2000 Conference on Information Sciences and Systems, Princeton University, March 15-17, 2000

Sequence Estimation over Linearly-Constrained Random Channels

Richard Perry and Kevin Buckley
ECE Department, Villanova University,
Villanova, PA 19085
perry@ece.villanova.edu,
(610)-519-4969, FAX: (610)-519-4436

Abstract — **This paper presents a new approach using EM (Expectation-Maximization) algorithms for ML (maximum likelihood) sequence estimation over unknown ISI (inter-symbol interference) channels with linearly–constrained random channel coefficients which may be fast time-varying. By using the EM formulation to marginalize over the underlying channel coefficient distribution, maximum-likelihood estimates of the transmitted sequence are obtained. The EM algorithms are shown to perform better, in terms of BER, than existing algorithms which perform jointly-optimal sequence and channel estimation, or which do not take into account fast time-varying channel effects.**

I. Introduction

Maximum Likelihood Sequence Estimation (MLSE) over an FIR channel with unknown coefficients can be formulated as either a channel estimation problem (followed by MLSE), a joint sequence/channel estimation problem, or a direct sequence estimation problem. Since MLSE is our primary concern, we consider only the latter two formulations here. For joint estimation of both the transmitted sequence and channel coefficients, optimum methods have been developed both for non-time-varying (e.g. [1, 2, 3]) and slowly time-varying (e.g. [4]) channels. This approach can be effective, and the use of non-time-varying block processing algorithms has been proposed for a fast-changing channel (under the assumption that the channel is constant over the processing block). Direct estimation of the sequence (only) can result in improved estimator performance. However, the ML formulation for optimal sequence estimation over random channels, which requires marginalization over the unknown channel parameters, is generally intractable to solve directly.

The Expectation Maximization (EM) method is an approach to development of iterative ML and Maximum a Posterior (MAP) estimation algorithms, which was introduced into the signal processing community by Dempster, et al. [5]. For digital communications, the EM approach has been applied to the channel estimation problem [6, 7, 8]. It has also been employed for MLSE. For example, Georghiades, et al. [9] use EM to account for unknown phase given unsynchronized reception. Nelson and Poor [10] employ EM in multiuser applications where the other users' symbols are unknown. Zeger and Kobayashi [11] estimate GMSK modulated symbols in the presence of unknown multipath parameters. In these examples, either simple prior distributions on the unknown secondary parameters are assumed in order to develop algorithms, or a high SNR approximation is derived.

In this paper we develop a general EM algorithmic approach to MLSE over unknown random ISI channels. We use EM to marginalize over unknown channel parameters, incorporating knowledge of the FIR channel coefficients in the form of linear constraints and a prior distribution on the channel parameters. However, it should be noted that improved bit-error-rate performance, relative to joint sequence/channel estimators, can be realized even with only a noninformative prior (and no constraints). Algorithms are developed for a variety of channel models, including several that account for channel variation over short symbol blocks.

II. Communication System Model

In the discrete-time FIR model of a time-varying noisy communications channel with inter-symbol interference, for a block of N received data values, the complex received data r_k at time k is given by:

$$r_k = \mathbf{a}_k^T \mathbf{h}_k + n_k, \;\; k = 1, \ldots, N, \tag{1}$$

where $\mathbf{a}_k^T$ is a complex row vector containing transmitted data $\{a_{k-i+1}, i = 1, \ldots, L\}$, L is the FIR channel length, $\mathbf{h}_k$ is a complex column vector containing the channel impulse response coefficients (which are unknown) at time k, and n_k is the white Gaussian complex noise at time k with variance σ^2. For $j \leq 0$, the transmitted data a_j may be either known (e.g. all 0's), unknown, or estimated with some associated probabilities from the end of a previous data block.

If the channel coefficients are known to be constrained linearly, such as having a zero at DC, the constraints can be expressed in general by an underdetermined matrix equation:

$$\mathbf{F}\mathbf{h}_k = \mathbf{f}. \tag{2}$$

The constraint parameters $\mathbf{F}$ and $\mathbf{f}$ may be time-varying; we are omitting the time subscript k on them only to simplify the notation. (2) implies the following orthogonal decomposition of $\mathbf{h}_k$:

$$\mathbf{h}_k = \mathbf{M}\mathbf{p}_k + \bar{\mathbf{M}}\bar{\mathbf{p}} \; . \tag{3}$$

Using the singular-value decomposition of $\mathbf{F}$:

$$\mathbf{F} = \mathbf{U}\mathbf{S}\mathbf{V}^{*T} = \mathbf{U}\,[\mathbf{S}_1 \mathbf{0}]\,[\mathbf{V}_1 \mathbf{V}_0]^{*T} \;\; , \tag{4}$$

with the matrices partitioned according to the rank of $\mathbf{F}$, we can then express M, $\bar{\mathrm{M}}$, and $\bar{\mathrm{p}}$ as:

$$\mathbf{M} = \mathbf{V}_0, \;\; \bar{\mathbf{M}} = \mathbf{V}_1, \;\; \bar{\mathbf{p}} = \mathbf{S}_1^{-1}\mathbf{U}^{*T}\mathbf{f} \tag{5}$$

where the columns of $\mathbf{M}$ represent an orthonormal basis for the null space of $\mathbf{F}$.

Note that the constraint parameters and M, $\bar{\mathrm{M}}$, and $\bar{\mathrm{p}}$ are assumed to be known, but $\mathbf{p}_k$ is unknown. Given values or a statistical description of $\mathbf{p}_k$ we can produce values or a statistical description of the channel coefficients $\mathbf{h}_k$ using (3). We will refer to $\mathbf{p}_k$ as the *underlying* channel coefficients.

Let $\mathbf{P} = [\mathbf{p}_1, \ldots, \mathbf{p}_N]$ represent the matrix of underlying channel coefficient vectors over time arranged by columns, and let $\mathbf{A} = [\mathbf{a}_1, \ldots, \mathbf{a}_N]^T$ represent the matrix of transmitted data arranged by rows. Also let $\mathbf{r} = [r_1, \ldots, r_N]^T$ represent the column vector of received data, and $\mathbf{n} = [n_1, \ldots, n_N]^T$ represent the column vector of noise over time. $\mathbf{A}$, $\mathbf{r}$, and $\mathbf{n}$ will be used in some matrix-vector formulas, but the notation for $\mathbf{P}$ as a matrix is simply a notational convenience to refer to the entire collection of underlying channel coefficients over the block of time.

With this notation, the probability density function of the received data, given $\mathbf{P}$ and $\mathbf{A}$, is:

$$f(\mathbf{r}|\mathbf{P},\mathbf{A}) = \frac{1}{(\pi\sigma^2)^N} \prod_{k=1}^{N} e^{-\frac{|r_k - \mathbf{a}_k^T \mathbf{h}_k|^2}{\sigma^2}} \quad . \tag{6}$$

For a non-time-varying channel, let $\mathbf{h} = \mathbf{h}_1 = \cdots = \mathbf{h}_N$ represent the single constant (and unknown) column vector of channel coefficients, and let $\mathbf{p} = \mathbf{p}_1 = \cdots = \mathbf{p}_N$ represent the underlying channel coefficient vector. In this case, using matrix-vector notation, (1) and (6) can be expressed more simply as:

$$\mathbf{r} = \mathbf{A}\mathbf{h} + \mathbf{n} \quad , \tag{7}$$

and

$$f(\mathbf{r}|\mathbf{P},\mathbf{A}) = \frac{1}{(\pi\sigma^2)^N} e^{-\frac{|\mathbf{r}-\mathbf{A}\mathbf{h}|^2}{\sigma^2}} \quad . \tag{8}$$

The basic equations above describe the communication system under consideration, based only on the facts that the additive noise is white and Gaussian, and the channel coefficients are linearly–constrained, but without making any additional assumptions about the properties of the channel itself. In the succeeding sections, additional known properties of the random channel will be considered, starting from the most general joint pdf case and proceeding to specific examples of underlying channel coefficient probability distribution functions.

III. EM and ML Sequence Estimation

To minimize BER, we must find $\mathbf{A}$ to maximize $f(\mathbf{A}|\mathbf{r})$. Using Bayes rule:

$$f(\mathbf{A}|\mathbf{r}) = \frac{f(\mathbf{r}|\mathbf{A})f(\mathbf{A})}{f(\mathbf{r})} = K_9 f(\mathbf{r}|\mathbf{A})f(\mathbf{A}) \doteq f(\mathbf{r}|\mathbf{A})f(\mathbf{A}) \ , \tag{9}$$

where K_9 is a constant which takes into account the effect of $f(\mathbf{r})$ for normalizing the distribution.[1]

MAP estimators minimize BER by maximizing $f(\mathbf{r}|\mathbf{A})f(\mathbf{A})$. If $f(\mathbf{A})$ is unknown or assumed to be uniform, then the ML estimator which maximizes the likelihood function $f(\mathbf{r}|\mathbf{A})$ can be used instead. All of the EM algorithms derived in the subsequent sections produce iterative solutions to the ML problem of maximizing $f(\mathbf{r}|\mathbf{A})$. But if $f(\mathbf{A})$ is known and not uniform, then this could be incorporated into the following EM algorithms as an additive term in the cost function, thus producing a MAP estimator.

The dependence of $f(\mathbf{r}|\mathbf{A})$ on the underlying random channel coefficients is:

$$f(\mathbf{r}|\mathbf{A}) = E[f(\mathbf{r}|\mathbf{P},\mathbf{A})] = \int_{\mathbf{P}} f(\mathbf{r}|\mathbf{P},\mathbf{A})f(\mathbf{P})d\mathbf{P}, \tag{10}$$

where $f(\mathbf{r}|\mathbf{P},\mathbf{A})$ is given by (6) for time-varying channels. Evaluating this multidimensional integral and then maximizing the result with respect to $\mathbf{A}$ seems to be an intractable problem in general. Yet this is exactly the problem which EM algorithms can solve iteratively.

The only apparent case where direct evaluation and maximization of $f(\mathbf{r}|\mathbf{A})$ from (10) seems tractable is if $\mathbf{P}$ is non-time-varying and deterministic. Then $f(\mathbf{P}) = \delta(\mathbf{p} - \mathbf{d})$, where $\mathbf{d}$ is an unknown constant, $\mathbf{h}_k = \mathbf{M}\mathbf{d} + \bar{\mathbf{M}}\bar{\mathbf{p}} = \mathbf{g}$ from (3), and (10) reduces to (8) with $\mathbf{h} = \mathbf{g}$, where $\mathbf{g}$ is the deterministic but unknown value of $\mathbf{h}$, as discussed further in Section IV. This provides a theoretical basis for the joint $(\mathbf{h},\mathbf{A})$ estimation algorithms such as [1, 12, 2, 4]. However for arbitrary $f(\mathbf{P})$ we must work with (10) directly, or use an EM algorithm, in order to produce optimal estimates of $\mathbf{A}$.

To maximize (10) using EM, define the auxiliary function [5]:

$$\begin{aligned} Q(\mathbf{A}|\mathbf{B}) &= E[\log(f(\mathbf{r},\mathbf{P}|\mathbf{A}))\,|\mathbf{r},\mathbf{B}] \\ &= \int_{\mathbf{P}} \log(f(\mathbf{r},\mathbf{P}|\mathbf{A}))f(\mathbf{P}|\mathbf{r},\mathbf{B})d\mathbf{P}, \end{aligned} \tag{11}$$

where we desire the ML estimate of $\mathbf{A}$, and $\mathbf{B}$ represents the current estimate of $\mathbf{A}$. In the terminology of EM, $\mathbf{r}$ is the observed data, $\mathbf{P}$ is the hidden data, and $(\mathbf{r},\mathbf{P})$ is the complete data.

The general EM algorithm for this problem is:

1. Initialize $\mathbf{B}$ to an estimate of $\mathbf{A}$.
2. E-step (Expectation): construct $Q(\mathbf{A}|\mathbf{B})$
3. M-step (Maximization): find $\mathbf{A}$ to maximize $Q(\mathbf{A}|\mathbf{B})$
4. Set $\mathbf{B} = \mathbf{A}$ and repeat steps 2 and 3 until converged.

In general, to construct $Q(\mathbf{A}|\mathbf{B})$, start with $f(\mathbf{r},\mathbf{P}|\mathbf{A})$:

$$f(\mathbf{r},\mathbf{P}|\mathbf{A}) = f(\mathbf{r}|\mathbf{P},\mathbf{A})f(\mathbf{P}|\mathbf{A}). \tag{12}$$

We can drop the $f(\mathbf{P}|\mathbf{A})$ term from this since $\mathbf{P}$ and $\mathbf{A}$ are independent so $f(\mathbf{P}|\mathbf{A})$ is not a function of $\mathbf{A}$ and will not affect the maximization in the M-step. $Q(\mathbf{A}|\mathbf{B})$ from (11) may then be written as:

$$\begin{aligned} Q(\mathbf{A}|\mathbf{B}) &\doteq E[\log(f(\mathbf{r}|\mathbf{P},\mathbf{A}))\,|\mathbf{r},\mathbf{B}] \\ &= \int_{\mathbf{P}} \log(f(\mathbf{r}|\mathbf{P},\mathbf{A}))f(\mathbf{P}|\mathbf{r},\mathbf{B})d\mathbf{P}. \end{aligned} \tag{13}$$

To evaluate $Q(\mathbf{A}|\mathbf{B})$, we need $f(\mathbf{P}|\mathbf{r},\mathbf{B})$, which can be written in terms of the apriori underlying channel coefficient density function $f(\mathbf{P})$ as:

$$f(\mathbf{P}|\mathbf{r},\mathbf{B}) = \frac{f(\mathbf{r}|\mathbf{P},\mathbf{B})f(\mathbf{P}|\mathbf{B})}{f(\mathbf{r}|\mathbf{B})} \quad , \tag{14}$$

where the denominator term $f(\mathbf{r}|\mathbf{B})$ can be treated as a constant since it simply serves to normalize the distribution, and is not a function of $\mathbf{P}$ or $\mathbf{A}$, so will not affect the maximization step. Also, $f(\mathbf{P}|\mathbf{B}) = f(\mathbf{P})$ since the underlying channel coefficients and transmitted data are independent. But note that given $\mathbf{r}$ and $\mathbf{B}$, $f(\mathbf{P}|\mathbf{r},\mathbf{B}) \neq f(\mathbf{P})$.

Therefore, for $f(\mathbf{P}|\mathbf{r},\mathbf{B})$ in (13) we may use:

$$f(\mathbf{P}|\mathbf{r},\mathbf{B}) \doteq f(\mathbf{r}|\mathbf{P},\mathbf{B})f(\mathbf{P}). \tag{15}$$

In the specific cases considered subsequently, we will show that the E-step reduces to computing sufficient statistics of $Q(\mathbf{A}|\mathbf{B})$, which are the mean and covariance of $\mathbf{P}$ conditioned on $\mathbf{r}$ and $\mathbf{B}$. This enables the Viterbi algorithm [13] to be used for the M-step. The Viterbi algorithm is a well-known efficient method for ML sequence estimation over known channels. It can be used here in the unknown channel case because of the estimates provided by the E-step. The E-step and M-step formulas will be derived in detail in subsequent sections for various specific underlying channel coefficient distributions.

IV. EM for Non-Time-Varying Channels

For a non-time-varying channel, using (8) in (15), (13) becomes:

$$Q(\mathbf{A}|\mathbf{B}) \doteq E[-|\mathbf{r}-\mathbf{A}\mathbf{h}|^2\ |\mathbf{r},\mathbf{B}] = \int_{\mathbf{p}} -|\mathbf{r}-\mathbf{A}\mathbf{h}|^2 f(\mathbf{p}|\mathbf{r},\mathbf{B})d\mathbf{p}, \tag{16}$$

[1] As a matter of notation, we will subscript normalization constants such as K_9 throughout this paper with the equation number in which they first appear. Also, we will use the symbol $\doteq$ to mean "equivalent within an additive and/or multiplicative constant" for equations in which such constants have no effect for purposes here.

with

$$f(\mathbf{p}|\mathbf{r},\mathbf{B}) = K_{17}\, e^{-\frac{|\mathbf{r}-\mathbf{B}\mathbf{h}|^2}{\sigma^2}} f(\mathbf{p}) \ , \tag{17}$$

where K_{17} can be determined if needed by setting the integral of $f(\mathbf{p}|\mathbf{r},\mathbf{B})$ to 1.

To evaluate (16) we only need the mean and covariance of the channel coefficients $\mathbf{h}$ given $\mathbf{r}$ and $\mathbf{B}$ using the distribution of (17). We will derive these for specific cases of $f(\mathbf{p})$ subsequently; for now, we define them symbolically as $\mathbf{g}$ and $\mathbf{G}$:

$$\mathbf{g} \ \triangleq \ E[\mathbf{h}|\mathbf{r},\mathbf{B}] \ = \ \mathbf{M}\mathbf{d} + \bar{\mathbf{M}}\bar{\mathbf{p}} \tag{18}$$

$$\mathbf{G} \ \triangleq \ E[(\mathbf{h}-\mathbf{g})(\mathbf{h}-\mathbf{g})^{*T}|\mathbf{r},\mathbf{B}] \ = \ \mathbf{M}\mathbf{D}\mathbf{M}^{*T} \ , \tag{19}$$

where $\mathbf{d}$ and $\mathbf{D}$ are the mean and covariance of the underlying channel coefficients $\mathbf{p}$ given $\mathbf{r}$ and $\mathbf{B}$:

$$\mathbf{d} \ \triangleq \ E[\mathbf{p}|\mathbf{r},\mathbf{B}] \tag{20}$$

$$\mathbf{D} \ \triangleq \ E[(\mathbf{p}-\mathbf{d})(\mathbf{p}-\mathbf{d})^{*T}|\mathbf{r},\mathbf{B}] \ . \tag{21}$$

The noise variance σ^2 is needed in general to compute $\mathbf{g}$ and $\mathbf{G}$. If σ^2 is not known it could be estimated from the variances of the transmitted and received data.

The norm-squared term in (16), for each candidate $\mathbf{A}$, can be represented as a path through a trellis, whose states at a given time are the states of the of the ISI channel. A path represents a summation of incremental state transition costs. The minimum cost path can be determined by the Viterbi algorithm.

Rewriting (16) in terms of the incremental costs:

$$Q(\mathbf{A}|\mathbf{B}) \doteq -\sum_{k=1}^{N} V_k \ , \tag{22}$$

where V_k is:

$$V_k = E[|r_k - \mathbf{a}_k^T\mathbf{h}|^2 \ |\mathbf{r},\mathbf{B}] \ . \tag{23}$$

The expected value is taken over $\mathbf{p}$ using (17) for $f(\mathbf{p}|\mathbf{r},\mathbf{B})$. Expanding the norm squared, and taking the expected value, yields:

$$V_k = |r_k - \mathbf{a}_k^T\mathbf{g}|^2 + \mathbf{a}_k^T\mathbf{G}\mathbf{a}_k^* \ . \tag{24}$$

Equation (24) represents the general form of the Viterbi algorithm incremental cost function for non-time-varying random ISI channels. To use (24) in an EM algorithm, $\mathbf{g}$ and $\mathbf{G}$ from (18) and (19) are computed in the M-step; these computations depend on the specific form of the apriori underlying channel distribution $f(\mathbf{p})$. The following subsections examine the specific cases of deterministic, uniform, and Gaussian $f(\mathbf{p})$.

DETERMINISTIC CHANNEL COEFFICIENTS

The case of deterministic channel coefficients is examined first because it is the simplest case, and because if there are no constraints it reduces to previously published algorithms [1, 12, 2] which do not use any known apriori statistics for the channel. Unlike the other cases considered subsequently, for the deterministic channel we do not know the apriori mean of the coefficients, and as shown below, the E-step of the EM algorithm can not use $E[\mathbf{h}|\mathbf{r},\mathbf{B}]$.

If $\mathbf{h}$ is not random, then let $f(\mathbf{p}) = \delta(\mathbf{p}-\mathbf{d})$, where $\mathbf{d}$ is an unknown constant, and $\mathbf{h} = \mathbf{M}\mathbf{d} + \bar{\mathbf{M}}\bar{\mathbf{p}} = \mathbf{g}$ from (3), where $\mathbf{g}$ is the deterministic but unknown value of $\mathbf{h}$. In this case, $E[\mathbf{h}] = \mathbf{g}$, and using (17) to estimate $\mathbf{d}$ does not provide anything useful since it reduces to $\mathbf{d}$, and $\mathbf{d}$ is unknown.

So to estimate $\mathbf{g}$ in this case, we must use an ad-hoc method. Note that given $\mathbf{r}$ and $\mathbf{B}$, $\mathbf{p}$ satisfies (7):

$$\mathbf{r} = \mathbf{B}\mathbf{M}\mathbf{p} + \mathbf{B}\bar{\mathbf{M}}\bar{\mathbf{p}} + \mathbf{n} \ . \tag{25}$$

Taking the expected value of the right-hand-side of (25) with respect to $\mathbf{p}$ and $\mathbf{n}$ yields:

$$\mathbf{r} = \mathbf{B}\mathbf{M}\mathbf{d} + \mathbf{B}\bar{\mathbf{M}}\bar{\mathbf{p}} \ , \tag{26}$$

which can be solved for a linear least-square-error estimate of $\mathbf{d}$ using the pseudo-inverse of $\mathbf{B}\mathbf{M}$:

$$\mathbf{d} = (\mathbf{B}\mathbf{M})^{+}(\mathbf{r} - \mathbf{B}\bar{\mathbf{M}}\bar{\mathbf{p}}) \ . \tag{27}$$

In this case, $\mathbf{G} = 0$, and the Viterbi algorithm incremental cost function from (24) becomes:

$$V_k = |r_k - \mathbf{a}_k^T\mathbf{g}|^2 \ . \tag{28}$$

For the E-step, (27) and (18) are used to estimate $\mathbf{g}$. For the M-step, the Viterbi algorithm is used with incremental cost function (28) to estimate $\mathbf{A}$. For EM initialization, we could select $\mathbf{B}$ randomly, or use all 1's, etc. But to reduce the number of iterations required, and to help the algorithm converge to a proper solution, it is better to initialize $\mathbf{B}$ with some reasonable estimate of $\mathbf{A}$. This can be done, for example, by initially using a pure delay channel, i.e. using $\mathbf{g} = [0 \ \ldots \ 0 \ 1 \ 0 \ \ldots \ 0]^T$ in (28), then using the Viterbi algorithm to estimate $\mathbf{A}$, then setting $\mathbf{B}$ to this estimate of $\mathbf{A}$.

Note that if there were no constraints, or if the constraints were unknown or simply not taken into account, then $\mathbf{M} = \mathbf{I}$, $\bar{\mathbf{M}} = \mathbf{0}$, and $\mathbf{g} = \mathbf{B}^{+}\mathbf{r}$; this estimate of $\mathbf{g}$ would then be a maximum-likelihood estimate of $\mathbf{h}$ given $\mathbf{r}$ and $\mathbf{B}$ since it maximizes $f(\mathbf{r}|\mathbf{h},\mathbf{B})$ (i.e. minimizes $|\mathbf{r} - \mathbf{B}\mathbf{h}|^2$). The EM algorithm in this case would be equivalent to previously published algorithms for optimal joint sequence and channel estimation [1, 12, 2].

UNIFORM UNDERLYING CHANNEL COEFFICIENTS

Using $f(\mathbf{p}) = 1$ in (17), $\mathbf{d}$ from (27), and assuming that $\mathbf{B}$ has full rank so that $(\mathbf{B}\mathbf{M})^{+} = (\mathbf{M}^{*T}\mathbf{B}^{*T}\mathbf{B}\mathbf{M})^{-1}\mathbf{M}^{*T}\mathbf{B}^{*T}$, then, dropping constants, we obtain:

$$f(\mathbf{p}|\mathbf{r},\mathbf{B})) \doteq e^{-\frac{(\mathbf{p}-\mathbf{d})^{*T}\mathbf{M}^{*T}\mathbf{B}^{*T}\mathbf{B}\mathbf{M}(\mathbf{p}-\mathbf{d})}{\sigma^2}} \ . \tag{29}$$

This shows that, given $\mathbf{r}$ and $\mathbf{B}$, $\mathbf{p}$ has a joint Gaussian distribution with mean $\mathbf{d}$ as in (27), and covariance:

$$\mathbf{D} = \left[\frac{\mathbf{M}^{*T}\mathbf{B}^{*T}\mathbf{B}\mathbf{M}}{\sigma^2}\right]^{-1} \ . \tag{30}$$

Therefore, for the case of uniformly distributed underlying channel coefficients, the EM algorithm is almost the same as for deterministic channel coefficients, except that the Viterbi algorithm incremental cost function from (24), which includes $\mathbf{G}$, is used instead of (28), and $\mathbf{G}$ is given by (30) and (19).

For the E-step, (27) and (30) are used in (18) and (19) to estimate $\mathbf{g}$ and $\mathbf{G}$. For the M-step, the Viterbi algorithm is used with incremental cost function (24) to estimate $\mathbf{A}$.

GAUSSIAN UNDERLYING CHANNEL COEFFICIENTS

Let $f(\mathbf{p})$ be a joint Gaussian distribution with mean $\mathbf{c}$ and covariance $\mathbf{C}$:

$$f(\mathbf{p}) \doteq e^{-(\mathbf{p}-\mathbf{c})^{*T}\mathbf{C}^{-1}(\mathbf{p}-\mathbf{c})} \ , \tag{31}$$

where $\mathbf{c}$ and $\mathbf{C}$ are known. Then (17) reduces to:

$$f(\mathbf{p}|\mathbf{r},\mathbf{B}) \doteq e^{-(\mathbf{p}-\mathbf{d})^{*T}\mathbf{D}^{-1}(\mathbf{p}-\mathbf{d})} \ , \tag{32}$$

with:

$$\mathbf{D} \ = \ \left[\frac{\mathbf{M}^{*T}\mathbf{B}^{*T}\mathbf{B}\mathbf{M}}{\sigma^2} + \mathbf{C}^{-1}\right]^{-1} \tag{33}$$

$$\mathbf{d} \ = \ \mathbf{D}\left(\frac{\mathbf{M}^{*T}\mathbf{B}^{*T}(\mathbf{r} - \mathbf{B}\bar{\mathbf{M}}\bar{\mathbf{p}})}{\sigma^2} + \mathbf{C}^{-1}\mathbf{c}\right) . \tag{34}$$

So in this case, the distribution of $\mathbf{p}$ given $\mathbf{r}$ and $\mathbf{B}$ is also Gaussian. Note that the formulas above for the mean and covariance of $\mathbf{p}$ given $\mathbf{r}$ and $\mathbf{B}$ do not use $(\mathbf{BM})^+$ as in the previous deterministic and uniform cases.

The Gaussian distribution of $\mathbf{p}$ is intermediate between the known deterministic channel case and uniform unknown $\mathbf{p}$. In the limit as $\mathbf{C}$ goes to 0, $\mathbf{D}$ approaches $\mathbf{C}$ and $\mathbf{d}$ approaches known $\mathbf{c}$, which corresponds to the Gaussian distribution approaching $\delta(\mathbf{p}-\mathbf{c})$. This is similar to the deterministic channel $\mathbf{p}$ case except that here the mean is known as opposed to being estimated using (27). In the limit as $\mathbf{C}$ goes to infinity, $\mathbf{D}$ approaches $\left[\frac{\mathbf{M}^{*T}\mathbf{B}^{*T}\mathbf{BM}}{\sigma^2}\right]^{-1}$ and $\mathbf{d}$ approaches $(\mathbf{BM})^+(\mathbf{r}-\mathbf{B}\bar{\mathbf{M}}\bar{\mathbf{p}})$, which corresponds to the Gaussian distribution approaching a uniform distribution.

Also, if $N >> L$ or the noise variance is small, the term involving $\mathbf{B}$ in (33) can dominate the $\mathbf{C}^{-1}$ term, and then $\mathbf{D}$ approaches 0, and $\mathbf{d}$ approaches $(\mathbf{BM})^+(\mathbf{r}-\mathbf{B}\bar{\mathbf{M}}\bar{\mathbf{p}})$, which corresponds to the Gaussian EM algorithm becoming equivalent to the algorithm for deterministic channel coefficients.

For the E-step, (33) and (34) are used in (18) and (19) to estimate $\mathbf{G}$ and $\mathbf{g}$. For the M-step, the Viterbi algorithm is used with incremental cost function (24) to estimate $\mathbf{A}$.

V. EM FOR TIME-VARYING CHANNELS

For a time-varying channel, using (6) in (15) and dropping constants, (13) becomes:

$$Q(\mathbf{A}|\mathbf{B}) \doteq E[-\sum_{k=1}^{N}|r_k-\mathbf{a}_k^T\mathbf{h}_k|^2\,|\mathbf{r},\mathbf{B}] = \int_{\mathbf{P}} -\sum_{k=1}^{N}|r_k-\mathbf{a}_k^T\mathbf{h}_k|^2\, f(\mathbf{P}|\mathbf{r},\mathbf{B})d\mathbf{P}, \quad (35)$$

with:

$$f(\mathbf{P}|\mathbf{r},\mathbf{B}) \doteq f(\mathbf{P})\prod_{k=1}^{N} e^{-\frac{|r_k-\mathbf{b}_k^T\mathbf{h}_k|^2}{\sigma^2}}. \quad (36)$$

In this time-varying case, in the E-step of the EM algorithm, time varying channel impulse response vector conditional means and covariances will be estimated. The details of the EM algorithm in this case depend on the form of $f(\mathbf{P})$. The following subsections examine the specific cases of independent and independent Gaussian underlying channel coefficient distributions.

INDEPENDENT UNDERLYING CHANNEL COEFFICIENTS

If the elements of $\mathbf{P}$ are independent over time, then $f(\mathbf{P})$ can be written as:

$$f(\mathbf{P}) = \prod_{k=1}^{N} f(\mathbf{p}_k), \quad (37)$$

and $f(\mathbf{P}|\mathbf{r},\mathbf{B})$ becomes:

$$f(\mathbf{P}|\mathbf{r},\mathbf{B}) = \prod_{k=1}^{N} f(\mathbf{p}_k|\mathbf{r},\mathbf{B}), \quad (38)$$

with:

$$f(\mathbf{p}_k|\mathbf{r},\mathbf{B}) \doteq f(\mathbf{p}_k)\, e^{-\frac{|r_k-\mathbf{b}_k^T\mathbf{h}_k|^2}{\sigma^2}}. \quad (39)$$

Using (38) and (39) in (35) we obtain:

$$Q(\mathbf{A}|\mathbf{B}) \doteq -\sum_{k=1}^{N}\int_{\mathbf{p}_k}|r_k-\mathbf{a}_k^T\mathbf{h}_k|^2\, f(\mathbf{p}_k)\, e^{-\frac{|r_k-\mathbf{b}_k^T\mathbf{h}_k|^2}{\sigma^2}}d\mathbf{p}_k = -\sum_{k=1}^{N} V_k, \quad (40)$$

with:

$$\mathbf{g}_k \triangleq E[\mathbf{h}_k|\mathbf{r},\mathbf{B}] = \mathbf{M}\mathbf{d}_k + \bar{\mathbf{M}}\bar{\mathbf{p}} \quad (41)$$

$$\mathbf{G}_k \triangleq E[(\mathbf{h}_k-\mathbf{g}_k)(\mathbf{h}_k-\mathbf{g}_k)^{*T}|\mathbf{r},\mathbf{B}] = \mathbf{M}\mathbf{D}_k\mathbf{M}^{*T} \quad (42)$$

$$V_k \triangleq |r_k-\mathbf{a}_k^T\mathbf{g}_k|^2 + \mathbf{a}_k^T\mathbf{G}_k\mathbf{a}_k^*. \quad (43)$$

V_k represents the Viterbi algorithm incremental cost function at time k. $\mathbf{d}_k$ and $\mathbf{D}_k$ are the mean and covariance of the underlying channel coefficients $\mathbf{p}_k$ given $\mathbf{r}$ and $\mathbf{B}$:

$$\mathbf{d}_k \triangleq E[\mathbf{p}_k|\mathbf{r},\mathbf{B}] \quad (44)$$

$$\mathbf{D}_k \triangleq E[(\mathbf{p}_k-\mathbf{d}_k)(\mathbf{p}_k-\mathbf{d}_k)^{*T}|\mathbf{r},\mathbf{B}] .. \quad (45)$$

In the E-step, (41) and (42) will be used to estimate $\mathbf{g}_k$ and $\mathbf{G}_k$, $k=1,\ldots,N$. In the M-step, the Viterbi algorithm with incremental cost function (43) will be used to estimate $\mathbf{A}$. To initialize $\mathbf{B}$ in step 1, we can initialize $\mathbf{g}_k$ and $\mathbf{G}_k$, $k=1,\ldots,N$, then use the Viterbi algorithm to estimate $\mathbf{A}$, then set $\mathbf{B}=\mathbf{A}$. To initialize and estimate $\mathbf{g}_k$ and $\mathbf{G}_k$ in steps 1 and 2 of the algorithm, we need to know the specific form of $f(\mathbf{p}_k)$. The next subsection derives the specific EM algorithm for Gaussian $\mathbf{p}_k$.

GAUSSIAN UNDERLYING CHANNEL COEFFICIENTS

If $f(\mathbf{p}_k)$ is Gaussian with mean $\mathbf{c}_k$ and covariance $\mathbf{C}_k$, then similar to the derivation of (32) for the non-time-varying Gaussian case, we obtain:

$$f(\mathbf{p}_k|\mathbf{r},\mathbf{B}) \doteq e^{-(\mathbf{p}_k-\mathbf{d}_k)^{*T}\mathbf{D}_k^{-1}(\mathbf{p}_k-\mathbf{d}_k)}, \quad (46)$$

with:

$$\mathbf{D}_k = \left[\frac{\mathbf{M}^{*T}\mathbf{b}_k^*\mathbf{b}_k^T\mathbf{M}}{\sigma^2} + \mathbf{C}_k^{-1}\right]^{-1} \quad (47)$$

$$\mathbf{d}_k = \mathbf{D}_k\left(\frac{\mathbf{M}^{*T}\mathbf{b}_k^*(r_k-\mathbf{b}_k^T\bar{\mathbf{M}}\bar{\mathbf{p}})}{\sigma^2} + \mathbf{C}_k^{-1}\mathbf{c}_k\right). \quad (48)$$

For the E-step, (47) and (48) are used in (41) and (42) to estimate $\mathbf{G}_k$ and $\mathbf{g}_k$, $k=1,\ldots,N$. For the M-step, the Viterbi algorithm is used with incremental cost function (43) to estimate $\mathbf{A}$.

GAUSS-MARKOV UNDERLYING CHANNEL COEFFICIENTS

For a fast time-varying channel with Gauss-Markov fading parameters, let α represent a complex first-order Markov factor such that $|\alpha| = e^{-\omega T}$ where T is the sampling period and $\frac{\omega}{\pi}$ is the Doppler spread [14]. Let the time-varying underlying channel coefficients follow a Gauss-Markov distribution such that, at time k:

$$\mathbf{p}_k = \alpha\mathbf{p}_{k-1} + \mathbf{v}_k, \quad (49)$$

where $\mathbf{v}_k$ is complex, white, and Gaussian with mean $\mathbf{c}$ and covariance $\mathbf{C}$. In this case, $f(\mathbf{p}_k|\mathbf{p}_{k-1})$ is Gaussian with mean $\mathbf{c}+\alpha\mathbf{p}_{k-1}$ and covariance $\mathbf{C}$. $Q(\mathbf{A}|\mathbf{B})$ from (35) can be written in terms of $f(\mathbf{p}_k|\mathbf{r},\mathbf{B})$:

$$Q(\mathbf{A}|\mathbf{B}) \doteq -\sum_{k=1}^{N}\int_{\mathbf{p}_k}|r_k-\mathbf{a}_k^T\mathbf{h}_k|^2\, f(\mathbf{p}_k|\mathbf{r},\mathbf{B})d\mathbf{p}_k = -\sum_{k=1}^{N} V_k, \quad (50)$$

and the definitions of $\mathbf{g}_k$, $\mathbf{G}_k$, and V_k from (41-43) still apply.

$f(\mathbf{P})$ can be written as:

$$f(\mathbf{P}) = f(\mathbf{p}_1)\prod_{k=2}^{N} f(\mathbf{p}_k|\mathbf{p}_{k-1}), \quad (51)$$

so $f(\mathbf{P}|\mathbf{r},\mathbf{B})$ from (36) becomes:

$$f(\mathbf{P}|\mathbf{r},\mathbf{B}) \doteq f(\mathbf{p}_1)\prod_{k=2}^{N} f(\mathbf{p}_k|\mathbf{p}_{k-1})\prod_{k=1}^{N} e^{-\frac{|r_k-\mathbf{b}_k^T\mathbf{h}_k|^2}{\sigma^2}}. \quad (52)$$

To determine $f(\mathbf{p}_k|\mathbf{r},\mathbf{B})$, and thereby $\mathbf{g}_k$ and $\mathbf{G}_k$, we must integrate (52) over $\mathbf{p}_i, i=1,\ldots,N, i\neq k$. Due to space limitations, the derivation of this integration result is not shown here. A derivation for the same problem, but without the linear constraints, appears in [15]. The result is that $f(\mathbf{p}_k|\mathbf{r},\mathbf{B})$ is Gaussian, with mean $\mathbf{d}_k$ and covariance $\mathbf{D}_k$ determined by the following recursive algorithm:

Initialize: for $k=N,N-1,\ldots,2$

$$\begin{aligned}\hat{\mathbf{D}}_k &= \Big[\frac{\mathbf{M}^{*T}\mathbf{b}_k^*\mathbf{b}_k^T\mathbf{M}}{\sigma^2}+\mathbf{C}^{-1}\\ &+ |\alpha|^2\left(\mathbf{C}^{-1}-\mathbf{C}^{-1}\hat{\mathbf{D}}_{k+1}\mathbf{C}^{-1}\right)\Big]^{-1}\\ \hat{\mathbf{q}}_k &= \frac{\mathbf{M}^{*T}\mathbf{b}_k^*(r_k-\mathbf{b}_k^T\bar{\mathbf{M}}\bar{\mathbf{p}})}{\sigma^2}+(1-\alpha^*)\mathbf{C}^{-1}\mathbf{c}\\ &+ \alpha^*\mathbf{C}^{-1}\hat{\mathbf{D}}_{k+1}\hat{\mathbf{q}}_{k+1} \quad (53)\end{aligned}$$

Iterate: for $k=1,2,\ldots,N$

$$\begin{aligned}\mathbf{D}_k &= \Big[\frac{\mathbf{M}^{*T}\mathbf{b}_k^*\mathbf{b}_k^T\mathbf{M}}{\sigma^2}+\mathbf{C}^{-1}\\ &+ |\alpha|^2\left(\mathbf{C}^{-1}-\mathbf{C}^{-1}\hat{\mathbf{D}}_{k-1}\mathbf{C}^{-1}-\mathbf{C}^{-1}\hat{\mathbf{D}}_{k+1}\mathbf{C}^{-1}\right)\Big]^{-1}\\ \mathbf{q}_k &= \frac{\mathbf{M}^{*T}\mathbf{b}_k^*(r_k-\mathbf{b}_k^T\bar{\mathbf{M}}\bar{\mathbf{p}})}{\sigma^2}+(1-\alpha^*)\mathbf{C}^{-1}\mathbf{c}\\ &+ \alpha\mathbf{C}^{-1}\hat{\mathbf{D}}_{k-1}\hat{\mathbf{q}}_{k-1}+\alpha^*\mathbf{C}^{-1}\hat{\mathbf{D}}_{k+1}\hat{\mathbf{q}}_{k+1} \quad (54)\\ \mathbf{d}_k &= \mathbf{D}_k\mathbf{q}_k,\ \mathbf{G}_k=\mathbf{M}\mathbf{D}_k\mathbf{M}^{*T},\ \mathbf{g}_k=\mathbf{M}\mathbf{d}_k+\bar{\mathbf{M}}\bar{\mathbf{p}}\end{aligned}$$

$\mathbf{G}_k$ and $\mathbf{g}_k$ from above will be used in the Viterbi algorithm incremental cost function for time k.

Update for next iteration:

$$\begin{aligned}\hat{\mathbf{D}}_k &= \Big[\frac{\mathbf{M}^{*T}\mathbf{b}_k^*\mathbf{b}_k^T\mathbf{M}}{\sigma^2}+\mathbf{C}^{-1}\\ &+ |\alpha|^2\left(\mathbf{C}^{-1}-\mathbf{C}^{-1}\hat{\mathbf{D}}_{k-1}\mathbf{C}^{-1}\right)\Big]^{-1}\\ \hat{\mathbf{q}}_k &= \frac{\mathbf{M}^{*T}\mathbf{b}_k^*(r_k-\mathbf{b}_k^T\bar{\mathbf{M}}\bar{\mathbf{p}})}{\sigma^2}+(1-\alpha^*)\mathbf{C}^{-1}\mathbf{c}\\ &+ \alpha\mathbf{C}^{-1}\hat{\mathbf{D}}_{k-1}\hat{\mathbf{q}}_{k-1}, \quad (55)\end{aligned}$$

with $\hat{\mathbf{D}}_0\equiv\hat{\mathbf{D}}_{N+1}\equiv\mathbf{C}$, $\hat{\mathbf{q}}_0\equiv\left(\frac{1-\alpha^*}{1-\alpha}\right)\mathbf{C}^{-1}\mathbf{c}$, and $\hat{\mathbf{q}}_{N+1}\equiv\mathbf{C}^{-1}\mathbf{c}$.

It is apparent from the above algorithm that $f(\mathbf{p}_k|\mathbf{r},\mathbf{B})$ at any time k depends on all of the received data and estimated symbols from time 1 to N.

For the E-step, the algorithm from (53-55) is used to estimate $\mathbf{G}_k$ and $\mathbf{g}_k$, $k=1,\ldots,N$. For the M-step, the Viterbi algorithm is used with incremental cost function (43) to estimate $\mathbf{A}$.

VI. Simulation Results

For all of the simulations shown here, 20,000 trials were performed for values of SNR from 0 to 10 in steps of 2, and two iterations of the EM algorithms were used. For each trial, $N=12$ random BPSK transmitted symbol values were generated using a uniform distribution, so the symbol values -1 and 1 were equally likely. A channel FIR length of $L=4$ was used, the channel was initialized to the known state [-1 -1 -1], and the true noise variance was used in the algorithms. The Viterbi algorithm was used to estimate the transmitted symbols and produce the BER results. For a reference BER, the Viterbi algorithm with known channel coefficients was used. The EM algorithms were initialized using the symbol values as predicted by running the Viterbi algorithm with the true channel coefficients.

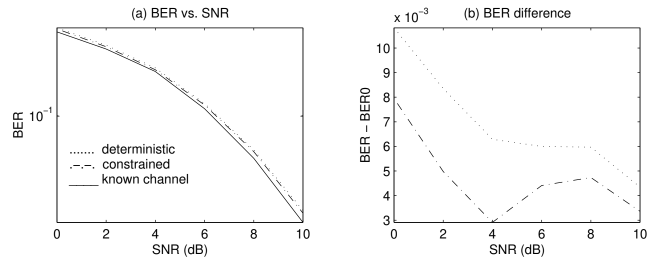

Figure 1: BER vs. SNR for deterministic and constrained deterministic EM algorithms using $N=12,\ L=4$: (a) absolute BER; (b) BER difference from known channel.

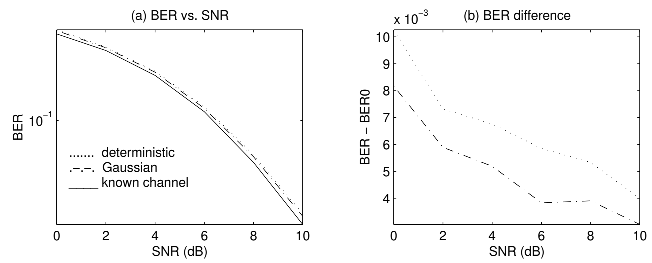

Figure 2: BER vs. SNR for deterministic and constrained Gaussian EM algorithms using $N=12,\ L=4$: (a) absolute BER; (b) BER difference from known channel.

The channel coefficients were constrained to have a double zero at DC using:

$$\mathbf{F}=\begin{bmatrix}1&1&1&1\\0&1&2&3\end{bmatrix},\qquad \mathbf{f}=\begin{bmatrix}0\\0\end{bmatrix}.$$

For the non-time-varying channel, the first set of simulation results confirm that the EM algorithm which incorporates knowledge of the linear constraints produces lower bit-error-rates in the estimated symbols. For the simulation results shown in Figure 1, $\mathbf{p}$ was fixed as $[1\ 1]^T$. The "deterministic" algorithm performs optimal joint sequence and channel estimation [1, 12, 2] without taking the constraints into account.

In the second set of simulations, the underlying channel coefficients for each trial were generated as Gaussian random values with mean 1 and covariance

$$\mathbf{C} = \begin{bmatrix}0.5&-0.5\\-0.5&1\end{bmatrix}.$$

The results shown in Figure 2 demonstrate that when the underlying channel coefficients have a non-time-varying Gaussian distribution, the Gaussian EM algorithm results in lower BER as compared with the EM algorithm for unconstrained deterministic channel coefficients.

For time-varying channel coefficients, Figure 3 shows results from a simulation using the Gaussian EM algorithm. For each trial, the underlying channel coefficients were generated as independent (over time) Gaussian random variables, with mean $\mathbf{c}_k$ changing linearly

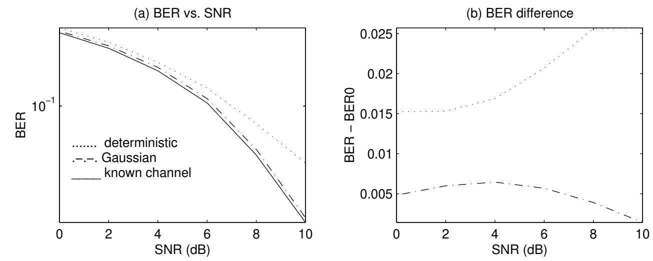

Fig. 3: BER vs. SNR for deterministic and constrained time-varying Gaussian EM algorithms using $N = 12,\ L = 4$: (a) absolute BER; (b) BER difference from known channel.

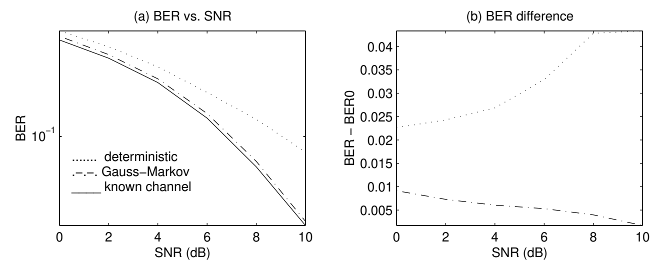

Figure 4: BER vs. SNR for deterministic and constrained Gauss-Markov EM algorithms using $N = 12,\ L = 4,\ \alpha = 0.8$: (a) absolute BER; (b) BER difference from known channel.

from 1 to 2 over the block of time, and fixed covariance

$$\mathbf{C} \quad = \quad \begin{bmatrix} 0.1 & -0.1 \\ -0.1 & 0.2 \end{bmatrix}.$$

The Gaussian EM algorithm performs almost as well as the known channel in this case, and outperforms the deterministic algorithm dramatically as SNR increases.

For Gauss-Markov channel coefficients, Figure 4 shows results from a typical simulation. The underlying channel coefficients were generated randomly using the Gauss-Markov formula (49) with $\mathbf{v}_k$ having mean 0.2 and the same covariance matrix as shown above for the time-varying Gaussian simulation. The Gauss-Markov EM algorithm performs almost as well as the known channel case, and much better than the deterministic algorithm.

VII. Summary and Conclusions

In this paper we have presented a new approach to MLSE over unknown ISI channels. We employ the EM algorithmic approach to marginalize over the channel parameters, given a prior distribution and linear constraints on the channel. Algorithms are developed for a variety of time-invariant and fast time-varying channels. Simulation results illustrate the advantage of this approach, relative to joint sequence/channel estimation.

Acknowledgments

This research was supported by ONR under Grant N00014-98-1-0892.

References

[1] M. Ghosh and C. L. Weber, “Maximum-likelihood blind equalization,” *Optical Engineering*, vol. 31, pp. 1224–1228, June 1992.

[2] N. Seshadri, “Joint data and channel estimation using blind trellis search techniques,” *IEEE Trans. on Communications*, vol. 42, pp. 1000–1011, April 1994.

[3] R. Raheli, A. Polydoros, and C. K. Tzou, “The principle of per-survivor processing: a general approach to approximate and adaptive MLSE,” *IEEE Trans. on Communications*, vol. 43, pp. 354–364, Feb/Mar/Apr 1995.

[4] F. R. Magee, Jr. and J. G. Proakis, “Adaptive maximum-likelihood sequence estimation for digital signaling in the presence of intersymbol interference,” *IEEE Trans. on Information Theory*, pp. 120–124, Jan 1973.

[5] A. P. Dempster, N. M. Laird, and D. B. Rubin, “Maximum likelihood from incomplete data via the EM algorithm,” *J. Roy. Stat. Soc.*, vol. 39, no. 1, pp. 1–38, 1977.

[6] G. K. Kaleh and R. Vallet, “Joint parameter estimation and symbol detection for linear or nonlinear unknown channels,” *IEEE Trans. on Communications*, vol. 42, pp. 2406–2413, Jul 1994.

[7] M. Erkurt and J. G. Proakis, “Joint data detection and channel estimation for rapidly fading channels,” *GLOBECOM’92 IEEE Global Telecommunications Conference, Orlando, Fla.*, pp. 910–914, Dec 1992.

[8] K. J. Molnar, “An iterative approach for joint channel and data estimation for time-varying multipath channels,” *1998 DSP Workshop, Bryce, Utah*, Aug 1998.

[9] C. N. Geroghiades and D. L. Snyder, “The expectation-maximization algorithm for symbol unsynchronized sequence detection,” *IEEE Trans. on Communications*, vol. 39, pp. 54–61, Jan 1991.

[10] L. B. Nelson and H. V. Poor, “Iterative multiuser receivers for CDMA channels: An EM-based approach,” *IEEE Trans. on Communications*, vol. 44, pp. 1700–1710, Dec 1996.

[11] L. M. Zeger and H. Kobayashi, “Sequence estimation in GSM using the EM algorithm,” *CISS’98 Conference on Information Sciences and Systems, Princeton, NJ*, 1998.

[12] K. Chang and C. N. Georghiades, “Iterative joint sequence and channel estimation for fast time-varying intersymbol interference channels,” *ICC’95 IEEE Int. Conf. on Communications*, June 1995.

[13] G. D. Forney, “The viterbi algorithm,” *Proc. of the IEEE*, vol. 61, pp. 268–278, Mar 1973.

[14] M. Stojanovic and Z. Zvonar, “Performance of multiuser diversity reception in rayleigh fading CDMA channels,” *IEEE Trans. on Communications*, vol. 47, pp. 356–359, Mar 1999.

[15] R. Perry, W. Berger, and K. Buckley, “EM algorithm for sequence estimation over Gauss–Markov ISI channels,” *ICC 2000 Conference*, June 2000.