Single-User Channel Estimation and Equalization

In digital communications, data signals are transmitted through linearly distortive analog channels such as telephone, cable, and wireless radio. In general, the single-user system model is an accurate description of point-to-point, time division multiple access (TDMA), and frequency division multiple access (FDMA) communication systems. Two major sources of linear channel distortion in (single-user) digital communications systems are multipath propagation and limited bandwidth. Linear channel distortion leads to intersymbol interference (ISI) at the receiver which, in turn, may lead to high error rates in symbol detection. Equalizers are designed to compensate for these channel distortions. One may directly design an equalizer given the received signal, or one may first estimate the channel impulse response and then design an equalizer based on the estimated channel. Traditionally, receivers or equalizers rely on a transmitter assisted training session to extract the desired reference signal for channel estimation and equalization. Such receivers continue to be highly important research subjects because of practical obstacles such as channel variation and nonlinearity.

More recently, there has been much interest in blind (self-recovering) channel estimation and blind equalization where no training sequences are available or used. In multipoint networks, whenever a link from the server to one of the tributary stations is interrupted, it is clearly not feasible (or desirable) for the server to start sending a training sequence to re-establish a particular link. In digital communications over fading/multipath channels, a restart is required following a temporary path interruption due to severe fading. During on-line transmission impairment monitoring, the training sequences are obviously not supplied by the transmitter. Consequently, the importance of blind channel compensation research is also strongly supported by practical needs.

In this article, we present a comprehensive summary of recent research development on single-user channel estimation and equalization, focusing on both training-based and blind approaches. Our emphasis is on linear time-invariant channels; linear time-varying as well as nonlinear channels are outside the scope of this article.

System Models

In this section we first describe the models that are used to characterize the wireless and mobile communications channels. Then we turn to a brief discussion of the various equalizer structures that are used to undo the signal distortions caused by the channel.

Channel Models

The propagation of signals through wireless channels (indoors or outdoors) results in the transmitted signal arriving at the receiver through multiple paths. These paths arise due to reflection, refraction, or diffraction in the channel. Multipath propagation results in a received signal that is a superposition of several delayed and scaled copies of the transmitted signal giv-

1053-5888/00/$10.00©2000IEEE

While the ML estimator is conceptually simple and it usually has good performance when the sample size is sufficiently large, the implementation of the ML estimator is sometimes computationally intensive.

ing rise to frequency-selective fading. Frequency-selective fading (defined as changes in the received signal level in time) is caused by destructive interference among multiple propagation paths. The environment around the transmitter and the receiver can change over time, particularly in a mobile setting, leading to variations in the channel response with time. This gives rise to time-selective fading. Also, the channels may have a dominant path (direct path in line-of-sight channels) in addition to several secondary paths, or they may be characterized as having multiple "random" paths with no single dominant path.

Multipath propagation leads to ISI at the receiver which, in turn, may lead to high error rates in symbol detection. Equalizers are designed to compensate for these channel distortions. One may directly design an equalizer given the received signal, or one may first estimate the channel impulse response and then design an equalizer based on the estimated channel. After some processing (matched filtering, for instance), the continuous-time received signals are sampled at the baud (symbol) or higher (fractional) rate before processing them for channel estimation and/or equalization. It is therefore convenient to work with a baseband-equivalent discrete-time channel model. Consider a baud-rate sampled system. Let $s[k]$ denote the kth information symbol, and let $r[k]$ denote the sampled received signal during the kth received symbol. Then the two are related via a time-varying linear system response as

$$r[n] = \sum_k h[n;k]s[n-k] + w[n] \tag{1}$$

where $h[n;k]$ is the channel response at time n to a unit input at time $n-k$ and $w[n]$ represents the additive noise (and interferences) at the receiver. Model (1) represents a time- and frequency-selective linear channel. A tapped delay line structure for this model is shown in Fig. 1. For a slowly (compared to the baud rate) time-varying system, one often simplifies (1) to a time-invariant system as

$$r[n] = \sum_k h[k]s[n-k] + w[n] \tag{2}$$

where $h[k] = h[0;k]$ is the time-invariant channel response to a unit input at time 0. Model (2) represents a frequency-selective linear channel with no time selectivity. It is the most commonly used model for receiver design.

Suppose that $h[n;k] = h[n]\delta[k,0]$ where $\delta[k,0]$ is the Kronecker delta located at 0, i.e., $\delta[k,0]=1$ for $k=0$ and $\delta[k,0]=0$ for $k \neq 0$. Then we have the time-selective and frequency-nonselective channel whose output is given by

$$r[n] = h[n]s[n] + w[n]. \tag{3}$$

Finally, a time-nonselective and frequency-nonselective channel is modeled as

$$r[n] = hs[n] + w[n] \tag{4}$$

where h is a random variable (or a constant).

All of the channel response functions (1)-(4) may be modeled as deterministic or random. Also, (1)-(4) result in a single-input single-output (SISO) complex discrete-time baseband-equivalent channel model. When the channel of (2) is deterministic, the output sequence $\{r[n]\}$ is discrete-time stationary. When there is excess channel bandwidth [bandwidth $> 1/2\times$ (baud rate)], baud rate sampling is below the Nyquist rate leading to aliasing and depending upon the symbol timing phase, in certain cases, causing deep spectral notches in the sampled, aliased channel transfer function [3]. Linear equalizers designed on the basis of the baud-rate sampled channel response are quite sensitive to symbol timing errors. Initially, in the trained case, fractional sampling was investigated to robustify the equalizer performance against timing errors. For linear time-invariant frequency-selective deterministic channels [as in (2)], when sampled at higher than the baud rate (typically an integer multiple, p, of baud rate), the sampled signal is discrete-time scalar cyclostationary, and equivalently, it may be represented as a discrete-time vector stationary sequence with an underlying single-input multiple-output (SIMO) model where we stack p consecutive received samples in the nth symbol duration to form a p-vector $\vec{r}[n]$:

$$\vec{r}[n] = \sum_k \vec{h}[k]s[n-k] + \vec{w}[n]. \tag{5}$$

In (5), $\vec{h}[k]$ is a p vector.

For more details on fading multipath channels, see [4] and [5]. For modeling saturation nonlinearities of power amplifiers, nonlinear channels of Volterra type have also been used [1]. A discussion of basis expansion models for time-varying channels may be found in [2] where, by a suitable selection of the basis functions, a time-varying channel can be "transformed" into a time-invariant channel.

Equalizer Structures

The most common channel equalizer structure is a linear transversal filter. Given the baud-rate sampled received signal [see (2)] $r[n]$, the linear transversal equalizer output $y[n]$ is an estimate of $s[n]$, given by

$$y[n]=\sum_{k=-N}^{N} c[k]r[n-k] \tag{6}$$

where $\{c[k]\}$ are the $(2N+1)$ tap weight coefficients of the equalizer; see Fig. 2. As noted earlier, linear equalizers designed on the basis of the baud-rate sampled received signal are quite sensitive to symbol timing errors [3]. Therefore, fractionally spaced linear equalizers (typically with twice the baud-rate sampling: oversampling by a factor of two) are quite widely used to mitigate sensitivity to symbol timing errors. A fractionally spaced equalizer (FSE) in the linear transversal structure has the output

$$y[n]=\sum_{k=-N}^{N} \vec{c}^{\,T}[k]\vec{r}[n-k] \tag{7}$$

where we have p samples per symbol, $\vec{r}[n]$ and $\vec{c}[n]$ are p-column vectors [cf. (5)], $\{\vec{c}[n]\}$ are the $(2N+1)$ vector tap [or $p(2N+1)$ scalar tap] weight coefficients of the FSE, and the superscript T denotes the transpose operation. Note that the FSE outputs data at the symbol rate. Various criteria and cost functions exist to design the linear equalizers in both batch and recursive (adaptive) form: we discuss these later. The linear equalizers can also be implemented as a lattice filter [4]. Lattice equalizers exhibit faster convergence and better numerical properties [4].

Linear equalizers do not perform well when the underlying channels have deep spectral nulls in the passband. Several nonlinear equalizers have been developed to deal with such channels. Two effective approaches are:

- *Decision Feedback Equalizer (DFE):* DFE is a nonlinear equalizer that employs previously detected symbols to eliminate the ISI due to the previously detected symbols on the current symbol to be detected. The use of the previously detected symbols makes the equalizer output a nonlinear function of the data. DFE can be symbol-spaced or fractionally spaced.
- *Maximum-Likelihood Sequence Detector:* This estimates the information sequence to maximize the joint probability of the received sequence conditioned on the information sequence.

A detailed discussion may be found in [4].

Channel Estimation

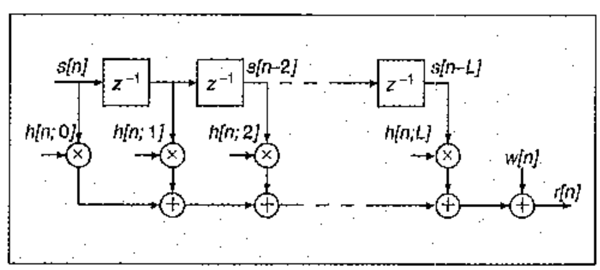
One of the objectives of receiver design is to minimize the detection error. In general, the design of optimal detector requires the knowledge of the channel. Often unknown in practice, channel parameters need to be estimated, preferably using only a limited amount of data samples. In communication applications, especially for packet transmissions, the efficiency (a measure of how effectively an algorithm utilizes the available data) of the estimator is particularly important.

We consider in this section three types of channel estimators based on the framework of maximizing the likelihood function. Referred to as the training-based channel estimation, the first type described consists of the classical techniques that estimate the channel from a known training sequence and its corresponding observation. The mode of operation is "train-before-transmit," which is effective when the channel does not have significant time variations as in the case of voiceband communication over telephone channels. For rapidly varying channels, however, such an approach is not efficient because training has to be performed repeatedly, which reduces the available time for transmitting information. Next we describe the approach of blind channel estimation, which means that the channel estimation is performed while information signals are being transmitted. In other words, the goal of blind channel estimation is "train-while-transmit." The major advantage of these techniques is the improved bandwidth utilization for time-varying channels. Finally, we consider the class of techniques that fall in between the training-based and blind channel estimation techniques. Referred to as the semiblind channel estimation, these techniques aim to estimate the channel using not only the known data in the transmitted signal and its corresponding observation, but also the observation corresponding to the unknown data. The semiblind channel estimation becomes training-based estimation when only the observation corresponding to the known data is used, and it becomes blind-channel estimation when the observation is restricted to that corresponding to the unknown part. Semiblind channel estimation is motivated by the fact that, in data transmission, there are always some known symbols that should be incorporated to improve the performance.

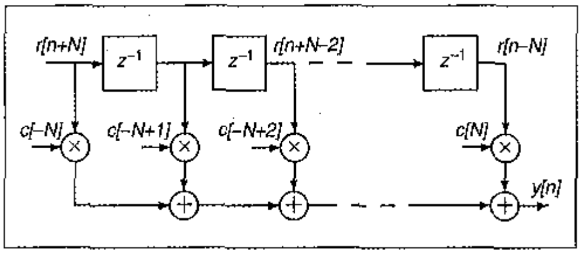
The Maximum-Likelihood Estimator

One of the most popular parameter estimation algorithms is the maximum-likelihood (ML) method. The ML estimators can be derived in a systematic way. Perhaps more importantly, the class of ML estimators are optimal asymptotically [7].

Let us consider the p-vector channel model given in (5) where we now assume that the channel has a finite im-



▲ 1. Tapped delay line model of frequency and time selective channel with finite impulse response. z^{-1} represents a unit (symbol duration) delay.



▲ *2. Structure of a baud-rate linear transversal equalizer.*

pulse response of order L. Suppose that we have collected N samples of the observation $\vec{r} = [\vec{r}[N-1],...,\vec{r}[0]]^T$. We then have the following linear model:

$$\vec{r} = \begin{pmatrix} s[N-1]I_p & s[N-2]I_p & \cdots & s[N-L-1]I_p \\ \vdots & \textit{Block} & \textit{Hankel} & \textit{Matrix} \\ s[0]I_p & s[-1]I_p & \cdots & s[-L]I_p \end{pmatrix} \cdot \begin{pmatrix} \vec{h}[0] \\ \vdots \\ \vec{h}[L] \end{pmatrix} + \begin{pmatrix} \vec{w}[N-1] \\ \vdots \\ \vec{w}[0] \end{pmatrix} = \mathcal{H}(\vec{s})_{N\times(L+1)}\vec{h} + \vec{w} \tag{8}$$

where I_p is a $p \times p$ identity matrix, $\vec{s}$ and $\vec{w}$ are vectors consisting of samples of the input sequence $s[n]$ and noise $\vec{w}[n]$, respectively, and $\vec{h}$ is the vector of the channel parameters.

Let $\vec{\theta}$ be the vector of unknown parameters that may include the channel parameter $\vec{h}$ and possibly the entire or part of the input vector $\vec{s}$. Given the probability space that describes jointly the noise vector $\vec{w}$ and possibly the input data vector $\vec{s}$, we can then obtain, in principle, the probability density function (pdf)—assuming it exists—of the observation $\vec{r}$. As a function of the unknown parameter $\vec{\theta}$, the pdf of the observation $f(\vec{r};\vec{\theta})$ is referred to as the likelihood function. The ML estimator is defined by the following optimization:

$$\hat{\vec{\theta}} = \arg\max_{\vec{\theta}\in\Theta} f(\vec{r};\theta) \tag{9}$$

where Θ defines the domain of the optimization.

While the ML estimator is conceptually simple and it usually has good performance when the sample size is sufficiently large, the implementation of ML estimator is sometimes computationally intensive. Furthermore, the optimization of the likelihood function in (9) is often hampered by the existence of local maxima. Therefore, it is desirable that effective initialization techniques are used in conjunction with the ML estimation.

We now apply the principle of maximizing the likelihood function to the three channel estimation problems: the training-based channel estimation, the blind channel estimation, and the semiblind channel estimation.

Training-Based Channel Estimation

The training-based channel estimation assumes the availability of the input vector $\vec{s}$ (as training symbols) and its corresponding observation vector $\vec{r}$. When the noise samples are zero mean, white Gaussian, i.e., $\vec{w}$ is a zero mean, Gaussian random vector with covariance $\sigma^2 I$, the ML estimator defined in (10), with $\vec{\theta} = \vec{h}$, is given by

$$\hat{\vec{h}} = \arg\min_{\vec{h}} \left\| \vec{r} - \mathcal{H}(\vec{s})\vec{h} \right\|^2 = \mathcal{H}^{\dagger}(\vec{s})\vec{r} \tag{10}$$

where $\mathcal{H}^{\dagger}(s)$ is the pseudo-inverse of the $\mathcal{H}(\vec{s})$ defined in (8). This is also the classical linear least squares estimator which can be implemented recursively, and it turns out to be the best (in terms of having minimum mean square error) among all unbiased estimators and it is the most efficient in the sense that it achieves the Cramér-Rao lower bound. Various adaptive implementations can be found in [17].

Blind Channel Estimation

Now suppose that both the input vector $\vec{s}$ and the channel vector $\vec{h}$ are unknown. The simultaneous estimation of the input vector and the channel appears to be ill-posed; how is it possible that the channel and its input can be distinguished using only the observation? The key in blind channel estimation is the utilization of qualitative information about the channel and the input. To this end, we consider two different types of ML techniques based on different models of the input sequence.

Stochastic Maximum-Likelihood Estimation

While the input vector $\vec{s}$ is unknown, it may be modeled as a random vector with a known distribution. In such a case, the likelihood function of the unknown parameter $\vec{\theta} = \vec{h}$ can be obtained by

$$f\left(\vec{r};\vec{h}\right) = \int f\left(\vec{r}\,|\,\vec{s};\vec{h}\right) f(\vec{s})\,d\vec{s} \tag{11}$$

where $f(\vec{s})$ is the marginal pdf of the input vector and $f(\vec{r}|\vec{s};\vec{h})$ is the likelihood function when the input is known. Assume, for example, that the input data symbol $s[n]$ takes, with equal probability, a finite number of values. Consequently, the input data vector $\vec{s}$ also takes values from the signal set $\{\vec{s}_1,...,\vec{s}_K\}$. The likelihood function of the channel parameter is then given by

$$f\left(\vec{r};\vec{h}\right) = \sum_{i=1}^{K} f\left(\vec{r}\,\middle|\,\vec{s}_i;\vec{h}\right) Prob(\vec{s} = \vec{s}_i) = C\sum_{i=1}^{K} \exp\left\{-\frac{\left\|\vec{r} - \mathcal{H}(\vec{s}_i)\vec{h}\right\|^2}{2\sigma^2}\right\} \tag{12}$$

where C is a constant, and the stochastic ML estimator is given by

$$\hat{\vec{h}} = \arg\min_{\vec{h}} \sum_{i=1}^{K} \exp\left\{ -\frac{\left\| \vec{r} - \mathcal{H}(\vec{s}_i)\vec{h} \right\|^2}{2\sigma^2} \right\}. \tag{13}$$

The maximization of the likelihood function defined in (11) is in general difficult because $f(\vec{r};\theta)$ is nonconvex. The expectation-maximization (EM) algorithm [6], [12] can be applied to transform the complicated optimization to a sequence of quadratic optimizations. Kaleh and Vallet [18] first applied the EM algorithm to the equalization of communication channels with input sequence having finite alphabet property. By using a hidden Markov model (HMM) model, they developed a batch (off-line) procedure that includes the so-called forward and backward recursions [21]. Unfortunately, the complexity of this algorithm increases exponentially with the channel memory.

To relax the memory requirements and facilitate channel tracking, "on-line" sequential approaches have been proposed in [30], [34], and [29] for general input and in [19] for input with finite alphabet properties under a HMM formulation. Given the appropriate regularity conditions [29] and a good initialization guess, it can be shown that these algorithms converge (almost surely and in the mean square sense) to the true channel value.

Deterministic Maximum-Likelihood Estimation

The deterministic ML approach assumes no statistical model for the input sequence $s[n]$. In other words, both the channel vector $\vec{h}$ and the input source vector $\vec{s}$ are parameters to be estimated. When the noise is zero-mean Gaussian with covariance $\sigma^2 I$, the ML estimates can be obtained by the nonlinear least squares optimization

$$\left\{ \hat{\vec{h}}, \hat{\vec{s}} \right\} = \arg\min \left\| \vec{r} - \mathcal{H}(\vec{s})\vec{h} \right\|_2^2. \tag{14}$$

The joint minimization of the likelihood function with respect to both the channel and the source parameter spaces is difficult. Fortunately, the observation vector $\vec{r}$ is linear in both the channel and the input parameters individually. In particular, we have

$$\vec{r} = \mathcal{H}(\vec{s})\vec{h} + \vec{w} = \mathcal{T}\left(\vec{h}\right)\vec{s} + \vec{w} \tag{15}$$

where

$$\mathcal{T}\left(\vec{h}\right) = \begin{pmatrix} \vec{h}[0] & \cdots & \vec{h}[L] & & \\ & \ddots & & \ddots & \\ & & \vec{h}[0] & \cdots & \vec{h}[L] \end{pmatrix} \tag{16}$$

is the so-called filtering matrix. We therefore have a separable nonlinear least squares problem that can be solved sequentially

$$\left\{ \hat{\vec{h}}, \hat{\vec{s}} \right\} = \arg\min_{\vec{s}} \left\{ \min_{\vec{h}} \left\| \vec{r} - \mathcal{H}(\vec{s})\vec{h} \right\|_2^2 \right\} \tag{17}$$

$$= \arg\min_{\vec{h}} \left\{ \min_{\vec{s}} \left\| \vec{r} - \mathcal{T}(\vec{h})\vec{s} \right\|_2^2 \right\}. \tag{18}$$

If we are only interested in estimating the channel, the above minimization can be rewritten as

$$\hat{\vec{h}} = \arg\min_{\vec{h}} \left\| \underbrace{\left(I - \mathcal{T}(\vec{h})\mathcal{T}^{\dagger}(\vec{h}) \right)}_{\mathcal{P}(\vec{h})} \vec{r} \right\|^2 = \arg\min_{\vec{h}} \left\| \mathcal{P}(\vec{h})\vec{r} \right\|^2 \tag{19}$$

where $\mathcal{P}(\vec{h})$ is a projection transform of $\vec{r}$ into the orthogonal complement of the range space of $\mathcal{T}(\vec{h})$, or the noise subspace of the observation. Discussions of algorithms of this type can be found in [31].

The finite alphabet properties of the input sequence, similar to the HMM for statistical ML approach, can also be incorporated into the deterministic ML methods. These algorithms, first proposed by Seshadri [22] and Ghosh and Weber [15], iterate between estimates of the channel and the input. At iteration k, with an initial guess of the channel $\vec{h}^{(k)}$, the algorithm estimates the input sequence $\vec{s}^{(k)}$ and the channel $\vec{h}^{(k+1)}$ for the next iteration by

$$\vec{s}^{(k)} = \arg\min_{\vec{s} \in \mathcal{S}} \left\| \vec{r} - \mathcal{T}\left(\vec{h}^{(k)}\right)\vec{s} \right\|^2 \tag{20}$$

$$\vec{h}^{(k+1)} = \arg\min_{\vec{h}} \left\| \vec{r} - \mathcal{H}\left(\vec{s}^{(k)}\right)\vec{h} \right\|^2 \tag{21}$$

where $\mathcal{S}$ is the (discrete) domain of $\vec{s}$. The optimization in (21) is a linear least squares problem whereas the optimization in (20) can be achieved by using the Viterbi algorithm [13]. The convergence of such approaches is not guaranteed in general.

The Method of Moments

Although the ML channel estimator usually provides better performance, the computation complexity and the existence of local optima are the two major difficulties. The method of moments, on the other hand, often has a closed-form identification by exploiting the relationship between the channel parameter and moments of the observation vector $\vec{r}$.

Second-Order Statistical Methods: In general, the second-order moment of the observation carries only the magnitude information of the channel. It is therefore insufficient for channel identification. For SIMO vector channels, however, the autocorrelation function of the observation is sufficient for the identification of the channel impulse response up to an unknown constant [32]. This observation led to a number of techniques under both statistical and deterministic assumptions of the input sequence [31]. By exploiting the multichannel aspects of the channel, many of these techniques lead to a constrained quadratic optimization

$$\hat{\vec{h}} = \arg\min_{\|\vec{h}\|=1} \vec{h}^t Q(\vec{r})\vec{h} \tag{22}$$

where $Q(\vec{r})$ is a positive definite matrix constructed from the observation. Asymptotically (either as the sample size increases to infinity or the noise variance approaches to zero), these estimates converge to true channel parameters.

Here we present a simple yet informative approach [35] that illustrates the basic idea. Suppose that we have only two channels with finite impulse responses $h_1[n]$ and $h_2[n]$, respectively. If there is no noise, the received signals from the two channels satisfy

$$r_1[n] = h_1[n] * s[n], \qquad r_2[n] = h_2[n] * s[n] \tag{23}$$

where * is the linear convolution. Consequently, we must have

$$r_1[n] * h_2[n] = r_2[n] * h_1[n]. \tag{24}$$

Since the convolution operation is linear with respect to the channel and $r_i[n]$ is available, (24) is equivalent to solving a homogeneous linear equation

$$R\vec{h} = 0 \tag{25}$$

where R is a matrix made of observations from the two channels. It can be shown that under certain identifiability conditions [31], the null space of R has dimension 1, which means that the channel can be identified up to a constant. When there is noise, the channel estimator can be obtained from a constrained quadratic optimization

$$\hat{\vec{h}} = \arg\min_{\|\vec{h}\|=1} \vec{h}^t R^t R\vec{h} \tag{26}$$

which implies that $\hat{\vec{h}}$ is the eigenvector corresponding to the smallest eigenvalue of $Q = R^t R$.

Alternatively, one can also exploit the subspace structure of the filtering matrix. For example, if it is possible to construct a matrix N, from data directly, such that

$$NT\left(\vec{h}\right) = 0 \tag{27}$$

due to the structure of $T(\vec{h})$, we then have

$$\mathcal{G}(N)\vec{h} = 0 \Rightarrow \hat{\vec{h}} = \arg\min_{\|\vec{h}\|=1} \vec{h}^t \left(\mathcal{G}(N)\mathcal{G}^t(N)\right)\vec{h}. \tag{28}$$

One such subspace technique was presented in [24].

More recently, the problem of blind channel identification has been formulated as problems of linear prediction [23], [14], [66] and smoothing [33] which have simple adaptive implementations [36].

Higher-Order Statistical (HOS) Methods: Given the mathematical model, there are two broad classes of approaches to channel estimation, the distinguishing feature among then being the choice of the optimization criterion. All of the approaches involve (more or less) a least-squares error measure. The error definition differs, however, as follows:

- *Fitting Error:* Match the model-based higher-order (typically fourth-order) statistics to the estimated (data-based) statistics in a least-squares sense to estimate the channel impulse response, as in [64] and [65], for example. This approach allows consideration of noisy observations. In general, it results in a nonlinear optimization problem. It requires availability of a good initial guess to prevent convergence to a local minimum. It yields estimates of the channel impulse response.
- *Equation Error:* It is based on minimizing an "equation error" in some equation which is satisfied ideally. The approaches of [69] and [68] (among others) fall in this category. In general, this class of approaches results in a closed-form solution for the channel impulse response so that a global extremum is always guaranteed provided that the channel length (order) is known. These approaches may also provide good initial guesses for the nonlinear fitting error approaches. Quite a few of these approaches fail if the channel length is unknown.

Further details may be found in [67] and references therein.

Semiblind Channel Estimation

Semiblind channel estimation has attracted considerable attention recently due to the need for fast and robust channel estimation and the fact that, for many packet transmission systems, there are embedded known symbols that can be exploited for channel estimation and tracking. We present here a brief discussion about the idea and refer the reader to a recent survey [9] for details.

Semiblind channel estimation assumes additional knowledge of the input sequence. Specifically, part of the input data vector is known. Both the statistical and deterministic ML estimators remain the same except that the likelihood function needs to be modified to incorporate the knowledge of the input [11], [10]. Semiblind channel estimation may offer significant performance improvement, however, over either the blind or the training-based methods as demonstrated in the evaluation of Cramér-Rao lower bound in [11].

There are many generalizations of blind channel estimation techniques to incorporate known symbols. In [8], Tsatsanis and Cirpan extended the approach of Kaleh and Vallet by restricting the transition of hidden Markov model. In [20], the knowledge of the known symbol is used to avoid the local maxima in the maximization of the likelihood function. A popular approach is to combine the objective function used to derive blind channel estimator with the least squares cost in the training-based channel estimation. For example, a weighted linear combination of the cost for blind channel estimator and that for the training-based estimator can be used [16].

Direct Equalization and Symbol Estimation

For the purpose of communicating, digital receivers need to recover channel input symbols from received signals that may suffer from noise and channel distortions. Direct channel equalization and symbol estimation are commonly adopted in practical systems. Recall that data communication input $s[k]$ comes from a known constellation S that has finite number of possible symbols. This important information forms the basis for many direct equalization and symbol estimation approaches to the channel equalization problem.

In this section, we describe several types of approaches to the problem of direct input signal recovery under linear time-invariant channels. First, we consider the classical approach of adaptive channel equalization based on training. This approach relies on an available sequence of training data that is transmitted by the transmitter during the setup stage and is known to the receiver. This training approach can be applied for T-spaced equalizers (TSE) as SISO feed-forward filters, for FSE as SIMO feed-forward filters, and for DFE. We then outline the basic principle of blind adaptive equalization based on implicit HOS criteria. Next, we explain the principle of some simple algorithms for blind symbol estimation exploiting second order statistics. Finally, we revisit the method of symbol estimation via iterative least square criterion and some variations.

Equalizer Adaptation Based on Training

Channel output (after matched filter) sampled at baud rate is

$$r[n] = \sum_{k=-\infty}^{\infty} h[k]s[n-k] + w[n]. \tag{29}$$

Problems occur when the original analog channel does not satisfy the Nyquist I criterion. Consequently, undesirable ISI is introduced as the channel output $r[n]$ depends on multiple symbols $\{s[n]\}$. ISI is usually caused by limited channel bandwidth, multipath, and channel fading. One of the simplest and most effect approach to recovering $s[n]$ from $r[n]$ is the use of linear channel equalization.

Following the successful application of adaptive filters by Lucky [37], equalization parameters are often updated through the minimum mean square error criteria. This requires that a known channel input sequence be transmitted initially. Equalization with training is common to many digital communication systems such as high speed telephone modem, satellite communication systems, and digital cellular systems.

The general structure of a channel equalizer is shown in Fig. 3. Adaptive channel equalizers begin adaptation with the assistance of a known training sequence transmitted during the initial stage by the transmitter. Since the input signal is available, adaptive algorithms can be used to adjust the equalizer parameters by minimizing a mean square error (MSE) between the equalizer output $y[n]$ and the known channel input with a delay $s[n-\nu]$. After training, equalizer parameters should be sufficiently close to the desired settings such that much of the ISI is removed. As the channel input can now be correctly recovered from the equalizer output through a memoryless decision device (slicer), the second stage of real data transmission can begin. In the operational stage, the receivers typically switch to a decision-directed mode where the equalized signal $y[n]$ is sent to a symbol detector and the detected symbols are used as a (pseudo-)training sequence to update equalizer coefficients. Feedforward TSE, FSE, as well as DFE can be updated. During either session, the equalizer filter parameters can be determined using the well known recursive least square (RLS) or least mean square (LMS) algorithm.

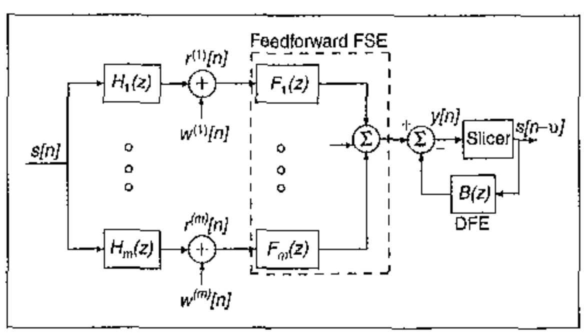
Maximum-Likelihood Sequence Estimation

Channel equalization followed by a symbol-by-symbol estimation slicer does not take into consideration the fact that the equalized noise is no longer white. Thus, performance loss is often encountered in feed-forward and feedback equalizers. A more effective but more costly approach is the use of Viterbi algorithm for ML estimation of the input sequence.

Assume that the SISO channel has finite impulse response

$$r[n] = \sum_{k=0}^{L} h[k]s[n-k] + w[n]. \tag{30}$$

As the noise $w[n]$ is often white Gaussian, ML estimate of the channel input $s[n]$ based on a sequence of channel output $r[n]$ can be obtained if the channel impulse response is known or has been estimated via training or blind channel estimation. The input sequence can be es-



▲ *3. Feedforward and decision feedback channel equalization filters.*

timated by maximizing the likelihood function or, equivalently, by minimizing

$$\sum_{n=L}^{\infty}\left|r[n]-\sum_{k=0}^{L}h[k]s[n-k]\right|^2. \quad (31)$$

Since $\mathcal{S}$ has only M symbols, the Viterbi algorithm can be implemented by denoting M^L states as all possible L-tuples of $(s[n],s[n-1],...,s[n-L+1])$. The trellis is determined by $\mathcal{S}$ while the metrics of the Viterbi algorithm depends on the estimated channel $h[k]$.

The ML Viterbi sequence estimator is optimum since it provides the minimum probability of symbol error under white Gaussian noise. It is a nonlinear equalizer, however, and is quite complex if the number of states M^L is large. The DFE can be considered as a suboptimum scheme that assumes all past decisions as correct and only estimates the most recent symbol. To obtain a method simpler than MLSE and yet more accurate than DFE, a reduced state Viterbi algorithm was proposed by Duel-Hallen and Heegard [38] that assumes some past decisions as correct while estimating several most recent symbols. This reduced state approach gives a nice compromise between complexity and performance. It provides good performance when the channel impulse response has long but small tails.

SISO Blind Equalization Based on HOS

In many communication systems, transmission of training sequences is either impractical or too costly. Blind adaptive channel equalization algorithms that do not rely on training signals have been developed. This property can be helpful in broadcast and multicast systems where training sequence for one new user can be disruptive to currently connected users.

Generally there are two types of approaches to this problem: blind channel estimation or direct blind equalization. Blind channel estimation issues have been discussed previously. Direct blind equalization seeks optimum parameter values for blind equalizer filters so that the eye pattern at the equalizer output is open to allow correct slicer decision. Because of the nonlinear nature of DFE, adaptive blind equalizers are generally implemented as feed-forward.

The key to designing a blind equalizer is to design rules of equalizer parameter adjustment. With the lack of training sequence, the receiver does not have access to the desired equalizer output $s[n]$ to adopt the traditional minimum mean square error criterion. Evidently, blind equalizer adaptation needs to minimize some special, non-MSE type cost function which implicitly involves higher order statistics of the channel output signal. The design of the blind equalizer thus translates into defining a mean cost function $E\{\Psi(y[n])\}$ where $\Psi(x)$ is a scalar function. Thus, the stochastic gradient descent minimization algorithm is easily determined by the derivative function $\psi(x)\triangleq\Psi'(x)$. Hence, a blind equalizer can either be defined by the cost function $\Psi(x)$ or, equivalently, by its derivative $\psi(x)$ function. Ideally, the function $\Psi(\cdot)$ should be selected such that local minima of the mean cost correspond to a significant removal of ISI in the equalizer output $y[n]$.

We summarize several blind adaptation algorithms designed for feed-forward equalizers.

Decision Directed Algorithm

The simplest blind equalization algorithm is the decision-directed algorithm without training sequence. It minimizes the mean square error between equalizer output $y[n]$ and the slicer output $\hat{s}[n-\nu]$. The performance of the decision-directed algorithm depends on how close the initial parameters are to their optimum settings. The closer they are, the more accurate the slicer output is to the true channel input $s[n-\nu]$. On the other hand, local convergence is highly likely if initial parameter values cause significant number of slicer errors [39], [40].

Sato Algorithm and Some Generalizations

The first truly blind algorithm was introduced by Sato [41]. For M-level PAM channel input, it is defined by

$$\psi(x)=x-R_1\operatorname{sgn}(x),\ \text{where}\ R_1\triangleq\frac{E|s[n]|^2}{E|s[n]|}. \quad (32)$$

The Sato algorithm was extended by Benveniste et al. [42] into a class of error functions given by

$$\psi_b(y[n])=\psi_a(y[n])-R_b\operatorname{sgn}(y[n]),$$
$$\text{where}\ R_b\triangleq\frac{E\{\psi_a(s_n)s_n\}}{E|s_n|}. \quad (33)$$

The generalization uses an odd function $\psi_a(x)$ whose second derivative is nonnegative for $x\geq 0$.

Stop-and-Go Algorithm

Another idea called the "stop-and-go" algorithm was introduced by Picchi and Prati [43] to allow adaptation "to go" only when several derivative functions agree in sign for the current output $y[n]$. Given several criteria for blind equalization, one can expect a more accurate descent direction when more than one of the existing algorithms agree on sign of their error functions. When the error signs differ for a particular output $y[n]$, parameter adaptation is "stopped" to maintain their current values. A similar idea was exploited in [53].

Bussgang Algorithm

The so-called Bussgang algorithms are derived from the maximum *a posteriori* (MAP) formulation [44], [45]. Define the impulse response of the channel-equalizer combination as $\zeta[k]=h[k]*c[k]$. If $\zeta\nu$ has the largest magnitude, then the equalizer output $y[k]$ is

$$y[n]=\sum_{i=0}^{\infty}\zeta[i]s[n-i]+w[n]$$
$$=\zeta[\nu]s[k-\nu]+\underbrace{\sum_{i\neq\nu}\zeta[i]s[n-i]+w[n]}_{\text{convolution noise}}.$$

Assuming that the probability distribution of noise is Gaussian, the MAP estimate of $s[n-\nu]$

$$\hat{s}[n-\nu]^{\text{MAP}}=\arg\max_{s[n-\nu]\in S} f(y[n]\,|\,s[n-\nu]) \tag{34}$$

can be used as a reference signal for LMS equalizer update in the Bussgang algorithm.

Constant Modulus Algorithm and Extension

The best known blind algorithms were presented in [46] and [48] with cost functions

$$\Psi_q(x)=\frac{1}{2q}\left(|x|^q-R_q\right)^2,$$
$$\text{where } R_q\triangleq\frac{E|s[n]|^{2q}}{E|s[n]|^q},\ q=1,2,\ldots \tag{35}$$

This class of Godard algorithms is indexed by the positive integer q. Using the stochastic gradient descent approach, equalizer parameters can be adapted accordingly.

For $q=2$, the special Godard algorithm was developed as the "constant modulus algorithm" (CMA) independently by Treichler and co-workers [48] using the philosophy of property restoral. For channel input signal that has a constant modulus $|s[n]|^2=R_2$, the CMA equalizer penalizes output samples $y[n]$ that do not have the desired constant modulus characteristics. The modulus error is simply $e(n)=|y[n]|^2-R_2$, and the squaring of this error yields the constant modulus cost function that is the identical to the Godard cost function with $q=2$.

This modulus restoral concept has a particular advantage in that it allows the equalizer to be adapted independent of carrier recovery. A carrier frequency offset of Δ_f causes a possible phase rotation of the equalizer output. Because the CMA cost function is insensitive to the phase of $y[n]$, the equalizer parameter adaptation can occur independently and simultaneously with the operation of the carrier recovery system. This property also allows CMA to be applied to analog modulation signals with constant amplitude such as those using frequency or phase modulation [48].

The methods of Shalvi-Weinstein [49] generalize CMA and are explicitly based on higher order statistics of the equalizer output. Define the kurtosis of the equalizer output signal $y[n]$ as

$$K_y\triangleq E|y[n]|^4-2E^2|y[n]|^2-\left|E\{y[n]^2\}\right|^2. \tag{36}$$

For the purpose of communicating, digital receivers need to recover channel input symbols from received signals that may suffer from noise and channel distortions.

The Shalvi-Weinstein algorithm maximizes $|K_y|$ subject to the power constraint $E\{|y[n]|^2\}=E\{|s[n]|^2\}$.

SIMO Equalization Symbol Estimation

Blind SIMO Linear Equalization

Any adaptive blind equalization algorithm can be easily adopted for linear SIMO equalizers [50]. SIMO blind equalization may offer a convergence advantage given the subchannel diversity [51]. While algorithms such as CMA in SISO equalization may suffer from local convergence [52], CMA and the super-exponential method [54] are shown to converge to complete ISI removal under noiseless channels [50], [55]. Furthermore, there is a close relationship between CMA and the nonblind minimum MSE equalizer [71], [72].

Blind Closed-Form Symbol Estimation

Joint blind channel estimation and symbol estimation methods based on a deterministic framework were presented earlier. It should be noted, however, that local convergence and high complexity are their notable disadvantages. A subspace method was presented in [56] that leads to closed-form solutions without high complexity.

Following (8) and (15), multiple snapshots of $\vec{r}$ can be collected as

$$\mathcal{H}(\vec{r})_{Np\times(L+N)}\triangleq\left[\vec{r}_n\vec{r}_{n-1}\cdots\vec{r}_{n-m}\right]=\mathcal{T}\left(\vec{h}\right)\underbrace{\left[\vec{s}_n\vec{s}_{n-1}\cdots\vec{s}_{n-m}\right]}_{\mathcal{H}(\vec{s}):(L+N)\times(m+1)}. \tag{37}$$

Under the condition that $\mathcal{T}(\vec{h})$ has full column rank, the span of $\mathcal{H}(\vec{r})$ is identical to the span of $\mathcal{H}(\vec{s})$. It was shown [56] that so long as $\mathcal{H}(\vec{s})$ has more modes than the rank of $\mathcal{T}(\vec{h})$, then $\mathcal{H}(\vec{s})$ can be uniquely determined from the nullspace U_n of $\mathcal{H}(\vec{r})$. The input symbols can be found by solving a set of linear equations

$$U_n^H\mathcal{H}(\vec{s})=0. \tag{38}$$

Some approaches do not require that the spectrum of $s[n]$ be known. A computationally more efficient row span intersection implementation of the above linear algorithm was given by [57].

Viterbi Algorithm for Blind Sequence Estimation

The Viterbi algorithm can also be applied for blind sequence estimation based on a statistical preprocessing step [58]. Assuming that $\mathcal{T}(\vec{h})$ is full rank and $s[n]$ is white

$$\mathbf{R}_r \triangleq E\{\vec{r}_n \vec{r}_n^H\} = \mathcal{T}(\vec{h})\mathcal{T}(\vec{h})^H + \sigma_w^2 \mathbf{I}. \tag{39}$$

When the channel is noiseless, $\sigma_w^2 = 0$ and singular value decomposition yields

$$\mathbf{R}_x = \mathbf{U}_s \underbrace{\mathbf{diag}(\lambda_1^2, \ldots, \lambda_d^2)}_{\Lambda_s^2} \mathbf{U}_s^H \tag{40}$$

where $d = L + N$ is rank of $\mathcal{T}(\vec{h})$. The Mahalanobis orthogonalization transform can be used to preprocess the received data vector by

$$\mathbf{y}_n \triangleq \Lambda_s^{-1} \mathbf{U}_s^H \vec{r}_n. \tag{41}$$

Thus, when there is no noise

$$q_n^{(k)} \triangleq \mathbf{y}_n^H \mathbf{y}_{n-k} = \vec{s}_n^H \vec{s}_{n-k} = \sum_{i=0}^{d-1} s[n-i]^* s[n-k-i]. \tag{42}$$

A direct application of the Viterbi algorithm to $q_n^{(k)}$ can be used to estimate the unknown data sequence $\{s[n], s[n-1], \ldots, s[n-k-d+1]\}$. To reduce the number of states, select $k = 1$ and estimate the unknown input sequence via Viterbi algorithm

$$\min_{s[m], s[m-1], \cdots, s[1]} \sum_{n=d+1}^{m} |q_n^{(1)} - \sum_{i=0}^{d-1} s[n-i]^* s[n-k-i]|^2. \tag{43}$$

Refer to [58] for a more generalized Mahalanobis transform when noise is present.

Iterative Blind Symbol Estimation

The iterative channel and symbol estimation method, as summarized earlier, also allows direct channel input estimation. Both the iterative least squares with enumeration (ILSE) and iterative least squares with projection (ILSP) exploit the finite alphabet nature of the channel input signals. Given that elements in $\vec{s}$ come from $\mathcal{S}$, the task of implementing

$$\min_{\vec{h}, \vec{s} \in \mathcal{S}} \left\| \mathcal{H}(\vec{r}) - \mathcal{T}(\vec{h}) \mathcal{H}(\vec{s}) \right\|^2 \tag{44}$$

can be iteratively implemented to improve the estimate in each step, as in (20) and (21). ILSP simply replaces the complex symbol estimation step of (20) by a simpler projection [59]

$$\vec{s}^{(k)} = \mathrm{proj}_{\mathcal{S}}\left(\mathcal{T}\left(\vec{h}^{(k)}\right)^{\dagger} \vec{r}\right). \tag{45}$$

In [60], a decision feedback method was presented. This method utilizes the intermediate smoothing error of the least squares smoothing (LSS) approach. Assuming past decisions are correct, decision of the latest symbol is based on the closeness (in terms of angle) between the linear prediction (least square smoothing) error and the projection of the input signal vector onto a pierced observation subspace $\mathcal{Z}_x(n)$.

Clapp and Godsill also successfully exploited the sequential importance sampling idea for blind sequence estimation [61].

Applications of Blind Equalization

Commercially, blind equalization has found new applications in the digital HDTV system and the digital cable modem [70]. More recently, promising results have also been reported on the application of blind equalization in the popular wireless GSM cellular system [62] using higher-order statistical deconvolution method [62] as well as CMA and second-order statistical channel identification [63].

Jitendra K. Tugnait (Fellow) is a Professor in the Department of Electrical and Computer Engineering, Auburn University, Auburn, AL. He received his Ph.D. degree from the University of Illinois, Urbana-Champaign, in 1978 in electrical engineering. From 1978 to 1982 he was an Assistant Professor of Electrical and Computer Engineering at the University of Iowa, Iowa City. He was with the Long Range Research Division of the Exxon Production Research Company, Houston, TX, from June 1982 to September 1989. He joined Auburn University in September 1989. His research interests are in statistical signal processing and stochastic systems analysis with applications to communications. Dr. Tugnait served as an Associate Editor of the *IEEE Transactions on Signal Processing* from December 1994 to December 1997. He is currently a member of the IEEE Signal Processing for Communications Technical Committee.

Lang Tong (Member) is an Associate Professor in the School of Electrical Engineering at Cornell University, Ithaca, NY. He received his Ph.D. degree in electrical engineering from the University of Notre Dame, Notre Dame, IN, in 1990. After being a Postdoctoral Research Affiliate at the Information Systems Laboratory, Stanford University, Stanford, CA, he joined the Department of Electrical and Computer Engineering, West Virginia University, Morgantown, and was with the University of Connecticut, Storrs. Since the fall of 1998, he has been with Cornell University. His research interests include statistical signal processing, wireless communication, and system theory. Dr. Tong received the Young Investigator Award from the Office of Naval Research in 1996 and the Outstanding Young Author Award from the IEEE Circuits and Systems Society. He is currently a member of the IEEE Signal Processing for Communications Technical Committee.

Zhi Ding (Senior Member) is an Associate Professor in the Department of Electrical and Computer Engineering at the University of Iowa. He received his Ph.D. degree from the School of Electrical Engineering, Cornell University, Ithaca, NY in August 1990. From 1990 to 1998, he was a faculty member in the Department of Electrical Engineering, Auburn University, Auburn, AL, first as an Assistant Professor and later as an Associate Professor. Dr. Ding has held visiting positions in the Australian National University, the Hong Kong University of Science and Technology, the NASA Lewis Research Center, and the USAF Wright Laboratory. He was an Associate Editor of the *IEEE Transactions on Signal Processing*. He currently serves as a member of the IEEE Signal Processing for Communications Technical Committee. His research covers a number of research issues involving statistical signal processing that include communications system design, signal detection and classification, as well as blind signal separation.

References

[1] G.B. Giannakis and E. Serpedin, "Linear multichannel blind equalizers of nonlinear FIR Volterra channels," *IEEE Trans. Signal Processing*, vol. 45, pp. 67-81, Jan. 1997.

[2] G.B. Giannakis and C. Tepedelenlioglu, "Basis expansion models and diversity techniques for blind identification and equalization of time-varying channels," *Proc. IEEE*, pp. 1969-1986, Nov. 1998.

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

[4] J.G. Proakis, *Digital Communications*, 3rd ed. New York: McGraw-Hill, 1995.

[5] T.S. Rappaport, *Wireless Communications: Principles and Practice*. Upper Saddle River, NJ: Prentice-Hall, 1996.

[6] L.E. Baum, T. Petrie, G. Soules, and N. Weiss. "A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains," *Ann. Math. Stat.*, vol. 41, pp. 164-171, 1970.

[7] P.J. Bickel and K.A. Doksum, *Mathematical Statistics: Basic Ideas and Selected Topics*. Englewood Cliffs, NJ: Prentice Hall, 1977.

[8] H.A. Cirpan and M.K. Tsatsanis, "Stochastic maximum likelihood methods for semi-blind channel estimation," *IEEE Signal Processing Lett.*, vol. 5, pp. 21-24, Jan. 1998.

[9] E. de Carvalho and D.T.M. Slock, "Semi-blind methods for FIR multichannel estimation," *Signal Processing Advances in Communications*, vol. I, G. Giannakis, Y. Hua, P. Stoica, and L. Tong, Eds. Englewood Cliffs, NJ: Prentice Hall, 2000.

[10] E. de Carvalho and D.T.M. Slock, "Cramér-Rao bounds for semi blind, blind and training sequence based channel estimation," in *Proc. 1st IEEE Signal Processing Workshop Signal Processing Advances in Wireless Communication*, vol 1, Paris, France, Apr. 1997, pp. 129-132.

[11] E. de Carvalho and D.T.M. Slock, "Maximum-likelihood blind FIR multi-channel estimation with Gaussian prior for the symbols," in *Proc. IEEE Int. Conf. Acoust. Speech, Signal Processing*, vol. 5, Munich, Germany, Apr. 1997, pp. 3593-3596.

[12] A.P. Dempster, N.M. Laird, and D.B. Rubin, "Maximum likelihood from incomplete data via EM algorithm," *J. Royal Statist. Soc.*, vol. 39 (Ser. B), 1977.

[13] G.D. Forney, "The Viterbi algorithm," *Proc. IEEE* , vol. 61, pp. 268-278, Mar. 1972.

[14] D. Gesbert and P. Duhamel, "Robust blind channel identification and equalization based on multi-step predictors," in *Proc. ICASSP-97*, Apr. 1997, pp. 3621-3624.

[15] M. Ghosh and C.L. Weber, "Maximum-likelihood blind equalization," *Optical Eng.*, vol. 31, no. 6, pp. 1224-1228, June 1992.

[16] A. Gorokhov and P. Loubaton, "Semi-blind second order identification of convolutive channels," in *Proc. IEEE Int. Conf. Acoustics Speech, and Signal Processing*, Munich, Germany, 1997, pp. 3905-3908.

[17] S. Haykin, *Array Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1984.

[18] G.K. Kaleh and R. Vallet, "Joint parameter estimation and symbol detection for linear or non linear unknown dispersive channels," *IEEE Trans. Comm.*, vol. 42, pp. 2406-2413, July 1994.

[19] V. Krishnamurphy and J. B. Moore, "On-line estimation of hidden Markov Model parameters based on Kullback-Leibler Information measure," *IEEE Trans. Signal Processing*, vol. 41, pp. 2557-2573, Aug. 1993.

[20] J. Laurila, K. Kopsa, and E. Bonek, "Semi-blind signal estimation for smart antennas using subspace tracking," in *Proc. 1999 SPAWC*, Annapolis, MD, May 1999.

[21] L. Rabiner, "A tutorial on hidden Markov Models and selected applications in speech recognition," *Proc. IEEE*, vol. 77, pp. 257-285, Feb. 1989.

[22] N. Seshadri, "Joint data and channel estimation using fast blind trellis search techniques," in *Proc. GLOBECOM'90*, 1991, pp. 1659-1663.

[23] D. Slock, "Blind fractionally-spaced equalization, perfect reconstruction filterbanks, and multilinear prediction," in *Proc. ICASSP'94 Conf.*, Adelaide, Australia, Apr. 1994.

[24] E. Moulines, et al., "Subspace-methods for the blind identification of multichannel FIR filters," *IEEE Trans. Signal Processing*, vol. 43, pp. 516-525, Feb. 1995.

[25] Y. Li and Z. Ding, "ARMA system identification based on second order cyclostationarity," *IEEE Trans. Signal Processing*, vol. 42, pp. 3483-3493, Dec. 1994.

[26] K.A. Meraim, et al. "Prediction error methods for time-domain blind identification of multichannel FIR filters," in *Proc. ICASSP*, May 1995, pp. 1968-1971.

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

[28] Z. Ding, "Matrix outer-product decomposition method for blind multiple channel identification," *IEEE Trans. Signal Processing*, vol. 45, pp. 3054-3061, 1997.

[29] D.M. Titterington, A.F.M. Smith, and U.E. Makov, *Statistical Analysis of Finite Mixture Distributions*. New York: Wiley, 1985.

[30] D.M. Titterington. "Recursive parameter estimation using incomplete data," *J. R. Statist. Soc. B*, vol. 46, no. 2, pp. 257-267, 1984.

[31] L. Tong and S. Perreau, "Multichannel blind channel estimation: From subspace to maximum likelihood methods," *Proc. IEEE*, vol. 86, pp. 1951-1968, Oct. 1998.

[32] 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, Nov. 1991.

[33] L. Tong and Q. Zhao, "Joint order detection and blind channel estimation by least squares smoothing," *IEEE Trans. Signal Processing*, vol. 47, pp. 2345-2355, Sept. 1999.

[34] E. Weinstein, M. Feder, and A. Oppenheim. "Sequential algorithms for parameter estimation based on the Kullback-Leibler information measure," *IEEE Trans. Signal Processing*, vol. 38, pp. 1652-1654, Sept. 1990.

[35] 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.

[36] Q. Zhao and L. Tong, "Adaptive blind channel estimation by least squares smoothing," *IEEE Trans. Signal Processing*, vol. 47, pp. 3000-3012, Nov. 1999.

[37] R.W. Lucky, "Techniques for adaptive equalization of digital communication systems,"*Bell Syst. Tech. J.*, vol. 45, pp. 255-286, Feb. 1966.

[38] A. Duel-Hallen and C.D. Heegard, "Delayed decision feedback sequence estimation," *IEEE Trans. Commun.*, vol. 37, pp. 428-436, May 1989.

[39] J.E. Mazo, "Analysis of decision-directed equalizer convergence," *Bell Syst. Tech. J.*, vol. 59, pp. 1857-1876, Dec. 1980.

[40] O. Macchi and E. Eweda, "Convergence analysis of self-adaptive equalizers,"*IEEE Trans. Inform. Theory*, vol. IT-30, pp. 162-176, Mar. 1983.

[41] Y. Sato, "A method of self-recovering equalization for multi-level amplitude modulation," *IEEE Trans. Commun.*, vol. COM-23, pp. 679-682, June 1975.

[42] 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. Contr.*, vol. AC-25, pp. 385-399, June 1980.

[43] G. Picchi and G. Prati, "Blind equalization and carrier recovery using a 'stop-and-go' decision-directed algorithm," *IEEE Trans. Commun.*, vol. COM-35, pp. 877-887, Sept. 1987.

[44] R. Godfrey and F. Rocca, "Zero memory non-linear deconvolution," *Geophys. Prospecting*, vol. 29, pp. 189-228, 1981.

[45] S. Bellini and F. Rocca, "Blind deconvolution: polyspectra or Bussgang techniques," in *Digital Communications*, E. Biglieri and G. Prati, Eds. North-Holland: Elsevier Science, 1986, pp. 251-263.

[46] D.N. Godard, "Self-recovering equalization and carrier tracking in two-dimensional data communication systems," *IEEE Trans. Commun.*, vol. COM-28, pp. 1867-1875, Nov. 1980.

[47] D. Donoho, "On minimum entropy deconvolution," in *Applied Time Series Analysis II*, New York: Academic, 1981, pp. 565-608.

[48] J.R. Treichler and M.G. Agee, "A new approach to multipath correction of constant modulus signals," *IEEE Trans. Acoustics, Speech, Signal Processing*, vol. ASSP-31, pp. 349-472, Apr. 1983.

[49] O. Shalvi and E. Weinstein, "New criteria for blind deconvolution of non-minimum phase systems (channels)," *IEEE Trans. Inform. Theory*, vol. 36, pp. 312-321, Mar. 1990.

[50] Y. Li and Z. Ding, "Global convergence of fractionally spaced Godard adaptive equalizers," *IEEE Trans. Signal Processing*, vol. 44, pp. 818-826, Apr. 1996.

[51] I. Fijalkow, A. Touzni, and J.R. Treichler, "Fractionally spaced equalization using CMA: Robustness to channel noise and lack of disparity," *IEEE Trans. Signal Processing*, vol. 44, pp. 56-66, Jan. 1997.

[52] Z. Ding, R.A. Kennedy, B.D.O. Anderson, and C.R. Johnson, Jr, "Ill-convergence of Godard blind equalizers in data communications," *IEEE Trans. Commun.*, vol. 39, pp. 1313-1328, Sept. 1991.

[53] D. Hatzinakos, "Blind equalization using stop-and-go criterion adaptation rules," *Opt. Eng.*, vol. 31, pp. 1181-1198, June 1992.

[54] O. Shalvi and E. Weinstein, "Super-exponential methods for blind deconvolution," *IEEE Trans. Inform. Theory*, vol. 39, pp. 504-519, Mar. 1993.

[55] Z. Ding, "On convergence analysis of fractionally spaced adaptive blind equalizers," *IEEE Trans. Signal Processing*, pp. 650-657, Mar. 1997.

[56] H. Liu and G. Xu, "Closed-form blind symbol estimation in digital communications," *IEEE Trans. Signal Processing*, vol. 43, pp. 2714-2723, Nov. 1995.

[57] 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. 45, pp. 173-190, Jan. 1997.

[58] L. Tong, "Blind sequence estimation," *IEEE Trans. Commun.*, vol. 43, pp. 2986-2994, Dec. 1995.

[59] S. Talwar, M. Viberg, and A. Paulraj, "Blind separation of synchronous co-channel digital signals using an antenna array - Part I: Algorithms," *IEEE Trans. Signal Processing*, vol. 44, pp. 1184-1197, May 1996.

[60] Q. Zhao and L. Tong, "Decision feedback blind symbol estimation by adaptive least square smoothing," in *Proc. ICASSP 1999*, Phoenix, AZ, 1999.

[61] T. C. Clapp and S. J. Godsill, "Fixed-lag blind equalization and sequence estimation in digital communications systems using sequential importance sampling," in *Proc. ICASSP 1999*, Phoenix, AZ, 1999.

[62] D. Boss, K.-D. Kammeyer, and T. Petermann, "Is blind channel estimation feasible in mobile communication systems? A study based on GSM," *IEEE J. Select. Areas Commun.*, vol. 16, pp.1480-1492, Oct. 1998.

[63] Z. Ding and G. Li, "Single channel blind equalization for GSM cellular systems," *IEEE J. Select. Areas Commun.*, vol. 16, pp.1493-1505, Oct. 1998.

[64] J.K. Tugnait, "Blind estimation and equalization of digital communication FIR channels using cumulant matching," *IEEE Trans. Commun.*, vol. 43, pp. 1240-1245, Feb./Mar./Apr. 1995.

[65] J.K. Tugnait, "Blind equalization and estimation of FIR communications channels using fractional sampling," *IEEE Trans. Commun.*, vol. 44, pp. 324-336, Mar. 1996.

[66] J.K. Tugnait, "Multistep linear predictors-based blind equalization of FIR/IIR single-input multiple-output channels with common zeros," *IEEE Trans. Signal Processing*, vol. 47, pp. 1689-1700, June 1999.

[67] J.K. Tugnait, "Channel estimation and equalization using higher-order statistics," *Signal Processing Advances in Communications*, vol. I, G. Giannakis, Y. Hua, P. Stoica, and L. Tong, Eds. Englewood Cliffs, NJ: Prentice Hall, 2000.

[68] D. Hatzinakos and C.L. Nikias, "Blind equalization using a tricepstrum based algorithm," *IEEE Trans. Commun.*, vol. 39, pp. 669-681, May 1991.

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

[70] J. Treichler and C.R. Johnson, Jr., "Blind fractionally spaced equalization of digital cable TV," in *Proc. 8th IEEE Workshop Statistical Signal Proc.*, Corfu, Greece, 1996, pp. 122-130.

[71] H. Zeng, L. Tong, and C.R. Johnson, Jr., "Relationships between CMA and Wiener receivers," *IEEE Trans. Inform. Theory*, vol. 44, pp. 1523-1538, July 1998.

[72] M. Gu and L. Tong, "Geometrical characterization of constant modulus receivers" *IEEE Trans. Signal Processing*, vol. 47, no. 10, pp. 2745-2756, Oct. 1999.