinakos and C. Nikias, "Estimation of multipath channel response in frequency selective channels", *IEEE J. Select. Areas Commun.*, Vol. S7, No. 1, January 1989, pp. 12–19.

[21] Y. Hua, "Fast maximum likelihood for blind identification of multiple FIR channels", *IEEE Trans. Signal Processing*, To appear.

[22] Y. Hua, "Fast maximum likelihood for blind identification of multiple FIR channels", *Proc. 28th Asilomar Conf. on Signals, Systems and Computers*, Pacific Grove, CA, November 1994.

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

[24] R. Kumaresan, L.L. Scharf and A.K. Shaw, "An algorithm for pole-zero modelling and spectral analysis", *IEEE Trans. Acoust. Speech Signal Process.*, Vol. 34, 1986, pp. 637–640.

[25] S.Y. Kung, T. Kailath and M. Morf, "A generalized resultant matrix for polynomial matrices", *Proc. IEEE Conf. on Decision and Control*, Florida, 1976, pp. 892–895.

[26] Y. Li and Z. Ding, "Blind channel identification based on second-order cyclostationary statistics", *Proc. Internat. Conf. Acoust. Speech Signal Process. '93*, Minneapolis, MN, April 1993, pp. 81–84.

[27] H. Liu and G. Xu, "A deterministic approach to blind symbol estimation", *IEEE Signal Processing Lett.*, Vol. 1. No. 12, December 1994, pp. 205–207.

[28] H. Liu and G. Xu, "Closed-form blind symbol estimation in digital communications", *IEEE Trans. Signal Processing*, Vol. SP-43, No. 11, November 1995, pp. 2714–2723.

[29] H. Liu and G. Xu, "Smart antennas in wireless systems: Multiuser blind channels and sequences identification", *IEEE Trans. Commun.*, Accepted.

[30] H. Liu, G. Xu and L. Tong, "A deterministic approach to blind equalization", *Proc. 27th Asilimar Conf. on Signals, Systems and Computers*, Pacific Grove, CA, November 1993, pp. 751–755.

[31] L. Ljung, *System Identification*, Prentice-Hall, Englewood Cliffs, NJ, 1987.

[32] E. Moulines, P. Dohamel, J. Cardoso and S. Mayrargue, "Subspace methods for the blind identification of multichannel FIR filters", *IEEE Trans. Signal Processing*, Vol. SP-43, No. 2, February 1995, pp. 516–525.

[33] B. Ottersten and M. Viberg, "Analysis of subspace fitting based methods for sensor array processing", *Proc. IEEE Internat. Conf. Acoust. Speech Signal Process. '89*, Glasgow, UK, May 1989, pp. 2807–2810.

[34] A.P. Petropulu and C.L. Nikias, "Blind deconvolution using signal reconstruction from partial higher order cepstral information", *IEEE Trans. SP*, Vol. 41, No. 6, June 1993, pp. 2088–2095.

[35] S.U. Pillai, *Array Signal Processing*, Springer, New York, 1989.

[36] B. Porat, *Digital Processing of Random Signals*, Prentice Hall, Englewood Cliffs, NJ, 1993.

[37] J. Rissanen, "Modeling by shortest data description", *Automatica*, Vol. 14, 1978, pp. 465–471.

[38] Y. Sato, "A method of self-recovering equalization for multilevel amplitude-modulation", *IEEE Trans. Commun.*, Vol. 23, No. 6, June 1975, pp. 679–682.

[39] S.V. Schell, D.L. Smith and S. Roy, "Blind channel identification using subchannel response matching", *Proc. 1994 Conf. Information Sciences and Systems*, Princeton, NJ, March 1994.

[40] R.O. Schmidt, "Multiple emitter location and signal parameter estimation", *Proc. RADC Spectral Estimation Workshop*, Griffiss AFB, NY, 1979, pp. 243–258.

[41] O. Shalvi and E. Weinstein, "New criteria for blind deconvolution of nonminimum phase systems (channels)", *IEEE Inform. Theory*, Vol. 36, No. 2, March 1990, pp. 312–320.

[42] M. Shao and C.L. Nikias, "An ML/MMSE approach to blind equalization", *Proc. IEEE Internat. Conf. Acoust. Speech Signal Process. '94*, April 1994, pp. IV569–IV572.

[43] D.T.M. Slock, "Blind fractionally-spaced equalization, perfect-reconstruction filter banks and multichannel linear prediction", *Proc. IEEE Internat. Conf. Acoust. Speech Signal Process. '94*, 30 April 1994, pp. IV585–IV588.

[44] D.T.M. Slock, "Blind joint equalization of multiple synchronous mobile users using oversampling and/or multiple antennas", *Proc. 28th Asilomar Conf. on Signals, Systems and Computers*, Pacific Grove, CA, November 1994, pp. 1154–1158.

[45] D.T.M. Slock and C.B. Papadias, "Further results on blind identification and equalization of multiple FIR channels", *Proc. IEEE Internat. Conf. Acoust. Speech Signal Process. '95*, April 1995, pp. 1964–1967.

[46] P. Stoica and A. Nehorai, "Performance study of conditional and unconditional direction of arrival estimation", *IEEE Trans. Acoust. Speech Signal Process.*, Vol. ASSP-38, No. 10, October 1990, pp. 1783–1795.

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

[48] L. Tong, G. Xu, B. Hassibi and T. Kailath, "Blind identification and equalization based on second-order statistics: A frequency domain approach", *IEEE Trans. Inform. Theory*, Vol. 41, No. 1, January 1995, pp. 329–333.

[49] L. Tong, G. Xu and T. Kailath, "A new approach to blind identification and equalization of multipath channel", *Proc. 25th Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, November 1991.

[50] L. Tong, G. Xu and T. Kailath, "Blind equalization of multipath fading channels", *Proc. 1992 Internat. Conf. on Communication*, Chicago, IL, June 1992, pp. 1513–1517.

[51] L. Tong, G. Xu and T. Kailath, "Blind identification and equalization using spectral measures, Part II: A time domain approach", in: W.A. Gardner, ed., *Cyclostationarity in Communications and Signal Processing*, IEEE Press, New York, 1993.

[52] J.K. Tugnait, "Identification of linear stochastic system via second and fourth-order cumulant matching", *IEEE Trans. Inform. Theory*, Vol. IT-33, May 1987, pp. 393–407.

[53] J.K. Tugnait, "On blind identifiability of multipath channels using fractional sampling and second-order cyclostationary statistics", *Proc. Global Telecom. Conf.*, 1993, pp. 2000–2005.

[54] G. Ungerboeck, "Fractional tap-spacing equalizer and consequences for clock recovery in data modems", *IEEE Trans. Commun.*, Vol. COM-24, August 1976, pp. 856–864.

[55] M. Viberg, B. Ottersten and T. Kailath, "Direction of arrival estimation and detection using weighted subspace fitting", *Proc. 23rd Asilomar Conf. on Signals, Systems and Computers*, Pacific Grove, CA, November 1989.

[56] M. Wax, Detection and estimation of superimposed signals, Ph.D. Thesis, Stanford University, Stanford, CA, 1985.

[57] G. Xu, L. Tong and H. Liu, "A new algorithm for fast blind equalization of wireless communication channels", *Proc. Internat. Conf. Acoust. Speech Signal Process. '94*, Adelaide, Australia, April 1994.

[58] G. Xu, H. Liu, L. Tong and T. Kailath, "A least-squares approach to blind channel identification", *IEEE Trans. Signal Processing*, Vol. SP-43, No. 12, December 1995.

[59] J. Yang and A. Swindlehurst, "DF directed multipath equalization", *Proc. 28th Asilomar Conf. on Signals, Systems and Computers*, Pacific Grove, CA, November 1994, pp. 1418–1422.

[60] D. Yellin and B. Porat, "Blind identification of FIR systems excited by discrete-alphabet inputs", *IEEE Trans. Signal Processing*, Vol. SP-41, No. 3, 1993, pp. 1331–1339.

[61] H. Zeng and L. Tong, "Some new results on blind channel estimation: performance and algorithms", *Proc. 29th Annual Conf. on Information Sciences and Systems*, March 1995.

[62] L.C. Zhao, P.R. Krishnaiah and Z.D. Bai, "Remarks on certain criteria for detection of number of signals", *IEEE Trans. Acoust. Speech Signal Process.*, Vol. 35, No. 2, February 1987, pp. 129–132.