EM Algorithm for Sequence Estimation over Gauss-Markov 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.villanova.edu, (610)-519-4969, FAX: (610)-519-4436

Abstract— **This paper presents a new algorithm, based on an EM (Expectation-Maximization) formulation, for ML (maximum likelihood) sequence estimation over unknown ISI (inter-symbol interference) channels with random channel coefficients which have a Gauss-Markov fast time-varying distribution. By using the EM formulation to marginalize over the channel coefficient distribution, maximum-likelihood estimates of the transmitted sequence are obtained. This EM algorithm is 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. Cozzo and Hughes [12] consider multiple transmitter/receiver antenna systems and memoryless channels, and investigate the use of interlaced pilot signals. Zamiri-Jafarian and Pasupathy [13] derive an interesting on-line, suboptimal, joint channel/symbol estimator, and relate it to per-survivor and per-branch processing. In these examples, either simple prior distributions on the unknown secondary parameters are assumed in order to develop algorithms, or approximations (e.g. high SNR or on-line) are derived.

In this paper we describe 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 a prior distribution on the channel parameters. In [14] we describe how to incorporate linear channel constraints into this EM formulation. Here, we focus on a Gauss-Markov model for random FIR channels which can have significant variation over a short temporal block.

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:

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

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.

Let $\mathbf{H} = [\mathbf{h}_1, \ldots, \mathbf{h}_N]$ represent the matrix of channel coefficient vectors over time arranged by columns, $\mathbf{A} = [\mathbf{a}_1, \ldots, \mathbf{a}_N]^T$ the matrix of transmitted data arranged by rows, and $\mathbf{r} = [r_1, \ldots, r_N]^T$ the column vector of received data. With this notation, the probability density function of the received data, given $\mathbf{H}$ and $\mathbf{A}$, is:

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

III. EM and ML Sequence Estimation

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

$$f(\mathbf{A}|\mathbf{r}) = \frac{f(\mathbf{r}|\mathbf{A}) f(\mathbf{A})}{f(\mathbf{r})} = K_3 f(\mathbf{r}|\mathbf{A}) f(\mathbf{A}) \doteq f(\mathbf{r}|\mathbf{A}) f(\mathbf{A}), \tag{3}$$

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

where K_3 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. The EM algorithm derived in the next section produces an iterative solution to the ML problem of maximizing $f(\mathbf{r}|\mathbf{A})$. But if $f(\mathbf{A})$ is known and not uniform, then it could be incorporated into the EM algorithm 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:

$$f(\mathbf{r}|\mathbf{A}) = E[f(\mathbf{r}|\mathbf{H},\mathbf{A})] = \int_{\mathbf{H}} f(\mathbf{r}|\mathbf{H},\mathbf{A})f(\mathbf{H})d\mathbf{H}, \quad (4)$$

where $f(\mathbf{r}|\mathbf{H},\mathbf{A})$ is given by (2). 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.

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

$$\begin{aligned} Q(\mathbf{A}|\mathbf{B}) &= E[\log(f(\mathbf{r},\mathbf{H}|\mathbf{A}))\ |\mathbf{r},\mathbf{B}] \\ &= \int_{\mathbf{H}} \log(f(\mathbf{r},\mathbf{H}|\mathbf{A}))f(\mathbf{H}|\mathbf{r},\mathbf{B})d\mathbf{H}, \quad (5) \end{aligned}$$

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.

It can be shown [14], [15] that $Q(\mathbf{A}|\mathbf{B})$ from (5) may be written as:

$$Q(\mathbf{A}|\mathbf{B}) \doteq \int_{\mathbf{H}} \log(f(\mathbf{r}|\mathbf{H},\mathbf{A}))f(\mathbf{r}|\mathbf{H},\mathbf{B})f(\mathbf{H})d\mathbf{H}. \quad (6)$$

In the specific case of the Gauss-Markov channel distribution, 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 [16] to be used for the M-step.

IV. EM for Gauss-Markov Channels

Using (2) in (6) and dropping constants:

$$Q(\mathbf{A}|\mathbf{B}) \doteq \int_{\mathbf{H}} -\sum_{k=1}^{N} |r_k - \mathbf{a}_k^T\mathbf{h}_k|^2\ f(\mathbf{H}|\mathbf{r},\mathbf{B})d\mathbf{H}, \quad (7)$$

[1] As a matter of notation, throughout this paper we 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:

$$f(\mathbf{H}|\mathbf{r},\mathbf{B}) \doteq f(\mathbf{H}) \prod_{k=1}^{N} e^{-\frac{|r_k - \mathbf{b}_k^T\mathbf{h}_k|^2}{\sigma^2}}. \quad (8)$$

As an intermediate development, first assume that the elements of $\mathbf{H}$ are independent over time. Then $f(\mathbf{H})$ can be written as $f(\mathbf{H}) = \prod_{k=1}^{N} f(\mathbf{h}_k)$, and $f(\mathbf{H}|\mathbf{r},\mathbf{B})$ becomes:

$$f(\mathbf{H}|\mathbf{r},\mathbf{B}) = \prod_{k=1}^{N} f(\mathbf{h}_k|\mathbf{r},\mathbf{B}), \quad (9)$$

with:

$$f(\mathbf{h}_k|\mathbf{r},\mathbf{B}) \doteq f(\mathbf{h}_k)\ e^{-\frac{|r_k - \mathbf{b}_k^T\mathbf{h}_k|^2}{\sigma^2}}. \quad (10)$$

Using (9) and (10) in (7) we obtain:

$$Q(\mathbf{A}|\mathbf{B}) \doteq -\sum_{k=1}^{N} V_k, \quad (11)$$

with:

$$\begin{aligned} \mathbf{g}_k &\triangleq E[\mathbf{h}_k|\mathbf{r},\mathbf{B}] & (12) \\ \mathbf{G}_k &\triangleq E[(\mathbf{h}_k - \mathbf{g}_k)(\mathbf{h}_k - \mathbf{g}_k)^{*T}|\mathbf{r},\mathbf{B}] & (13) \\ V_k &\triangleq |r_k - \mathbf{a}_k^T\mathbf{g}_k|^2 + \mathbf{a}_k^T\mathbf{G}_k\mathbf{a}_k^*. & (14) \end{aligned}$$

V_k represents the Viterbi algorithm incremental cost function at time k.

Now, 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 [17]. Let the time-varying channel coefficients follow a Gauss-Markov distribution such that, at time k:

$$\mathbf{h}_k = \alpha\mathbf{h}_{k-1} + \mathbf{v}_k, \quad (15)$$

where $\mathbf{v}_k$ is complex, white, and Gaussian with mean $\mathbf{d}$ and covariance $\mathbf{C}$. In this case, $f(\mathbf{h}_k|\mathbf{h}_{k-1})$ is Gaussian with mean $\mathbf{d}+\alpha\mathbf{h}_{k-1}$ and covariance $\mathbf{C}$. Additional properties of this distribution are derived in the appendix.

(11) and the definitions from (12-14) still apply, with:

$$f(\mathbf{H}|\mathbf{r},\mathbf{B}) \doteq f(\mathbf{h}_1) \prod_{k=2}^{N} f(\mathbf{h}_k|\mathbf{h}_{k-1}) \prod_{k=1}^{N} e^{-\frac{|r_k - \mathbf{b}_k^T\mathbf{h}_k|^2}{\sigma^2}}. \quad (16)$$

To determine $f(\mathbf{h}_k|\mathbf{r},\mathbf{B})$, and thereby $\mathbf{g}_k$ and $\mathbf{G}_k$, we must integrate (16) over $\mathbf{h}_i, i = 1,\ldots,N, i \neq k$. The derivation of this integration result is shown in the appendix. The result is that $f(\mathbf{h}_k|\mathbf{r},\mathbf{B})$ is Gaussian, with mean $\mathbf{g}_k$ and covariance $\mathbf{G}_k$ determined by the following recursive algorithm:

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

$$\begin{aligned} \hat{\mathbf{G}}_k &= \left[\frac{\mathbf{b}_k^*\mathbf{b}_k^T}{\sigma^2} + \mathbf{C}^{-1} \right. \\ &\left. \quad + |\alpha|^2 \left(\mathbf{C}^{-1} - \mathbf{C}^{-1}\hat{\mathbf{G}}_{k+1}\mathbf{C}^{-1}\right)\right]^{-1} \end{aligned}$$

$$\hat{\mathbf{q}}_k = \frac{r_k \mathbf{b}_k^*}{\sigma^2} + (1-\alpha^*)\mathbf{C}^{-1}\mathbf{d} + \alpha^* \mathbf{C}^{-1}\hat{\mathbf{G}}_{k+1}\hat{\mathbf{q}}_{k+1} \qquad (17)$$

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

$$\mathbf{G_k} = \left[\frac{\mathbf{b}_k^* \mathbf{b}_k^T}{\sigma^2} + \mathbf{C}^{-1} + |\alpha|^2 \left(\mathbf{C}^{-1} - \mathbf{C}^{-1}\hat{\mathbf{G}}_{k-1}\mathbf{C}^{-1} - \mathbf{C}^{-1}\hat{\mathbf{G}}_{k+1}\mathbf{C}^{-1}\right)\right]^{-1}$$

$$\mathbf{q_k} = \frac{r_k \mathbf{b}_k^*}{\sigma^2} + (1-\alpha^*)\mathbf{C}^{-1}\mathbf{d} + \alpha \mathbf{C}^{-1}\hat{\mathbf{G}}_{k-1}\hat{\mathbf{q}}_{k-1} + \alpha^* \mathbf{C}^{-1}\hat{\mathbf{G}}_{k+1}\hat{\mathbf{q}}_{k+1} \qquad (18)$$

$$\mathbf{g_k} = \mathbf{G_k}\mathbf{q_k}$$

$\mathbf{G}_k$ and $\mathbf{g}_k$ from (18) will be used in the Viterbi algorithm incremental cost function for time k. Update for next iteration:

$$\hat{\mathbf{G}}_k = \left[\frac{\mathbf{b}_k^* \mathbf{b}_k^T}{\sigma^2} + \mathbf{C}^{-1} + |\alpha|^2 \left(\mathbf{C}^{-1} - \mathbf{C}^{-1}\hat{\mathbf{G}}_{k-1}\mathbf{C}^{-1}\right)\right]^{-1}$$

$$\hat{\mathbf{q}}_k = \frac{r_k \mathbf{b}_k^*}{\sigma^2} + (1-\alpha^*)\mathbf{C}^{-1}\mathbf{d} + \alpha \mathbf{C}^{-1}\hat{\mathbf{G}}_{k-1}\hat{\mathbf{q}}_{k-1}, \qquad (19)$$

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

It is apparent from the above algorithm that $f(\mathbf{h_k}|\mathbf{r}, \mathbf{B})$ at any time k depends on all of the received data and estimated symbols from time 1 to N. In the EM algorithm E-step, (17-19) are 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 (14) is used to estimate $\mathbf{A}$.

V. Simulation Results

For each trial in the simulations, 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, 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. For a reference BER, the Viterbi algorithm with known channel coefficients was used.

For comparison purposes, the EM algorithm from [6] 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. The Viterbi algorithm is then used for MLSE. Also for comparison, the algorithm referred to as "deterministic" is that from [1], [2],

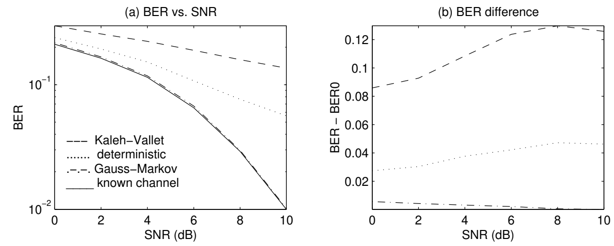

Fig. 1. BER vs. SNR for deterministic, Gauss-Markov, and Kaleh–Vallet EM algorithms using $N = 10$, $L = 3$, $\alpha = 0.8$: (a) absolute BER; (b) BER difference from known channel.

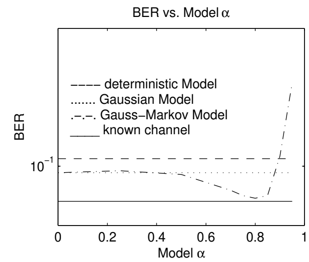

Fig. 2. BER vs. model α for Gaussian, Gauss-Markov, and Known Channel using $N = 10$, $L = 3$, true $\alpha = 0.8$:

[18] and is equivalent to optimal joint sequence and channel estimation . Our EM algorithm for the Gauss-Markov distribution is referred to as "Gauss-Markov", and produces optimal estimates of the symbols by marginalizing over the channel coefficient distribution.

The deterministic and Gauss-Markov 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 was initialized using the expected value (theoretical average) of the channel coefficients.

For Gauss-Markov channel coefficients, Figure 1 shows results from a typical simulation, and demonstrates that the Gauss-Markov EM algorithm performs almost as well as the known channel case, and much better than the deterministic and Kaleh–Vallet algorithms. In this simulation, the channel coefficients were generated randomly using (15) with $\mathbf{v}_k$ having mean $\mathbf{d} = [0.2\ 0.2\ 0.2]^T$ and covariance $\mathbf{C} = \begin{bmatrix} 0.1 & -0.1 & 0 \\ -0.1 & 0.2 & -0.1 \\ 0 & -0.1 & 0.2 \end{bmatrix}$.

The second simulation illustrates the effect of using an incorrect prior distribution for the channel coefficients. Using the same parameters as above to generate the data, with SNR fixed at 6dB, the Gauss-Markov model used the correct mean and covariance, while α was varied from 0 to 0.95. Figure 2 shows BER as a function of model α. Also shown are BER for the time-independent Gaussian [15], deterministic, and known channel estimators (which are not a function of model α). For model α too small or large, performance is degraded. This should be expected since the marginalization over $\mathbf{H}$ is then based on a prior that does not represent the actual channel coefficients. However, for a broad range of α, results are better then for the deterministic model. For model $\alpha = 0$, the Gauss-Markov model is equivalent to the time-independent Gaussian model.

VI. 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. In particular, we consider a Gauss-Markov model for random channels which can have significant variation over a short temporal block. Simulation results illustrate the advantage of this approach, relative to joint sequence/channel estimation.

Appendix

From Section IV we have $\mathbf{h}_k = \alpha \mathbf{h}_{k-1} + \mathbf{v}_k$, where $\mathbf{v}_k$ is Gaussian with mean $\mathbf{d}$ and covariance $\mathbf{C}$. Therefore $f(\mathbf{h}_k|\mathbf{h}_{k-1})$ is Gaussian with mean $\mathbf{d} + \alpha\mathbf{h}_{k-1}$ and covariance $\mathbf{C}$. To derive the formulas for $\mathbf{g}_k$ and $\mathbf{G}_k$ in this case, we start by examining the forms of $f(\mathbf{h}_k)$ and $f(\mathbf{h}_{k-1}|\mathbf{h}_k)$.

Recursively writing $\mathbf{h}_k$ in terms of $\mathbf{h}_{k-1}$, $\mathbf{h}_{k-2}$, ..., we obtain:

$$\mathbf{h}_k = \sum_{j=0}^{\infty} \alpha^j \mathbf{v}_{k-j}. \tag{20}$$

Assuming that $|\alpha| < 1$, the result of the summation in (20) is that $f(\mathbf{h}_k)$ is Gaussian with mean $\frac{\mathbf{d}}{1-\alpha}$ and covariance $\frac{\mathbf{C}}{1-|\alpha|^2}$.

Using Bayes rule, $f(\mathbf{h}_{k-1}|\mathbf{h}_k) = \frac{f(\mathbf{h}_k|\mathbf{h}_{k-1}) f(\mathbf{h}_{k-1})}{f(\mathbf{h}_k)} \doteq e^{-E_{k-1}}$, with:

$$\begin{aligned} E_{k-1} &= (\mathbf{h}_k - \mathbf{d} - \alpha\mathbf{h}_{k-1})^{*T}\, \mathbf{C}^{-1}\, (\mathbf{h}_k - \mathbf{d} - \alpha\mathbf{h}_{k-1}) \\ &+ \left(\mathbf{h}_{k-1} - \frac{\mathbf{d}}{1-\alpha}\right)^{*T} (1-|\alpha|^2)\mathbf{C}^{-1} \left(\mathbf{h}_{k-1} - \frac{\mathbf{d}}{1-\alpha}\right) \\ &- \left(\mathbf{h}_k - \frac{\mathbf{d}}{1-\alpha}\right)^{*T} (1-|\alpha|^2)\mathbf{C}^{-1} \left(\mathbf{h}_k - \frac{\mathbf{d}}{1-\alpha}\right). \end{aligned}$$

Algebraic manipulation yields:

$$E_{k-1} = (\mathbf{h}_{k-1} - \beta\mathbf{d} - \alpha^*\mathbf{h}_k)^{*T}\, \mathbf{C}^{-1}\, (\mathbf{h}_{k-1} - \beta\mathbf{d} - \alpha^*\mathbf{h}_k),$$

with $\beta \triangleq \frac{1-\alpha^*}{1-\alpha}$. So $f(\mathbf{h}_{k-1}|\mathbf{h}_k)$ is Gaussian with mean $\beta\mathbf{d} + \alpha^*\mathbf{h}_k$ and covariance $\mathbf{C}$.

Note that $f(\mathbf{H})$ can be written in terms of $f(\mathbf{h}_k)$ for any value of k. (16) shows this for $k = 1$. In general:

$$\begin{aligned} f(\mathbf{H}) &= f(\mathbf{h}_N|\mathbf{h}_{N-1}) \cdots f(\mathbf{h}_{k+1}|\mathbf{h}_k) f(\mathbf{h}_k) f(\mathbf{h}_{k-1}|\mathbf{h}_k) \\ &\cdot f(\mathbf{h}_{k-2}|\mathbf{h}_{k-1}) \cdots f(\mathbf{h}_1|\mathbf{h}_2) \end{aligned} \tag{21}$$

To evaluate $f(\mathbf{h}_k|\mathbf{r}, \mathbf{B})$, we can use the above form of $f(\mathbf{H})$ in (16) and integrate over $\mathbf{h}_i, i = 1, \ldots, N; i \neq k$.

Consider two of these integrations, say the integrals over $\mathbf{h}_1$ and $\mathbf{h}_2$. By examining the form of this result, we will be able to produce the general result. Showing just the terms involving $\mathbf{h}_1$, $I_1 = \int_{\mathbf{h}_1} e^{-E_1} d\mathbf{h}_1$, with:

$$\begin{aligned} E_1 &= (\mathbf{h}_1 - \beta\mathbf{d} - \alpha^*\mathbf{h}_2)^{*T}\mathbf{C}^{-1}(\mathbf{h}_1 - \beta\mathbf{d} - \alpha^*\mathbf{h}_2) \\ &+ \frac{|r_1 - \mathbf{b}_1^T\mathbf{h}_1|^2}{\sigma^2}. \end{aligned}$$

E_1 can be reduced, dropping constant terms, to:

$$\begin{aligned} E_1 &\doteq (\mathbf{h}_1 - \hat{\mathbf{g}}_1)^{*T}\hat{\mathbf{G}}_1^{-1}(\mathbf{h}_1 - \hat{\mathbf{g}}_1) \\ &- \hat{\mathbf{g}}_1^{*T}\hat{\mathbf{G}}_1^{-1}\hat{\mathbf{g}}_1 + |\alpha|^2\mathbf{h}_2^{*T}\mathbf{C}^{-1}\mathbf{h}_2 \\ &+ \alpha\beta\mathbf{h}_2^{*T}\mathbf{C}^{-1}\mathbf{d} + \alpha^*\beta^*\mathbf{d}^{*T}\mathbf{C}^{-1}\mathbf{h}_2, \end{aligned}$$

where $\hat{\mathbf{G}}_1$ and $\hat{\mathbf{g}}_1$ will be defined below. Since the last four terms in E_1 are not functions of $\mathbf{h}_1$, they can be factored out of the integral. And the integral over $\mathbf{h}_1$ involving the first term from E_1 produces a constant which can be ignored. Thus, the result is:

$$\begin{aligned} I_1 &\doteq e^{\hat{\mathbf{g}}_1^{*T}\hat{\mathbf{G}}_1^{-1}\hat{\mathbf{g}}_1 - |\alpha|^2\mathbf{h}_2^{*T}\mathbf{C}^{-1}\mathbf{h}_2} \\ &\cdot e^{-\alpha\beta\mathbf{h}_2^{*T}\mathbf{C}^{-1}\mathbf{d} - \alpha^*\beta^*\mathbf{d}^{*T}\mathbf{C}^{-1}\mathbf{h}_2}, \end{aligned} \tag{22}$$

with:

$$\begin{aligned} \hat{\mathbf{G}}_1 &= \left[\frac{\mathbf{b}_1^*\mathbf{b}_1^T}{\sigma^2} + \mathbf{C}^{-1}\right]^{-1} \\ \hat{\mathbf{q}}_1 &= \frac{r_1\mathbf{b}_1^*}{\sigma^2} + \beta\mathbf{C}^{-1}\mathbf{d} \\ \hat{\mathbf{g}}_1 &= \hat{\mathbf{G}}_1(\hat{\mathbf{q}}_1 + \alpha^*\mathbf{C}^{-1}\mathbf{h}_2) \end{aligned} \tag{23}$$

Expanding the first term from the exponent in (22), and dropping constant terms:

$$\begin{aligned} \hat{\mathbf{g}}_1^{*T}\hat{\mathbf{G}}_1^{-1}\hat{\mathbf{g}}_1 &= \left(\hat{\mathbf{q}}_1^{*T} + \alpha\mathbf{h}_2^{*T}\mathbf{C}^{-1}\right)\hat{\mathbf{G}}_1\left(\hat{\mathbf{q}}_1 + \alpha^*\mathbf{C}^{-1}\mathbf{h}_2\right) \\ &\doteq |\alpha|^2\mathbf{h}_2^{*T}\mathbf{C}^{-1}\hat{\mathbf{G}}_1\mathbf{C}^{-1}\mathbf{h}_2 \\ &+ \alpha^*\hat{\mathbf{q}}_1^{*T}\hat{\mathbf{G}}_1\mathbf{C}^{-1}\mathbf{h}_2 + \alpha\mathbf{h}_2^{*T}\mathbf{C}^{-1}\hat{\mathbf{G}}_1\hat{\mathbf{q}}_1. \end{aligned}$$

So we now have:

$$\begin{aligned} I_1 &\doteq e^{|\alpha|^2\mathbf{h}_2^{*T}\mathbf{C}^{-1}\hat{\mathbf{G}}_1\mathbf{C}^{-1}\mathbf{h}_2 + \alpha^*\hat{\mathbf{q}}_1^{*T}\hat{\mathbf{G}}_1\mathbf{C}^{-1}\mathbf{h}_2} \\ &\cdot e^{\alpha\mathbf{h}_2^{*T}\mathbf{C}^{-1}\hat{\mathbf{G}}_1\hat{\mathbf{q}}_1 - |\alpha|^2\mathbf{h}_2^{*T}\mathbf{C}^{-1}\mathbf{h}_2} \\ &\cdot e^{-\alpha\beta\mathbf{h}_2^{*T}\mathbf{C}^{-1}\mathbf{d} - \alpha^*\beta^*\mathbf{d}^{*T}\mathbf{C}^{-1}\mathbf{h}_2} \end{aligned}$$

This result shows how the integral over $\mathbf{h}_1$ produces terms involving $\mathbf{h}_2$ which must be included when performing the integral over $\mathbf{h}_2$. Showing just the terms involving $\mathbf{h}_2$ for this next integral, $I_2 = \int_{\mathbf{h}_2} e^{-E_2} d\mathbf{h}_2$, with:

$$\begin{aligned} E_2 &= (\mathbf{h}_2 - \beta\mathbf{d} - \alpha^*\mathbf{h}_3)^{*T}\mathbf{C}^{-1}(\mathbf{h}_2 - \beta\mathbf{d} - \alpha^*\mathbf{h}_3) \\ &+ \frac{|r_2 - \mathbf{b}_2^T\mathbf{h}_2|^2}{\sigma^2} - |\alpha|^2\mathbf{h}_2^{*T}\mathbf{C}^{-1}\hat{\mathbf{G}}_1\mathbf{C}^{-1}\mathbf{h}_2 \\ &- \alpha^*\hat{\mathbf{q}}_1^{*T}\hat{\mathbf{G}}_1\mathbf{C}^{-1}\mathbf{h}_2 - \alpha\mathbf{h}_2^{*T}\mathbf{C}^{-1}\hat{\mathbf{G}}_1\hat{\mathbf{q}}_1 \\ &+ |\alpha|^2\mathbf{h}_2^{*T}\mathbf{C}^{-1}\mathbf{h}_2 + \alpha\beta\mathbf{h}_2^{*T}\mathbf{C}^{-1}\mathbf{d} \\ &+ \alpha^*\beta^*\mathbf{d}^{*T}\mathbf{C}^{-1}\mathbf{h}_2 \end{aligned}$$

E_2 can be reduced, dropping constant terms, to:

$$\begin{aligned} E_2 &\doteq (\mathbf{h}_2-\hat{\mathbf{g}}_2)^{*T}\hat{\mathbf{G}}_2^{-1}(\mathbf{h}_2-\hat{\mathbf{g}}_2)-\hat{\mathbf{g}}_2^{*T}\hat{\mathbf{G}}_2^{-1}\hat{\mathbf{g}}_2 \\ &\quad +|\alpha|^2\mathbf{h}_3^{*T}\mathbf{C}^{-1}\mathbf{h}_3+\alpha\beta\mathbf{h}_3^{*T}\mathbf{C}^{-1}\mathbf{d} \\ &\quad +\alpha^*\beta^*\mathbf{d}^{*T}\mathbf{C}^{-1}\mathbf{h}_3, \end{aligned} \tag{24}$$

where $\hat{\mathbf{G}}_2$ and $\hat{\mathbf{g}}_2$ will be defined below. Since the last four terms in (24) are not functions of $\mathbf{h}_2$, they can be factored out of the integral. And the integral over $\mathbf{h}_2$ involving the first term produces a constant which can be ignored. Thus, the result is:

$$\begin{aligned} I_2 &\doteq e^{\hat{\mathbf{g}}_2^{*T}\hat{\mathbf{G}}_2^{-1}\hat{\mathbf{g}}_2-|\alpha|^2\mathbf{h}_3^{*T}\mathbf{C}^{-1}\mathbf{h}_3} \\ &\quad \cdot e^{-\alpha\beta\mathbf{h}_3^{*T}\mathbf{C}^{-1}\mathbf{d}-\alpha^*\beta^*\mathbf{d}^{*T}\mathbf{C}^{-1}\mathbf{h}_3} \end{aligned} \tag{25}$$

with:

$$\begin{aligned} \hat{\mathbf{G}}_2 &= \left[\frac{\mathbf{b}_2^*\mathbf{b}_2^T}{\sigma^2}+\mathbf{C}^{-1}+|\alpha|^2\left(\mathbf{C}^{-1}-\mathbf{C}^{-1}\hat{\mathbf{G}}_1\mathbf{C}^{-1}\right)\right]^{-1} \\ \hat{\mathbf{q}}_2 &= \frac{r_2\mathbf{b}_2^*}{\sigma^2}+(1-\alpha^*)\mathbf{C}^{-1}\mathbf{d}+\alpha\mathbf{C}^{-1}\hat{\mathbf{G}}_1\hat{\mathbf{q}}_1 \qquad (26) \\ \hat{\mathbf{g}}_2 &= \hat{\mathbf{G}}_2(\hat{\mathbf{q}}_2+\alpha^*\mathbf{C}^{-1}\mathbf{h}_3) \end{aligned}$$

From the forms of (22) and (25) we can easily deduce the general form of the result of integrating over $\mathbf{h}_i$.

(26) indicates the general form of the equations for $\hat{\mathbf{G}}_i$ and $\hat{\mathbf{g}}_i$. These equations have more terms than (23) because in deriving (23) there were no previous integration results to incorporate. However, if we define $\hat{\mathbf{G}}_0=\mathbf{C}$ and $\hat{\mathbf{q}}_0=\beta\mathbf{C}^{-1}\mathbf{d}$, then (23) may also be written in the same form as (26).

The preceding derivations used (21) to integrate (16) over $\mathbf{h}_1$ and $\mathbf{h}_2$, and show the recursive forms of the update equations which appear in (19). If we start at the "other end" of (21), and integrate over $\mathbf{h}_N$ first, then $\mathbf{h}_{N-1}$, etc., we obtain similar recursive equations which appear in (17).

Finally, to produce $f(\mathbf{h}_k|\mathbf{r},\mathbf{B})$ after integrating over $\mathbf{h}_i, i=1,\ldots,N, i\neq k$, we have $f(\mathbf{h}_k|\mathbf{r},\mathbf{B})\doteq e^{-E_k}$, with:

$$\begin{aligned} E_k &= \left(\mathbf{h}_k-\frac{\mathbf{d}}{1-\alpha}\right)^{*T}\left(1-|\alpha|^2\right)\mathbf{C}^{-1}\left(\mathbf{h}_k-\frac{\mathbf{d}}{1-\alpha}\right) \\ &\quad +\frac{|r_k-\mathbf{b}_k^T\mathbf{h}_k|^2}{\sigma^2} \\ &\quad +|\alpha|^2\mathbf{h}_k^{*T}\mathbf{C}^{-1}\mathbf{h}_k+\alpha^*\mathbf{h}_k^{*T}\mathbf{C}^{-1}\mathbf{d}+\alpha\mathbf{d}^{*T}\mathbf{C}^{-1}\mathbf{h}_k \\ &\quad -\alpha\mathbf{q}_{k+1}^{*T}\hat{\mathbf{G}}_{k+1}\mathbf{C}^{-1}\mathbf{h}_k-\alpha^*\mathbf{h}_k^{*T}\mathbf{C}^{-1}\hat{\mathbf{G}}_{k+1}\mathbf{q}_{k+1} \\ &\quad -|\alpha|^2\mathbf{h}_k^{*T}\mathbf{C}^{-1}\hat{\mathbf{G}}_{k+1}\mathbf{C}^{-1}\mathbf{h}_k \\ &\quad +|\alpha|^2\mathbf{h}_k^{*T}\mathbf{C}^{-1}\mathbf{h}_k+\alpha\beta\mathbf{h}_k^{*T}\mathbf{C}^{-1}\mathbf{d} \\ &\quad +\alpha^*\beta^*\mathbf{d}^{*T}\mathbf{C}^{-1}\mathbf{h}_k-\alpha^*\mathbf{q}_{k-1}^{*T}\hat{\mathbf{G}}_{k-1}\mathbf{C}^{-1}\mathbf{h}_k \\ &\quad -\alpha\mathbf{h}_k^{*T}\mathbf{C}^{-1}\hat{\mathbf{G}}_{k-1}\mathbf{q}_{k-1} \\ &\quad -|\alpha|^2\mathbf{h}_k^{*T}\mathbf{C}^{-1}\hat{\mathbf{G}}_{k-1}\mathbf{C}^{-1}\mathbf{h}_k \end{aligned}$$

where the third through fifth lines above show the cumulative contributions of the integrals over $\mathbf{h}_i$ for $i>k$, and the last four lines show the contributions from $i<k$. Combining terms and dropping constants, this can be written as:

$$E_k \doteq (\mathbf{h}_k-\mathbf{g}_k)^{*T}\mathbf{G}_k(\mathbf{h}_k-\mathbf{g}_k)$$

indicating that $f(\mathbf{h}_k|\mathbf{r},\mathbf{B})$ has a Gaussian distribution, with $\mathbf{G}_k$ and $\mathbf{g}_k$ defined in (18).

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, 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] C. Cozzo and B. Hughes, "An iterative receiver for space-time communications," *CISS'2000, Princeton, NJ*, pp. WA8a–7 – WA8a–11, March 2000.

[13] H. Zamiri-Jafarian and S. Pasupathy, "Adaptive MLSDE using the EM algorithm," *IEEE Trans. on Communications*, vol. 47, pp. 1181 – 1193, August 1999.

[14] R. Perry and K. Buckley, "Sequence estimation over linearly–constrained random channels," *CISS'2000, Princeton, NJ*, pp. FA7–7 – FA7–12, March 2000.

[15] R. Perry, W. A. Berger, and K. Buckley, "EM algorithms for sequence estimation over random ISI channels," *Proc. 33-rd Asilomar Conference*, Oct 1999.

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

[17] 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.

[18] K. Chang and C. N. Georghiades, "Iterative joint sequence and channel estimation for fast time-varying intersymbol interference channels," *ICC'95*, June 1995.