EM Algorithms for Sequence Estimation over Random ISI Channels*

Richard Perry, W. Andrew Berger† and Kevin Buckley
ECE Department, Villanova University, Villanova, PA 19085
†Physics/EE Department, University of Scranton, Scranton, PA 18510
perry@ece.vill.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 random channel coefficients. By using the EM formulation to marginalize over the 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.

1 Introduction

In this paper we address the problem of estimation of a sequence of digital communication symbols transmitted over random ISI channels. For a known FIR channel, it is well known that the Viterbi algorithm [1] solves the ML sequence estimation (MLSE) problem. Although the computationally efficiency of this algorithm has lead to its broad use, the Viterbi algorithm requires knowledge of the channel (e.g. its impulse response) [2]. Here we are concerned with solving the blind MLSE problem, when the channel is unknown and no training data is employed.

Sequence estimation over a channel with unknown coefficients can be formulated as either a joint sequence/channel estimation problem, a channel estimation problem (followed by MLSE), or a direct sequence estimation problem. For joint estimation of both the transmitted sequence and channel coefficients, optimum methods have been developed (e.g. [3]). In this paper we develop an algorithm which accounts for random channel parameters and we pursue MLSE by marginalizing over the channel coefficients in order to improve sequence estimation performance over joint sequence/channel estimators.

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. [4]. For digital communications, the EM approach has been applied to the channel estimation problem [5, 6]. It has also been employed for MLSE with unknown nuisance channel, interference or receiver parameters. For example, Georghiades, et al. [7] use EM to account for unknown phase given unsynchronized reception. Zeger and Kobayashi [8] estimate GMSK modulated symbols in the presence of unknown multipath parameters. In these examples, either simple prior distributions on the nuisance parameters are assumed in order to develop algorithms, or a high SNR approximation is derived.

This paper presents a new approach using an EM derived algorithm for MLSE over channels having unknown FIR impulse coefficients. The ML formulation for optimal sequence estimation over random channels is generally intractable to solve directly. However, the EM algorithm is a technique which can effectively be used to solve such intractable ML problems. Specifically, by using EM to marginalize over the channel coefficient distribution, optimal estimates of the transmitted sequence are derived which can achieve lower bit-error-rates than methods which jointly estimate both the transmitted sequence and the channel coefficients, and methods based on first optimally estimating the channel coefficients.

2 Communication System Model

In the discrete-time FIR model of a 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:

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

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, $\mathbf{h}$ is a complex column vector containing the channel impulse response coefficients (which are unknown), 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. Using matrix-vector notation, the above equations can then be expressed more simply as:

$$\mathbf{r} = \mathbf{A}\mathbf{h} + \mathbf{n}, \tag{2}$$

and

$$f(\mathbf{r}|\mathbf{h}, \mathbf{A}) = \frac{1}{(\pi\sigma^2)^N} \, e^{-\frac{|\mathbf{r}-\mathbf{A}\mathbf{h}|^2}{\sigma^2}}. \tag{3}$$

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

3 EM and ML Sequence Estimation

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

$$\begin{aligned} f(\mathbf{A}|\mathbf{r}) &= \frac{f(\mathbf{r}|\mathbf{A})f(\mathbf{A})}{f(\mathbf{r})} \\ &= K_4 f(\mathbf{r}|\mathbf{A})f(\mathbf{A}) \doteq f(\mathbf{r}|\mathbf{A})f(\mathbf{A}), \end{aligned} \tag{4}$$

where K_4 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 random channel coefficients is:

$$f(\mathbf{r}|\mathbf{A}) = E[f(\mathbf{r}|\mathbf{h},\mathbf{A})] = \int_{\mathbf{h}} f(\mathbf{r}|\mathbf{h},\mathbf{A})f(\mathbf{h})d\mathbf{h}, \tag{5}$$

where $f(\mathbf{r}|\mathbf{h},\mathbf{A})$ is given by (3). 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 (5) seems tractable is if $\mathbf{h}$ is deterministic. Then $f(\mathbf{h}) = \delta(\mathbf{h}-\mathbf{g})$ and (5) reduces to (3) with $\mathbf{h}=\mathbf{g}$, where $\mathbf{g}$ is the deterministic but unknown value of $\mathbf{h}$, as discussed further in Section 4.1. This provides a theoretical basis for the joint $(\mathbf{h},\mathbf{A})$ estimation algorithms such as [3]. However for arbitrary $f(\mathbf{h})$ we must work with (5) directly, or use an EM algorithm, in order to produce optimal estimates of $\mathbf{A}$.

To maximize (5) using EM, define the auxiliary function [4]:

$$\begin{aligned} Q(\mathbf{A}|\mathbf{B}) &= E[\log(f(\mathbf{r},\mathbf{h}|\mathbf{A}))\,|\mathbf{r},\mathbf{B}] \\ &= \int_{\mathbf{h}} \log(f(\mathbf{r},\mathbf{h}|\mathbf{A}))f(\mathbf{h}|\mathbf{r},\mathbf{B})d\mathbf{h}, \end{aligned} \tag{6}$$

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, $\mathbf{h}$ is the hidden data, and $(\mathbf{r},\mathbf{h})$ 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.

[1] As a matter of notation, we will subscript normalization constants such as K_4 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.

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

$$f(\mathbf{r},\mathbf{h}|\mathbf{A}) = f(\mathbf{r}|\mathbf{h},\mathbf{A})f(\mathbf{h}|\mathbf{A}). \tag{7}$$

We can drop the $f(\mathbf{h}|\mathbf{A})$ term from this since $\mathbf{h}$ and $\mathbf{A}$ are independent so $f(\mathbf{h}|\mathbf{A})$ is not a function of $\mathbf{A}$ and will not affect the maximization in the M-step. $Q(\mathbf{A}|\mathbf{B})$ from (6) may then be written as:

$$\begin{aligned} Q(\mathbf{A}|\mathbf{B}) &\doteq E[\log(f(\mathbf{r}|\mathbf{h},\mathbf{A}))\,|\mathbf{r},\mathbf{B}] \\ &= \int_{\mathbf{h}} \log(f(\mathbf{r}|\mathbf{h},\mathbf{A}))f(\mathbf{h}|\mathbf{r},\mathbf{B})d\mathbf{h}. \end{aligned} \tag{8}$$

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

$$f(\mathbf{h}|\mathbf{r},\mathbf{B}) = \frac{f(\mathbf{r}|\mathbf{h},\mathbf{B})f(\mathbf{h}|\mathbf{B})}{f(\mathbf{r}|\mathbf{B})}, \tag{9}$$

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 $\mathbf{h}$ or $\mathbf{A}$, so will not affect the maximization step. Also, $f(\mathbf{h}|\mathbf{B}) = f(\mathbf{h})$ since the channel coefficients and transmitted data are independent. But note that given $\mathbf{r}$ and $\mathbf{B}$, $f(\mathbf{h}|\mathbf{r},\mathbf{B}) \neq f(\mathbf{h})$.

Therefore, for $f(\mathbf{h}|\mathbf{r},\mathbf{B})$ in (8) we may use:

$$f(\mathbf{h}|\mathbf{r},\mathbf{B}) \doteq f(\mathbf{r}|\mathbf{h},\mathbf{B})f(\mathbf{h}). \tag{10}$$

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 $\mathbf{h}$ conditioned on $\mathbf{r}$ and $\mathbf{B}$. This enables the Viterbi algorithm [1] 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 channel coefficient distributions.

4 EM Algorithms for Specific Random Channels

Using (3) in (10), (8) becomes:

$$Q(\mathbf{A}|\mathbf{B}) \doteq E[-|\mathbf{r}-\mathbf{A}\mathbf{h}|^2\,|\mathbf{r},\mathbf{B}] = \int_{\mathbf{h}} -|\mathbf{r}-\mathbf{A}\mathbf{h}|^2 f(\mathbf{h}|\mathbf{r},\mathbf{B})d\mathbf{h}, \tag{11}$$

with

$$f(\mathbf{h}|\mathbf{r},\mathbf{B}) = K_{12}\, e^{-\frac{|\mathbf{r}-\mathbf{B}\mathbf{h}|^2}{\sigma^2}} f(\mathbf{h}), \tag{12}$$

where K_{12} can be determined if needed by setting the integral of $f(\mathbf{h}|\mathbf{r},\mathbf{B})$ to 1.

To evaluate (11) we only need the mean and covariance of $\mathbf{h}$ given $\mathbf{r}$ and $\mathbf{B}$ using the distribution of (12). We will derive these for specific cases of $f(\mathbf{h})$ subsequently; for now, we define them symbolically as $\mathbf{g}$ and $\mathbf{G}$:

$$\mathbf{g} \triangleq E[\mathbf{h}|\mathbf{r},\mathbf{B}] \tag{13}$$

$$\mathbf{G} \triangleq E[(\mathbf{h}-\mathbf{g})(\mathbf{h}-\mathbf{g})^{*T}\,|\mathbf{r},\mathbf{B}]. \tag{14}$$

The noise variance σ^2 is needed in general to compute 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 (11), 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 (11) in terms of the incremental costs:

$$Q(\mathbf{A}|\mathbf{B}) \doteq -\sum_{k=1}^{N} V_k, \tag{15}$$

where V_k is:

$$V_k = E[|r_k - \mathbf{a}_k^T \mathrm{h}|^2 \, |\mathrm{r}, \mathbf{B}]. \tag{16}$$

The expected value is taken over h using (12) for $f(\mathrm{h}|\mathrm{r}, \mathbf{B})$.

Expanding the norm squared:

$$\begin{aligned} |r_k - \mathbf{a}_k^T \mathrm{h}|^2 &= |r_k|^2 - r_k \mathbf{a}_k^{*T} \mathrm{h}^* - r_k^* \mathbf{a}_k^T \mathrm{h} + \mathbf{a}_k^T \mathrm{h}\mathrm{h}^{*T} \mathbf{a}_k^* \\ &= |r_k|^2 - r_k \mathbf{a}_k^{*T} \mathrm{h}^* - r_k^* \mathbf{a}_k^T \mathrm{h} \\ &\quad + \mathbf{a}_k^T (\mathrm{h} - \mathrm{g})(\mathrm{h} - \mathrm{g})^{*T} \mathbf{a}_k^* \\ &\quad + \mathbf{a}_k^T \mathrm{h}\mathrm{g}^{*T} \mathbf{a}_k^* + \mathbf{a}_k^T \mathrm{g}\mathrm{h}^{*T} \mathbf{a}_k^* - \mathbf{a}_k^T \mathrm{g}\mathrm{g}^{*T} \mathbf{a}_k^*. \end{aligned} \tag{17}$$

Taking the expected value over h:

$$\begin{aligned} V_k &= |r_k|^2 - r_k \mathbf{a}_k^{*T} \mathrm{g}^* - r_k^* \mathbf{a}_k^T \mathrm{g} + \mathbf{a}_k^T \mathbf{G} \mathbf{a}_k^* + \mathbf{a}_k^T \mathrm{g}\mathrm{g}^{*T} \mathbf{a}_k^* \\ &= |r_k - \mathbf{a}_k^T \mathrm{g}|^2 + \mathbf{a}_k^T \mathbf{G} \mathbf{a}_k^*. \end{aligned} \tag{18}$$

Equation (18) represents the general form of the Viterbi algorithm incremental cost function for random ISI channels. To use (18) in an EM algorithm, g and $\mathbf{G}$ from (13) and (14) are computed in the M-step; these computations depend on the specific form of the apriori channel distribution $f(\mathrm{h})$. The following subsections examine the specific cases of deterministic, uniform, and Gaussian $f(\mathrm{h})$.

4.1 Deterministic Channel Coefficients

The case of deterministic channel coefficients is examined first because it is the simplest case, and because it corresponds to the assumptions made by previously published algorithms [3] 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 $E[\mathrm{h}|\mathrm{r}, \mathbf{B}]$.

If h is not random, then let $f(\mathrm{h}) = \delta(\mathrm{h} - \mathrm{g})$ where g is the deterministic but unknown value of h, and $\delta()$ represents an impulse function. In this case, $E[\mathrm{h}] = \mathrm{g}$, and using (12) to estimate g does not provide anything useful since it reduces to:

$$\begin{aligned} E[\mathrm{h}|\mathrm{r}, \mathbf{B}] &= \int_{\mathrm{h}} \mathrm{h} \; K_{12} \, e^{-\frac{|\mathrm{r} - \mathbf{B}\mathrm{h}|^2}{\sigma^2}} \delta(\mathrm{h} - \mathrm{g}) d\mathrm{h} \\ &= \mathrm{g} \, K_{12} \, e^{-\frac{|\mathrm{r} - \mathbf{B}\mathrm{g}|^2}{\sigma^2}} = \mathrm{g}, \end{aligned} \tag{19}$$

and g is unknown.

So to estimate g in this case, we must use an ad-hoc method. Note that given r and $\mathbf{B}$, h satisfies (2):

$$\mathrm{r} = \mathbf{B}\mathrm{h} + \mathrm{n}. \tag{20}$$

Taking the expected value of the right-hand-side of (20) with respect to h and n yields:

$$\mathrm{r} = \mathbf{B}\mathrm{g}, \tag{21}$$

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

$$\mathrm{g} = \mathbf{B}^+ \mathrm{r}. \tag{22}$$

Note that this estimate of g is also a maximum-likelihood estimate of h given r and $\mathbf{B}$ since it maximizes $f(\mathrm{r}|\mathrm{h}, \mathbf{B})$ (i.e. minimizes $|\mathrm{r} - \mathbf{B}\mathrm{h}|^2$). Although producing an optimal estimate of the channel coefficients was not a goal in this EM algorithm, it does so anyway as a byproduct of the E-step.

In this case, $\mathbf{G} = 0$, and the Viterbi algorithm incremental cost function from (18) becomes:

$$V_k = |r_k - \mathbf{a}_k^T \mathrm{g}|^2 = |r_k - \mathbf{a}_k^T \mathbf{B}^+ \mathrm{r}|^2. \tag{23}$$

Referring to the general EM algorithm described above, in the E-step use (22) to estimate g and in the M-step use the Viterbi algorithm with incremental cost function (23) to estimate $\mathbf{A}$.

In step 1 of the EM algorithm, we could initialize $\mathbf{B}$ randomly, or to 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 $\mathrm{g} = [0 \; \ldots \; 0 \; 1 \; 0 \; \ldots \; 0]^T$ in (23), then using the Viterbi algorithm to estimate $\mathbf{A}$, then setting $\mathbf{B}$ to this estimate of $\mathbf{A}$.

This EM algorithm is equivalent to previously published algorithms for optimal joint sequence and channel estimation [3], but the derivation here provides a mathematical basis for understanding this algorithm in the context of MLSE by marginalizing over the channel parameters.

4.2 Uniform Channel Coefficients

The previous section derived an EM algorithm for the case of deterministic channel coefficients, using an impulse function for $f(\mathrm{h})$. In this section we analyze a similar situation, where no information about h is available, but here we use a uniform distribution for $f(\mathrm{h})$. Although impulse and uniform distributions for $f(\mathrm{h})$ do not seem similar, they are similar in this application because the location of the impulse (mean of h) is unknown in the previous section.

If $f(\mathrm{h})$ is uniform over some finite interval, then computation of g and $\mathbf{G}$ could be performed numerically using $f(\mathrm{h}|\mathrm{r}, \mathbf{B})$ from (12). But if we let the domain of h be infinite, then we can compute g and $\mathbf{G}$ analytically. In this case $f(\mathrm{h})$ is improper, since it would have to be zero for the area under it to be 1. Nevertheless, assuming that $f(\mathrm{h})$ is uniform and of infinite extent is useful, since it is practically equivalent to letting $f(\mathrm{h})$ be uniform over some range of several standard deviations for which the exponential term in (12) is significant. Outside of some finite region, the exponential term in (12) is practically zero.

So we propose that letting $f(\mathrm{h}) = 1$ in (12) is reasonable, which reduces (12) to:

$$f(\mathrm{h}|\mathrm{r}, \mathbf{B}) = K_{12} \, e^{-\frac{|\mathrm{r} - \mathbf{B}\mathrm{h}|^2}{\sigma^2}}. \tag{24}$$

To put this into a more useful form, expand the norm-squared term from (24):

$$\begin{aligned}|\mathbf{r}-\mathbf{B}\mathbf{h}|^2 &= \mathbf{r}^{*T}\mathbf{r}-\mathbf{r}^{*T}\mathbf{B}\mathbf{h}-\mathbf{h}^{*T}\mathbf{B}^{*T}\mathbf{r}+\mathbf{h}^{*T}\mathbf{B}^{*T}\mathbf{B}\mathbf{h} \\ &= \mathbf{r}^{*T}\mathbf{r}-\mathbf{r}^{*T}\mathbf{B}\mathbf{h}-\mathbf{h}^{*T}\mathbf{B}^{*T}\mathbf{r} \\ &\quad +(\mathbf{h}-\mathbf{g})^{*T}\mathbf{B}^{*T}\mathbf{B}(\mathbf{h}-\mathbf{g}) \\ &\quad +\mathbf{h}^{*T}\mathbf{B}^{*T}\mathbf{B}\mathbf{g}+\mathbf{g}^{*T}\mathbf{B}^{*T}\mathbf{B}\mathbf{h}-\mathbf{g}^{*T}\mathbf{B}^{*T}\mathbf{B}\mathbf{g},\end{aligned} \quad (25)$$

for any $\mathbf{g}$.

Now, if we let $\mathbf{g}=\mathbf{B}^{+}\mathbf{r}$, and if $\mathbf{B}$ has full rank so that $\mathbf{B}^{+}=(\mathbf{B}^{*T}\mathbf{B})^{-1}\mathbf{B}^{*T}$, then (25) reduces to:

$$|\mathbf{r}-\mathbf{B}\mathbf{h}|^2=\mathbf{r}^{*T}\mathbf{r}-\mathbf{g}^{*T}\mathbf{B}^{*T}\mathbf{r}+(\mathbf{h}-\mathbf{g})^{*T}\mathbf{B}^{*T}\mathbf{B}(\mathbf{h}-\mathbf{g}), \quad (26)$$

and (24) may be written as:

$$f(\mathbf{h}|\mathbf{r},\mathbf{B}) \doteq e^{-\frac{(\mathbf{h}-\mathbf{g})^{*T}\mathbf{B}^{*T}\mathbf{B}(\mathbf{h}-\mathbf{g})}{\sigma^2}}. \quad (27)$$

This shows that, given $\mathbf{r}$ and $\mathbf{B}$, $\mathbf{h}$ has a joint Gaussian distribution with mean $\mathbf{g}=\mathbf{B}^{+}\mathbf{r}$ and covariance:

$$\mathbf{G}=\left[\frac{\mathbf{B}^{*T}\mathbf{B}}{\sigma^2}\right]^{-1}. \quad (28)$$

If $\mathbf{B}$ does not have full rank then some simplification of (25) into a form similar to (26) may be performed, but that is not pursued here. In the case of data values which do not contain 0's, e.g. (-1,1) data values, $\mathbf{B}$ will always have full rank if $a_j=0$ for $j\le 0$, due to the 0's in the first $L-1$ rows which make the columns of $\mathbf{B}$ linearly independent.

Therefore, for the case of uniformly distributed channel coefficients, the EM algorithm is almost the same as for deterministic channel coefficients, except that the Viterbi algorithm incremental cost function from (18), which includes $\mathbf{G}$, is used instead of (23), and $\mathbf{G}$ is given by (28).

To initialize $\mathbf{B}$, we can initialize $\mathbf{g}$ and $\mathbf{G}$, then use the Viterbi algorithm to estimate $\mathbf{A}$, then set $\mathbf{B}$ to this estimate of $\mathbf{A}$. The initial estimate of $\mathbf{G}$ depends on the set of transmitted sequence values and initial channel conditions. For example, in the case of (-1,1) transmitted data values, with $a_j=0$ for $j\le 0$, $\mathbf{G}=\text{diag}\left[\frac{\sigma^2}{N-i+1}\right]$, $i=1,\ldots,L$ may be used, based on just the diagonal elements of $\mathbf{B}^{*T}\mathbf{B}$.

4.3 Gaussian Channel Coefficients

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

$$f(\mathbf{h})=\frac{1}{\pi^N|\mathbf{C}|}e^{-(\mathbf{h}-\mathbf{d})^{*T}\mathbf{C}^{-1}(\mathbf{h}-\mathbf{d})}, \quad (29)$$

where $\mathbf{d}$ and $\mathbf{C}$ are known. In this case, (12) reduces to:

$$\begin{aligned}f(\mathbf{h}|\mathbf{r},\mathbf{B}) &= K_{12}\,e^{-\frac{|\mathbf{r}-\mathbf{B}\mathbf{h}|^2}{\sigma^2}}\,\frac{1}{\pi^N|\mathbf{C}|}\,e^{-(\mathbf{h}-\mathbf{d})^{*T}\mathbf{C}^{-1}(\mathbf{h}-\mathbf{d})} \\ &\doteq e^{-(\mathbf{h}-\mathbf{g})^{*T}\mathbf{G}^{-1}(\mathbf{h}-\mathbf{g})},\end{aligned} \quad (30)$$

with:

$$\mathbf{G} = \left[\frac{\mathbf{B}^{*T}\mathbf{B}}{\sigma^2}+\mathbf{C}^{-1}\right]^{-1} \quad (31)$$

$$\mathbf{g} = \mathbf{G}\left(\frac{\mathbf{B}^{*T}\mathbf{r}}{\sigma^2}+\mathbf{C}^{-1}\mathbf{d}\right) \quad (32)$$

So in this case, the distribution of $\mathbf{h}$ given $\mathbf{r}$ and $\mathbf{B}$ is also Gaussian. Note that the formula in (32) for the mean of $\mathbf{h}$ given $\mathbf{r}$ and $\mathbf{B}$ does not use $\mathbf{B}^{+}\mathbf{r}$ as in the previous deterministic and uniform cases.

The Gaussian distribution of $\mathbf{h}$ is intermediate between the known deterministic channel case and uniform unknown $\mathbf{h}$. In the limit as $\mathbf{C}$ goes to 0, $\mathbf{G}$ approaches $\mathbf{C}$ and $\mathbf{g}$ approaches known $\mathbf{d}$, which corresponds to the Gaussian distribution approaching $\delta(\mathbf{h}-\mathbf{d})$. This is similar to the case of deterministic $\mathbf{h}$ from Section 4.1 except that here the mean is known whereas in Section 4.1 it was estimated using (22). In the limit as $\mathbf{C}$ goes to infinity, $\mathbf{G}$ approaches $\left[\frac{\mathbf{B}^{*T}\mathbf{B}}{\sigma^2}\right]^{-1}$ and $\mathbf{g}$ approaches $\mathbf{B}^{+}\mathbf{r}$, 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 (31) can dominate the $\mathbf{C}^{-1}$ term, so then $\mathbf{G}$ approaches $\left[\frac{\mathbf{B}^{*T}\mathbf{B}}{\sigma^2}\right]^{-1}$ which approaches 0, and $\mathbf{g}$ approaches $\mathbf{B}^{+}\mathbf{r}$, which corresponds to the Gaussian EM algorithm becoming equivalent to the algorithm for deterministic channel coefficients.

Therefore, for the case of Gaussian distributed channel coefficients, the EM algorithm is almost the same as for uniform channel coefficients, except that the E-step uses (31) and (32) to estimate $\mathbf{G}$ and $\mathbf{g}$.

To use the Viterbi algorithm to initialize $\mathbf{B}$ in step 1, we could, for example, initialize $\mathbf{G}=\mathbf{C}$ and $\mathbf{g}=\mathbf{d}$.

5 Simulation Results

For the simulations shown here, 20,000 trials were performed for values of SNR from 0 to 10 in steps of 2, and 4 iterations of the EM algorithms were used. For each trial, 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 $L=3$ was used, the channel coefficients were generated randomly for each trial, and the channel was initialized to the known state [-1 -1]. The true noise variance was used for all algorithms.

The Viterbi algorithm, with truncation of the Viterbi trellis at $N-L$, was used to estimate the transmitted symbols and produce the BER results. This means that for $N=10$ and $L=3$, only errors in the first 7 symbols were used in computing BER. For a reference BER, the Viterbi algorithm with known channel coefficients was used.

For comparison purposes, the EM algorithm from [5] was used, and is referred to as "Kaleh–Vallet" in the simulation results. The Kaleh–Vallet algorithm produces optimal estimates of the channel coefficients for data blocks which are short enough such that the channel coefficients are non-time-varying over the block. Also for comparison, the algorithm referred to as "deterministic" is that from Section 4.1, and is equivalent to optimal joint sequence and

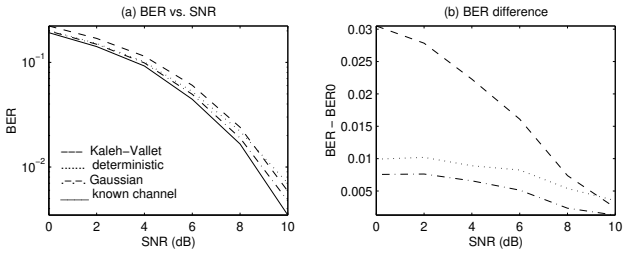

Figure 1. BER vs. SNR for deterministic, Gaussian, and Kaleh–Vallet EM algorithms using $N = 10$, $L = 3$: (a) absolute BER; (b) BER difference from known channel.

channel estimation [3]. Our EM algorithm for Gaussian channel coefficient distributions (from Section 4.3 is referred to as "Gaussian", and produces optimal estimates of the symbols by marginalizing over the channel coefficient distribution.

The deterministic and Gaussian EM algorithms require an initial estimate of the transmitted symbols. For the simulations shown here, these EM algorithms were initialized using the symbol values as predicted by running the Viterbi algorithm with the true channel coefficients. The Kaleh–Vallet EM algorithm requires an initial estimate of the channel coefficients, and was initialized using the true channel coefficients. In the simulations presented here, the EM algorithms converged sufficiently within just about two or three iterations, so four iterations were used for all of the Monte-Carlo simulation runs.

The simulation results in Figure 1 confirm that knowledge of $f(\mathbf{h})$ as used in the EM algorithm for Gaussian $\mathbf{h}$ can produce lower bit-error-rates in the estimated symbols. When $\mathbf{h}$ has a Gaussian distribution, the EM algorithm for Gaussian $\mathbf{h}$ from Section 4.3 results in lower BER as compared with the EM algorithm for deterministic $\mathbf{h}$ from Section 4.1, and also performs better than the Kaleh–Vallet algorithm. For this simulation, the channel coefficients were generated as Gaussian random values with mean 1 and covariance:

$$\mathbf{C} = \begin{bmatrix} 0.5 & -0.5 & 0 \\ -0.5 & 1 & -0.5 \\ 0 & -0.5 & 1 \end{bmatrix}.$$

6 Summary and Conclusions

In this paper we have derived EM algorithms for MLSE over random channels. This new approach produces optimal estimates of the transmitted symbols by marginalizing over the random channel coefficient distribution. Simulation results demonstrate a BER performance improvement when using our EM algorithms.

References

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

[2] J. G. Proakis, *Digital Communications, 3–rd ed.* McGraw-Hill, 1995.

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

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] 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.