Wireless Networks 8, 649–658, 2002

A Simplified EM Algorithm for Detection of CPM Signals in a Fading Multipath Channel

LINDA M. ZEGER *
Lucent Technologies, Room 14C-327, 67 Whippany Road, Whippany, NJ 07981, USA

HISASHI KOBAYASHI
Department of Electrical Engineering, Princeton University, Princeton, NJ 08540, USA

Abstract. Application of the EM (Expectation–Maximization) algorithm to sequence estimation in an unknown channel can in principle produce MLSE (maximum likelihood sequence estimates) that are not dependent on a particular channel estimate. The Expectation step of this algorithm cannot be directly performed for continuous phase modulated (CPM) signals transmitted in a time varying multipath channel. We therefore derive a simplification of the EM algorithm for CPM signals in this channel. Simulations applied to the Global System for Mobile Communications (GSM) show that the simplified EM algorithm significantly decreases the amount of training data needed for the channel model considered, and removes the majority of the bit errors that are due to imperfect knowledge of the channel.

Keywords: EM algorithm, sequence estimation, mobile communication, multipath channels

1. Introduction

There has recently been much interest in improving equalization techniques in mobile radio. A number of recent works including [1,17] study equalizers for CPM signals that are subject to multipath. Many current receivers compute an estimate $\widehat{\beta}$ of the fading multipath channel, denoted here by β, and then produce the estimate of the transmitted data sequence $\mathbf{C}$ that maximizes the likelihood function $f(\mathbf{y} \mid \mathbf{C}, \widehat{\beta})$, where $\mathbf{y}$ denotes samples of the received signal. This sequence estimate then depends on the channel estimate $\widehat{\beta}$, which is often determined from training data. Recent works have explored variants of this receiver. For example, in [6] the complexity of the VA (Viterbi Algorithm) in the maximization of $f(\mathbf{y} \mid \mathbf{C}, \widehat{\beta})$ is reduced, while in [3] blind channel estimation is investigated. Both [6] and [3] require slightly more power to achieve the same BER than conventional equalizers. Because all of these receivers maximize $f(\mathbf{y} \mid \mathbf{C}, \widehat{\beta})$ with respect to $\mathbf{C}$, they incur bit errors due to the difference between the estimate $\widehat{\beta}$ and the true channel β. Such bit errors could be reduced by a receiver that instead maximizes $f(\mathbf{y} \mid \mathbf{C})$, which is the desired receiver since the channel β is not of direct interest to the mobile user. We propose and implement an algorithm which can perform this maximization for CPM signals by simultaneously estimating the sequence $\mathbf{C}$ and handling the time varying multipath channel. Our algorithm thus improves performance of reception of such signals relative to current conventional equalizers.

Combined data detection and channel estimation for signals that are transmitted with modulations other than CPM have been recently studied in works such as [8,10]. In these studies another likelihood function $f(\mathbf{y} \mid \mathbf{C}, \beta)$ is maximized jointly with respect to $\mathbf{C}$ and β, by alternating between estimation of $\mathbf{C}$ and β. The EM (Expectation–Maximization) algorithm is used for the channel estimation portions of these methods.

In [4] combined detection and decoding of CPM signals is done by using the EM algorithm for channel estimation only; this EM algorithm is then embedded in the iterations of another algorithm that performs detection and decoding. In contrast, in our work, the entire process of handling the channel and detecting the CPM signal is accomplished within a single EM algorithm.

The EM algorithm [9,15] is an iterative method for maximizing a likelihood function in the presence of unobserved data. This algorithm has been traditionally applied to parameter estimation in many works such as [18]. The EM algorithm has also been applied to sequence estimation [11,12,16], where either interfering users' bits or unknown parameters, such as symbol timing, are treated as the unobserved data.

The EM algorithm has recently been applied to perform sequence estimation when the channel is unknown [11]. In this case, it averages over possible realizations of the channel to produce the most likely transmitted sequence $\mathbf{C}$, given only the received signal samples $\mathbf{y}$. The desired likelihood function $f(\mathbf{y} \mid \mathbf{C})$ is maximized with respect to $\mathbf{C}$. This study shows examples in which the channel is either a random phase or a random amplitude fading channel; the average over these channel models was performed analytically.

In this paper, we apply the EM algorithm to estimate data sequences transmitted instead by continuous phase modulation (CPM) in a multipath channel, in which each path contains a random phase, amplitude, and time delay. In this case, the EM algorithm's averaging over the unknown channel cannot be performed analytically, and is computationally intractable. Therefore, we derive a simplified version of the

* Corresponding author.

EM algorithm for this problem. We show that this simplified EM algorithm is equivalent to the original EM algorithm and it maximizes $f(\mathbf{y} \mid \mathbf{C})$, when an adequate amount of training data is used. Maximization of this desired likelihood function, rather than other commonly used likelihood functions, produces performance improvement; we demonstrate that our algorithm applied to GSM can significantly reduce the number of training bits needed relative to current methods.

In section 2 the modulation and channel models are specified. The EM algorithm is applied to sequence estimation for these models in section 3. We present a simplified version of this EM algorithm in section 4, and prove its equivalence to the original EM algorithm in appendix A. Section 5 discusses initialization of the algorithm and use of training data. Simulation results are presented in section 6, and conclusions are drawn in section 7.

2. The model

The transmitted data sequence is denoted by C_n for $n = 1, 2, 3, \ldots$, where $C_n \in \{-1, 1\}$, and the sequence is denoted collectively by $\mathbf{C}$. The transmitted signal uses some form of CPM, and is thus given by

$$X(t) = \cos\left(\omega_c t + \sum_n C_n q(t - nT)\right), \tag{1}$$

where T denotes the bit period, and ω_c is the carrier frequency. The continuous function $q(t)$ can be represented as the integral of a baseband pulse

$$q(t) = \frac{\pi h_f}{T} \int_{-\infty}^{t} g(t')\, dt'. \tag{2}$$

For example, in Gaussian minimum shift keying (GMSK), which is used in GSM and GPRS (General Packet Radio Service), the baseband frequency pulse $g(t)$ spreads each transmitted symbol over several symbol periods, and the modulation index is $h_f = 0.5$.

We consider a general multipath model with M paths. The received signal is then

$$W(t) = \sum_{i=1}^{M} \alpha_i(t) X\big(t - \tau_i(t), \theta_i(t)\big), \tag{3}$$

where the phase shift, amplitude, and delay of the ith path are denoted respectively by $\theta_i(t)$, $\alpha_i(t)$, and $\tau_i(t)$. White Gaussian noise is added to the multipath fading model (3), and the resulting received signal can then be represented as

$$V(t) = \Re\big\{Y(t)\exp(j\omega_c t)\big\}, \tag{4}$$

where the complex envelope $Y(t)$ of the received signal is

$$Y(t) = \sum_{i=1}^{M} \alpha_i(t) \exp\big[j\big(\Phi_i(t, \mathbf{C})\big)\big] + n_I(t) + jn_Q(t). \tag{5}$$

The phase in (5) is

$$\Phi_i(t, \mathbf{C}) = \theta_i(t) - \omega_c \tau_i(t) + \sum_n C_n q\big(t - \tau_i(t) - nT\big), \tag{6}$$

and the inphase and quadrature noise components n_I and n_Q are independent white Gaussian noise processes with double-sided power spectral density N_o. The inphase and quadrature samples of the received signal $Y(t)$ are denoted collectively by $\mathbf{y}$. In our simulations of GSM the sampling rate is $2/T$, and an ideal filter of bandwidth $1/T$ is assumed at the receiver.

We denote the channel parameters collectively by β:

$$\beta = \beta(t) = \big\{\alpha_1(t), \theta_1(t), \tau_1(t), \ldots, \alpha_M(t), \theta_M(t), \tau_M(t)\big\}. \tag{7}$$

We note that $\mathbf{C}$ is divided into subsequences of symbols such that the channel varies little over the length of a subsequence. Assuming a constant channel during each subsequence, we use the EM algorithm separately within each subsequence to compute the MLSE of that subsequence. Hereafter, $\mathbf{y}$, $\mathbf{C}$, and β refer to a single such subsequence, which for GSM corresponds to a time slot.

3. The EM algorithm for sequence estimation

The quantity to be estimated here is the sequence $\mathbf{C}$. The "missing data" of the EM algorithm is the unknown channel β, while the "complete data" are $\{\mathbf{y}, \beta\}$. Thus, an expectation over the missing data amounts to taking an average over the unknown channel [11]. We first summarize application of the EM algorithm to sequence estimation in the presence of a general unknown channel β, as was described in [11]. We next focus on application to the modulation and channel models specified in section 2.

The EM algorithm here consists of repeating two steps. The E (Expectation) step computes the expected value of the "complete data" log likelihood $\ln f(\mathbf{y}, \beta \mid \mathbf{C})$ as a function of the sequence estimate $\mathbf{C}$:

$$Q\big(\mathbf{C} \mid \widehat{\mathbf{C}}^p\big) = E\big[\ln f(\mathbf{y}, \beta \mid \mathbf{C}) \mid \mathbf{y}, \widehat{\mathbf{C}}^p\big]. \tag{8}$$

The $(p+1)$st E step is computed in (8) by using the density $f(\beta \mid \mathbf{y}, \widehat{\mathbf{C}}^p)$ to take the expectation over β, where the sequence estimate $\widehat{\mathbf{C}}^p$ of $\mathbf{C}$ is obtained from the previous (pth) M step. In the $(p+1)$st M (Maximization) step, the transmitted sequence $\widehat{\mathbf{C}}^{p+1}$ that maximizes $Q(\mathbf{C} \mid \widehat{\mathbf{C}}^p)$ is found. This sequence estimate will then be used in the E step in the $(p+2)$nd iteration in $f(\beta \mid \mathbf{y}, \widehat{\mathbf{C}}^{p+1})$. The E and M steps are repeated until convergence is achieved, which for the problem here of sequence estimation occurs when $\widehat{\mathbf{C}}^{p+1} = \widehat{\mathbf{C}}^p$.

Equation (8) can be simplified, as was done in [11] by omitting normalization constants and terms independent of $\mathbf{C}$ hereafter, which will drop out in the M step:

$$\begin{aligned} Q\big(\mathbf{C} \mid \widehat{\mathbf{C}}^p\big) &= E\big[\ln f(\mathbf{y} \mid \beta, \mathbf{C}) \mid \mathbf{y}, \widehat{\mathbf{C}}^p\big] \\ &= \int \ln f(\mathbf{y} \mid \beta, \mathbf{C}) f\big(\beta \mid \mathbf{y}, \widehat{\mathbf{C}}^p\big)\, d\beta, \end{aligned} \tag{9}$$

where the posterior density of β can be expressed as, omitting the normalization constant,

$$f\big(\beta \,\big|\, \mathbf{y}, \widehat{\mathbf{C}}^p\big) \doteq f\big(\mathbf{y} \,\big|\, \beta, \widehat{\mathbf{C}}^p\big)\rho(\beta), \tag{10}$$

and $\rho(\beta)$ is the prior density of the channel parameters.

The EM algorithm is now applied to CPM and the time varying multipath channel model described in section 2. The density $f(\mathbf{y} \mid \beta, \widehat{\mathbf{C}}^p)$ can be obtained from (5) and is

$$\begin{aligned} f(\mathbf{y} \mid \beta, \mathbf{C}) &= \exp\left\{-\frac{1}{2N_0}\sum_{l=1}^{K}\left\{\left[y_{\mathrm{I}}(t_l) - \sum_{i=1}^{M}\alpha_i(t_l)\cos\big(\Phi_i(t_l, \mathbf{C})\big)\right]^2\right.\right. \\ &\quad \left.\left. + \left[y_{\mathrm{Q}}(t_l) - \sum_{i=1}^{M}\alpha_i(t_l)\sin\big(\Phi_i(t_l, \mathbf{C})\big)\right]^2\right\}\right\}, \end{aligned} \tag{11}$$

where K is the number of samples of the received signal, and the inphase and quadrature components of the received signal samples taken at time t_l are denoted by $y_{\mathrm{I}}(t_l)$ and $y_{\mathrm{Q}}(t_l)$, respectively. Omitting the constant factor $1/(2N_0)$, we denote the negative logarithm of (11) by the following "distance" function:

$$\begin{aligned} \lambda(\beta, \mathbf{y}, \mathbf{C}) &= -\ln f(\mathbf{y} \mid \beta, \mathbf{C}) \\ &= \sum_{l=1}^{K}\left\{\left[y_{\mathrm{I}}(t_l) - \sum_{i=1}^{M}\alpha_i(t_l)\cos\big(\Phi_i(t_l, \mathbf{C})\big)\right]^2\right. \\ &\quad \left. + \left[y_{\mathrm{Q}}(t_l) - \sum_{i=1}^{M}\alpha_i(t_l)\sin\big(\Phi_i(t_l, \mathbf{C})\big)\right]^2\right\}. \end{aligned} \tag{12}$$

Equation (12) is substituted into (9) to yield the algorithm at the $(p+1)$st iteration:

$$\text{E step:}\quad Q\big(\mathbf{C} \,\big|\, \widehat{\mathbf{C}}^p\big) = -\int \lambda(\beta, \mathbf{y}, \mathbf{C}) f\big(\beta \,\big|\, \mathbf{y}, \widehat{\mathbf{C}}^p\big)\,\mathrm{d}\beta, \tag{13}$$

$$\text{M step:}\quad \widehat{\mathbf{C}}^{p+1} = \arg\max_{\mathbf{C}} Q\big(\mathbf{C} \,\big|\, \widehat{\mathbf{C}}^p\big). \tag{14}$$

Equation (10) is used in (13), where the density $f(\mathbf{y} \mid \beta, \widehat{\mathbf{C}}^p)$ can be obtained from equation (11), and construction of the prior density $\rho(\beta)$ is discussed in section 5. We will later refer to the steps (13)–(14) as the "complete" EM algorithm.

The EM algorithm could be directly implemented if the multiple integral in (13) could be performed analytically, as it can for the modulation and channel models considered in the examples in [11], which produce a likelihood function $f(\mathbf{y} \mid \beta, \mathbf{C})$ that is simpler than that of (11). However, calculation of $Q(\mathbf{C} \mid \widehat{\mathbf{C}}^p)$ with the likelihood function (11) would require numerical integration for every realization of the sequence $\mathbf{C}$, which would be computationally intractable.

4. Simplified EM algorithm

We present a simplification of the EM algorithm in section 4.1 to enable evaluation of the otherwise intractable step (13). We explicitly calculate the E step of the simplified algorithm in section 4.2.

4.1. Reduction of the EM algorithm

The idea behind simplification of the E step (13) is that most of the contribution to the integral comes from a limited range of β values near the peak of $f(\beta \mid \mathbf{y}, \widehat{\mathbf{C}}^p)$, and the symbols in the sequence $\mathbf{C}$ can take on only discrete values, so that the same sequence $\mathbf{C}$ will be optimal for all values of β in this limited region. We state the simplified EM algorithm below, and prove its equivalence to the complete EM algorithm in appendix A.

We denote

$$\widehat{\beta}^p = \widehat{\beta}^p\big(\mathbf{y}, \widehat{\mathbf{C}}^p\big) = \arg\max_{\beta} f\big(\beta \,\big|\, \mathbf{y}, \widehat{\mathbf{C}}^p\big). \tag{15}$$

We assume $\widehat{\beta}^p$ is unique; this assumption is removed at the end of appendix B. In section 4.2 we show how to compute $\widehat{\beta}^p$. The simplified EM algorithm at the $(p+1)$st iteration for every $p \geqslant 0$ is

$$\text{E step:}\quad Q^{\mathrm{simp}}\big(\mathbf{C} \,\big|\, \widehat{\mathbf{C}}^p\big) = -\lambda\big(\widehat{\beta}^p, \mathbf{y}, \mathbf{C}\big), \tag{16}$$

$$\text{M step:}\quad \widehat{\mathbf{C}}^{p+1} = \arg\max_{\mathbf{C}} Q^{\mathrm{simp}}\big(\mathbf{C} \,\big|\, \widehat{\mathbf{C}}^p\big), \tag{17}$$

where the M step can be performed with the VA. The E step (16) of the simplified EM algorithm replaces the expectation $Q(\mathbf{C} \mid \widehat{\mathbf{C}}^p)$ of the log likelihood in the complete algorithm (13) with the log likelihood $Q^{\mathrm{simp}}(\mathbf{C} \mid \widehat{\mathbf{C}}^p)$ evaluated at the single value of the channel parameters $\widehat{\beta}^p$. Equivalence of the simplified EM algorithm (16), (17) to the complete EM algorithm (13), (14) is, therefore, achieved if $\arg\max_{\mathbf{C}}[-\int \lambda(\beta, \mathbf{y}, \mathbf{C}) f(\beta \mid \mathbf{y}, \widehat{\mathbf{C}}^p)\,\mathrm{d}\beta] = \arg\max_{\mathbf{C}}[-\lambda(\widehat{\beta}^p, \mathbf{y}, \mathbf{C})]$. A sufficient condition for this equivalence is given by (A.4) of appendix A. In the remainder of this paper the sequence $\widehat{\mathbf{C}}^{p+1}$ is defined by (17).

The probability p_{notEM} that the simplified EM algorithm at iteration $p+1$ is not equivalent to the complete EM algorithm is

$$p_{\mathrm{notEM}} = \Pr\big(\arg\max_{\mathbf{C}} Q\big(\mathbf{C} \,\big|\, \widehat{\mathbf{C}}^p\big) \neq \widehat{\mathbf{C}}^{p+1}\big), \tag{18}$$

where $Q(\mathbf{C} \mid \widehat{\mathbf{C}}^p)$ is defined by (13). Generally, the probability p_{notEM} decreases as the SNR increases, since $f(\beta \mid \mathbf{y}, \widehat{\mathbf{C}}^p)$ is more peaked with a higher SNR, and thus approximation of the integral (13) by using only $\beta = \widehat{\beta}^p$ becomes more accurate. The effect of the amount of training data transmitted on the size of p_{notEM} is discussed in section 5.

The initial E step (16) of the simplified EM algorithm estimates the channel using training data, for example, as described in section 5. The information symbol estimates derived from this channel estimate in the M step (17) are then used in subsequent iterations to improve the channel estimate. The improved channel estimate in turn results in the elimination of bit errors due to the initial imperfect channel estimate.

Conventionally, the parameter which is being estimated by an EM algorithm takes on a continuum of values, in which case the algorithm has been shown [9] to converge to the correct local maximum of the likelihood function if the initial value of that parameter is close enough to this maximum. In contrast, we use the EM algorithm for estimation of the sequence $\mathbf{C}$, which consists of a series of symbols that take on discrete values. Each M step in essence through the VA considers every possible sequence $\mathbf{C}$; the issue of convergence to the global maximum is different here. In appendix B we prove convergence of our simplified EM algorithm to the global maximum of $f(\mathbf{y} \mid \mathbf{C})$. A sufficient criterion (B.7) for this convergence requires that the prior density be peaked; use of enough training data for the given SNR should enable this criterion to be satisfied, as discussed in section 5.

The actual steps of the simplified EM algorithm are similar to those of other algorithms such as joint sequence and parameter estimation [13], as well as [8,10]. However, the simplified EM algorithm differs from these other algorithms in two ways. First, at each iteration of the simplified EM algorithm channel estimates derived from prior information such as training data are combined with those derived from current estimates of information symbols, as we show explicitly in section 4.2. Secondly, as we show in appendix B, this combination makes the simplified EM algorithm produce the sequence estimate that maximizes the pertinent likelihood function $f(\mathbf{y} \mid \mathbf{C})$, which does not depend on the unknown channel, rather than other commonly used likelihood functions such as $f(\mathbf{y} \mid \mathbf{C}, \beta)$. As a result, the simplified EM algorithm removes most bit errors due to imperfect knowledge of the channel, as we demonstrate in section 6.

4.2. E step of simplified algorithm

Explicit calculation of the E step (16) requires calculation of $\widehat{\beta}^p$ through (15), and thus calculation of the posterior density $f(\beta \mid \mathbf{y}, \widehat{\mathbf{C}}^p)$. The posterior density (10) is the product of the prior density and $f(\mathbf{y} \mid \beta, \widehat{\mathbf{C}}^p)$; below we show how both of these densities can be treated as Gaussian.

We first consider the prior density. Any prior density $\rho(\beta)$ with tails that have sufficiently low probability such that (A.4) is satisfied with high probability, or less restrictively, that renders the probability (18) small, allows the simplification of the E step done in section 4.1. The exact form of the prior density used will have little effect on the final results of the algorithm; as long as the tails of $\rho(\beta)$ are small enough that $\rho(\beta)$ satisfies (B.7), the algorithm will converge to the global maximum of $f(\mathbf{y} \mid \mathbf{C})$, as shown in appendix B.

In this section we consider a Gaussian $\rho(\beta)$, in order to illustrate how information from training data is combined with information from tentative symbol estimates in the E step (16). We derive a Gaussian approximation to (10), which will be used by the receiver in section 6, where it is seen that this approximation in the receiver yields quite good performance, although it differs from the density of the simulated true channel. Thus, the simplified EM algorithm demonstrates robustness with respect to the particular channel prior density assumed by the receiver.

We now consider a Gaussian prior density $\rho(\beta)$:

$$\rho(\beta) = \exp\left[-\frac{(\beta - \overline{\beta}_\rho)^{\mathrm{t}}\mathbf{B}(\beta - \overline{\beta}_\rho)}{2}\right], \tag{19}$$

where the superscript t denotes transpose. Derivation of the mean $\overline{\beta}_\rho$ and inverse variance matrix $\mathbf{B}$ is discussed in section 5.

We next derive a Gaussian approximation to $f(\mathbf{y} \mid \beta, \widehat{\mathbf{C}}^p)$. We consider the estimate, denoted by $\widetilde{\beta}^p$, at iteration $p+1$ of the true β that maximizes the likelihood function $f(\mathbf{y} \mid \beta, \widehat{\mathbf{C}}^p)$:

$$\widetilde{\beta}^p = \widetilde{\beta}^p(\mathbf{y}, \widehat{\mathbf{C}}^p) = \arg\min_{\beta} \lambda(\beta, \mathbf{y}, \widehat{\mathbf{C}}^p), \tag{20}$$

where the negative log likelihood $\lambda(\beta, \mathbf{y}, \widehat{\mathbf{C}}^p)$ is obtained from (12). Common techniques to perform numerical minimization would require taking numerical derivatives with respect to β of (20), although methods that do not involve derivatives [2] are available. In order to compute these derivatives, we therefore require CPM so that the phase $\Phi_i(t_l, \widehat{\mathbf{C}}^p)$, as given by (6), is continuous with respect to the multipath parameter $\tau_i(t_l)$.

When β is close to $\widetilde{\beta}^p$, a Taylor series expansion of the exponent in (11) about $\widetilde{\beta}^p$ yields

$$f(\mathbf{y} \mid \beta, \widehat{\mathbf{C}}^p) \approx \exp\left[-\frac{(\beta - \widetilde{\beta}^p)^{\mathrm{t}}\mathbf{A}^p(\beta - \widetilde{\beta}^p)}{2}\right], \tag{21}$$

where the constant factor $\exp(-\lambda(\widetilde{\beta}^p, \mathbf{y}, \widehat{\mathbf{C}}^p)/(2N_o))$ has been omitted because it is irrelevant in the M step. The matrix elements of $\mathbf{A}^p$ are

$$\mathbf{A}^p_{ij} = \frac{1}{2N_o}\frac{\partial^2\lambda(\beta, \mathbf{y}, \widehat{\mathbf{C}}^p)}{\partial\beta_i\partial\beta_j}\bigg|_{\widetilde{\beta}^p}. \tag{22}$$

The posterior density (10) for the channel parameters is, therefore, the product of the Gaussian densities (19) and (21) for β near $\widetilde{\beta}^p$, and thus, in this region is also a Gaussian density that can be expressed as

$$f(\beta \mid \mathbf{y}, \widehat{\mathbf{C}}^p) \approx \exp\left[-\frac{(\beta - \widehat{\beta}^p)^{\mathrm{t}}\widehat{\mathbf{A}}^p(\beta - \widehat{\beta}^p)}{2}\right], \tag{23}$$

where the estimated posterior mean $\widehat{\beta}^p$ and the inverse posterior variance matrix $\widehat{\mathbf{A}}^p$ at iteration $p+1$ are

$$\widehat{\beta}^p = (\widehat{\mathbf{A}}^p)^{-1}(\mathbf{A}^p\widetilde{\beta}^p + \mathbf{B}\overline{\beta}_\rho), \tag{24}$$

$$\widehat{\mathbf{A}}^p = \mathbf{A}^p + \mathbf{B}. \tag{25}$$

Equation (24) is, thus, a weighted sum of the estimate $\overline{\beta}_\rho$ obtained from the prior density $\rho(\beta)$, which, for example, is derived from training data, and the estimate $\widetilde{\beta}^p$, which is derived through (20) from the estimate $\widehat{\mathbf{C}}^p$ at iteration p of the information sequence. At each iteration, the estimates $\widetilde{\beta}^p$ and $\mathbf{A}^p$ derived from $f(\mathbf{y} \mid \beta, \widehat{\mathbf{C}}^p)$ are updated, while the estimates $\overline{\beta}_\rho$ and $\mathbf{B}$ derived from $\rho(\beta)$ remain fixed. Thus, (24) and (25) can be used to calculate $\widehat{\beta}^p$ and thereby the E step (16) of the simplified EM algorithm.

5. Initialization

We start the algorithm with the E step by constructing the initial channel parameter density, denoted here by $f(\beta \mid \mathbf{y}, \widehat{\mathbf{C}}^0)$, although there is actually no previous sequence estimate $\widehat{\mathbf{C}}^0$. Therefore, we equate

$$f\big(\beta \,\big|\, \mathbf{y}, \widehat{\mathbf{C}}^0\big) = \rho(\beta). \tag{26}$$

The initial M step will then produce the first sequence estimate $\widehat{\mathbf{C}}^1$.

We now describe methods for constructing $\rho(\beta)$, which will be used in (26), as well as in (24) which is used in subsequent iterations of the E step (16). We consider a Gaussian prior density for the purpose of illustration. We assume the channel changes rapidly enough that it is not highly correlated between a given user's consecutive subsequences or time slots, so that we use training data in each subsequence to derive the parameters of $\rho(\beta)$. Given a training data sequence $\mathbf{C}_T$ of symbols in a time slot and the corresponding samples of the received signal denoted collectively by $\mathbf{y}_T$, we let the mean value $\overline{\beta}_\rho = \overline{\beta}_T$ of β, equal the maximum likelihood estimate of β:

$$\overline{\beta}_T = \overline{\beta}_T(\mathbf{y}_T, \mathbf{C}_T) = \arg\min_\beta \lambda(\beta, \mathbf{y}_T, \mathbf{C}_T), \tag{27}$$

where $\lambda(\beta, \mathbf{y}_T, \mathbf{C}_T)$ is given by (12). We choose the variance of β in the prior density by considering how likely β is to be close to the estimate $\overline{\beta}_T$, or equivalently, what is the variance of the estimate $\overline{\beta}_T$ for a given β. Hence, using the Cramer–Rao inequality [7] as a guide, we choose the variance of $\rho(\beta)$ to be greater than the inverse of the Fisher information matrix $\mathbf{J}(\overline{\beta}_T)$, which in this case has matrix elements

$$J_{ij}(\overline{\beta}_T) = E\left[\frac{\partial \ln f(\mathbf{y}_T \mid \overline{\beta}_\mathbf{T}, \mathbf{C}_T)}{\partial \overline{\beta}_{Ti}} \frac{\partial \ln f(\mathbf{y}_T \mid \overline{\beta}_\mathbf{T}, \mathbf{C}_T)}{\partial \overline{\beta}_{Tj}}\right], \tag{28}$$

where E denotes the expectation over the possible values of $\mathbf{y}_T$ for the given N_o. Therefore, the Gaussian prior density based on training data has mean and inverse variance equal to

$$\begin{aligned} \overline{\beta}_\rho &= \overline{\beta}_T, \\ \mathbf{B} &= f\mathbf{J}\big(\overline{\beta}_T\big), \end{aligned} \tag{29}$$

where $f < 1$. From equations (24) and (25), it is seen that the greater f is the greater the weighting of the channel estimate $\overline{\beta}_\rho$ derived from training data relative to that $\widetilde{\beta}^p$ derived from observations of the signal, in the computations of $\widehat{\beta}^p$. Thus, the choice of f determines this relative weighting. As discussed at the start of section 4.2 and proven in appendix B, the exact form of the prior density $\rho(\beta)$ will have little effect on the final results of the algorithm, as long as $\rho(\beta)$ has small enough tail probability to ensure convergence to the sequence that maximizes $f(\mathbf{y} \mid \mathbf{C})$, for example, by satisfying (B.7). We, therefore, suggest choosing f close to 1 to make $\rho(\beta)$ more peaked; the simulations of section 6 show good performance of the simplified EM algorithm for such values. Alternatively, the variance of $\overline{\beta}_T$, and hence $\mathbf{B}$, could be determined more precisely from repeated numerical experiments calculating $\overline{\beta}_T$ from the training data.

In the first ($p = 0$) iteration (26) indicates that (24) and (25) reduce to

$$\widehat{\beta}^0 = \overline{\beta}_\rho. \tag{30}$$

$$\widehat{\mathbf{A}}^0 = \mathbf{B}. \tag{31}$$

The first iteration of our simplified EM algorithm is analogous to the current equalization method used in GSM. First, the channel estimate $\overline{\beta}_T$ is derived from training data, and is used in the E step (16) to form $Q^{\mathrm{simp}}(\mathbf{C} \mid \widehat{\mathbf{C}}^p)$ with $\widehat{\beta}^0 = \overline{\beta}_T$. Next, the transmitted information sequence is estimated from this imperfect estimate of the channel using the M step (17).

This M step is equivalent by (12) and (16) to maximizing $f(\mathbf{y} \mid \overline{\beta}_T, \mathbf{C})$ with respect to $\mathbf{C}$. While in GSM the analogous estimate of the transmitted sequence from training data is the final estimate, our simplified EM algorithm uses this first sequence estimate to improve the estimate of the channel parameters through (20) and (24), which in turn is used to improve the sequence estimate through (16) and (17).

The sequence estimate based on even our first EM iteration is expected to be an improvement over that obtained in GSM due to our use of a discrete multipath channel model (3), rather than the finite impulse response used in GSM. Since estimation of a finite impulse response from training data requires a linear approximation to the nonlinearly modulated GMSK signal, this estimation incurs error. It was shown [5] that use of a multiray model to parameterize the channel, similar to our model, instead of a finite impulse response model, helps eliminate this linearization error and improves performance by 2 to 4 dB.

Computation of the matrix elements of $\mathbf{B}$ using (11) in (28) shows that they grow with the length of the training sequence and are inversely proportional to the power spectral density N_o of the noise. Hence, the length of the training sequence can in principle be selected so that for the lowest desired SNR the prior density (19) will with high probability be peaked enough that it satisfies the condition (A.4) for $p = 0$, which guarantees equivalence of the simplified and complete EM algorithms, or similarly that the probability p_{notEM} in (18) is small. Likewise, use of enough training data will ensure that (19) with high probability satisfies (B.7), which guarantees convergence of the simplified EM algorithm to the maximum of $f(\mathbf{y} \mid \mathbf{C})$. Reduction of the amount of transmitted training data is investigated in section 6.

We now consider equivalence of the simplified and complete EM algorithms for subsequent ($p > 0$) iterations. Equivalence of the algorithms depends on if evaluation of (13) can be replaced by evaluation of $-\lambda(\widehat{\beta}^p, \mathbf{y}, \mathbf{C})$ where $\widehat{\beta}^p$ maximizes $f(\beta \mid \mathbf{y}, \widehat{\mathbf{C}}^p)$. Equation (25) indicates that $f(\beta \mid \mathbf{y}, \widehat{\mathbf{C}}^p)$ for $p > 0$ is more sharply peaked than $f(\beta \mid \mathbf{y}, \widehat{\mathbf{C}}^0) = \rho(\beta)$. Thus, if

$$p_{\mathrm{notEM}} \leqslant p^*_{\mathrm{notEM}} \tag{32}$$

for some small desired probability p^*_{notEM} for the initial ($p = 0$) iteration, then it is expected that subsequent iterations ($p > 0$) will also satisfy (32), for large enough eigenvalues of $\mathbf{B}$ and $\widehat{\beta}^p$ not too far from $\widehat{\beta}^0$.

We have thus far assumed that the channel is uncorrelated between a user's successive time slots of data, so that the channel needs to be estimated independently in each time slot. When the channel fades more slowly, information from one time slot can be used to form an initial estimate of the channel in the next time slot, as is discussed in [19]. An intermediate scenario would take advantage of the fact that the time delays $\tau_i(t)$ and phase shifts $\theta_i(t)$ are slowly varying functions of time at vehicular speeds in most situations. Thus, only the parameters $\alpha_i(t)$ would need to be completely reestimated in every time slot, as is done in [5], while estimation of $\tau_i(t)$ and $\theta_i(t)$ could use information from preceding time slots for the initial estimate in the current time slot.

6. Simulations

Simulations of our simplified EM algorithm were performed with the GMSK modulation of the GSM system for various lengths of the training data sequence. Each subsequence over which the EM algorithm was run corresponds to a GSM time slot. The channel realization is chosen independently in each time slot, so that the simulation's time slots represent, for example, a user's consecutive time slots in a rapidly fading channel. Initially, a two path multipath model is used for the true channel, and the receiver also assumes the channel consists of two multipath components. Additional multipath models are considered at the end of this section. Each multipath component undergoes independent Rayleigh fading. It is assumed that the time delay of the first path is known, so that we set that delay to zero and consider the relative time delay between the two paths. The time delay between the paths was randomly picked from a uniform distribution with a range from 0.9 μs to 7.2 μs. Although this probability density of the true channel parameters differs from the Gaussian prior density assumed by the receiver, as described in section 4.2, the simplified EM algorithm performs quite well, as expected.

Equation (29) is used to form a Gaussian prior density (19) from training data, and (24), (25) are used in the E step in each iteration. The VA is used to find the optimal sequence $\widehat{\mathbf{C}}^{p+1}$ through (17) in each M step of the simplified EM algorithm. No channel coding was used, and the transmitted data is an independently identically distributed binary sequence with -1 and $+1$ equally likely.

The probability of bit error is plotted as a function of SNR in figures 1 and 2 when 14 and 26 bits, respectively, of training data per time slot are used by the receiver. The SNR in each curve in each figure refers to the actual signal to noise ratio at the receiver before processing. The BER for each curve in each figure is based on 116 information bits that are estimated in each time slot. The simulations corresponding to both figures 1 and 2 use the same realization of the random channel model.

The upper (dashed) curve in each figure displays the BER when sequence estimation is performed by the "training data method". In this case, training data alone are used to estimate the channel, followed by estimation of the information

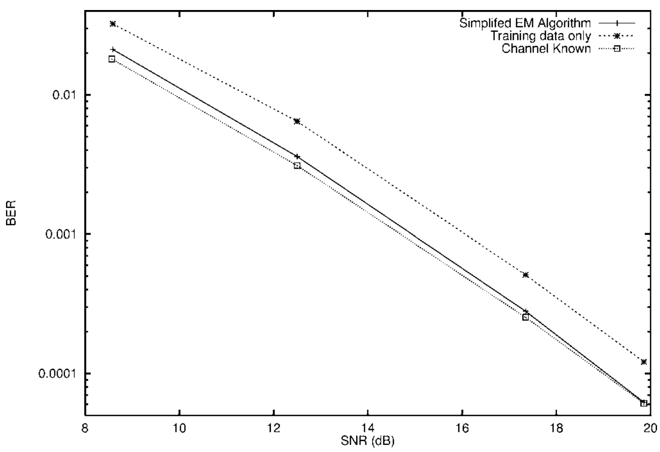

Figure 1. BER versus SNR when only 14 bits of training data per time slot are used.

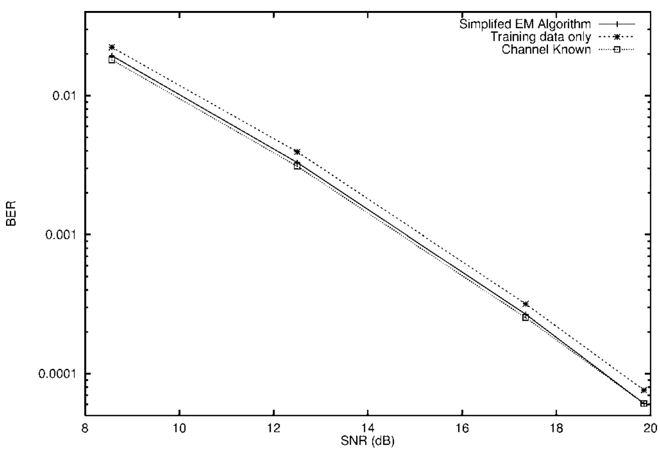

Figure 2. BER versus SNR when 26 bits of training data per time slot are used.

sequence, given this channel estimate, through one pass of the VA. The sequence estimate derived from the training data method is, thus, analogous to the method used in [5] (which has computational complexity much less than that of the finite impulse response model currently used in GSM). The BER from the current GSM system, if it did not use channel coding, would thus be higher than that of this upper curve due to the parametric multipath channel model, as explained in section 5.

The middle (solid) curve in each figure displays the BER when the simplified EM algorithm is used. The difference between the upper curve and the solid middle curve shows the improvement obtained by the simplified EM algorithm relative to the training data method. Since the training data provide a better channel estimate when 26 bits are used instead of 14, the upper (dashed) curve based on the training data method shows better performance in figure 2 than in figure 1. The EM algorithm decreased the BER by 35% to 49% relative to that of the training data method when 14 bits of training data are used, with the largest decreases at the highest SNR.

The lower (dotted) curve in each figure displays a lower bound on the lowest possible BER that can be achieved for the simulated channel: it is the BER when the channel is known exactly by the receiver, and is thus identical in figures 1 and 2. The EM algorithm essentially removes the bit errors of the training data method due to uncertainty in the channel, as seen by the fact that in both figures the EM algorithm's BER is almost as low as the BER when the channel is exactly known. At high SNR the performance of the simplified EM algorithm approaches that of a receiver to which the channel is exactly known.

By comparing the differences in each figure of the solid middle curve to the common lower bound of an exactly known channel, we see that the simplified EM algorithm performs nearly as well when 14 bits of training data per time slot are used as when 26 bits of training data per time slot are used. Furthermore, a comparison of the middle curve in figure 1 to the upper curve in figure 2 shows that use of the simplified EM algorithm with 14 training bits per time slot outperforms the training data method with use of 26 bits of training data per time slot. Since the SNR in both figures is based on 26 bits of data being transmitted during the training period, if only 14 bits of training data were transmitted, further power savings could be achieved.

Further simulations showed that as the number of training bits per time slot is decreased below 14, the BERs of both the training data method and the simplified EM algorithm increase substantially. We note that all of the above results were obtained from transmission of one of the specified 26 bit training data sequences used in GSM. These training sequences were selected for their optimal correlation properties to be used with a matched filter in estimating the channel impulse response. However, since we use a discrete multipath channel model, and directly estimate the channel through (15), other training sequences optimized for this method may find even fewer training bits needed by the receiver than reported here.

The simulation results discussed thus far have been based on a two-path channel model, as well as a receiver that assumes a two path channel. We now consider cases in which the channel can have multiple multipath reflections and the receiver may assume a different number of multipath reflections than that found in the actual channel. In figure 3, we display results of the case in which the receiver assumes a two path model, whereas the actual channel has three multipath components. The extra path of the actual channel acts as an extra unmodeled noise source, as seen from (5). In this case, the EM algorithm again offers improvement over using training data alone; however, the performance does not closely approach that of the case in which the complete three-path channel is known by the receiver.

When the receiver assumes $M = 2$ paths, as it does in figures 1–3, most of the decrease in the BER in the EM algorithm takes place in the $p = 1$ iteration. While there is some additional decrease in the next ($p = 2$) iteration, little improvement is seen in subsequent iterations. Therefore, a practical receiver could require only one or two iterations following the initial sequence estimate ($p = 0$) based on training data alone. In figures 1–3 the BER plotted for the simplified EM algorithm is taken after the $p = 2$ iteration. In figure 3, the BER is also plotted after the $p = 1$ iteration; this curve is seen to lie very slightly above the simplified EM algorithm curve, which uses two iterations. The simplified EM algorithm's computation time is dominated by the VA rather than the channel updating. Hence, a simplified EM algorithm which is terminated after the $p = 1$ iteration should take approximately twice as much computation time as the training data method.

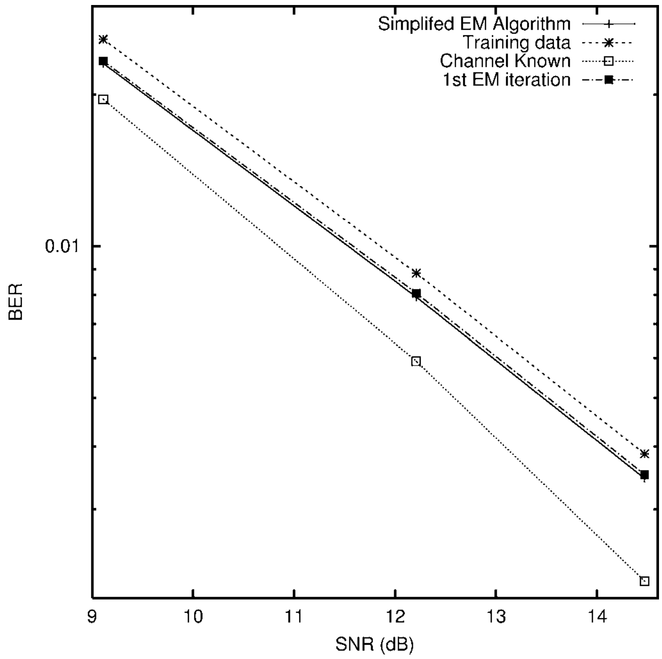

Figure 3. BER versus SNR when 26 bits of training data per time slot are used, and the channel has 3 multipath components, whereas the receiver assumes a 2 path channel.

We now consider the event in which the true channel has M paths, and the receiver also assumes an M path channel. Performance of the training data method, when it is used with a fixed amount of training data, will worsen with increasing M since as the number of channel parameters to be estimated increases, the error in estimating them will increase. Similarly, since the initial iteration of the simplified EM algorithm is equivalent to the training data method, for large M more training data will be needed to provide an accurate enough initial estimate of the channel for the simplified EM algorithm. If enough training data are used so that the channel prior density satisfies (B.7), the simplified EM algorithm will converge to the sequence that maximizes $f(\mathbf{y} \mid \mathbf{C})$, as discussed in section 4.1, which was based on general M, and at convergence, should offer significant improvement over using training data alone. However, as M increases the number of iterations until convergence may increase. In conventional uses of the EM algorithm the number of iterations until convergence is known to increase with the amount of "missing data" [9]; the analogous missing data in our model are the channel parameters. The convergence rate of the simplified EM algorithm for larger values of M remains a question for further investigation.

If the receiver uses a channel model which has more paths than the actual channel does, initial estimation of the channel from training data will result in overfitting and a higher BER, because the channel will be fit to random noise fluctuations. In addition, at this initial iteration, as well as at E steps in subsequent iterations, the receiver's use of extra multipath components could render the channel parameters unidentifiable, in which case the optimizer may be unable to maximize (15). When this problem occurs, the number of paths used by the receiver should be reduced. Furthermore, the receiver should begin by assuming a minimal number of multipath components. Assuming the receiver can tabulate its frame or bit error rates, the receiver could vary the number of paths it assumes, if the current model is giving too high a frame error rate.

7. Summary

Application of the EM algorithm to find the MLSE of CPM signals in a fading multipath channel cannot be directly implemented. We derive a simplified version of the EM algorithm to enable implementation. We show that for suitable initial conditions, with high probability, the simplified EM algorithm is equivalent to the complete EM algorithm, and it maximizes the desired likelihood function $f(\mathbf{y} \mid \mathbf{C})$. Simulations for GSM show that the simplified EM algorithm produces a BER almost as low as that of an ideal receiver that has perfect knowledge of the channel for the model considered, and reduces the required training data from 26 to 14 bits per time slot. The simplified EM algorithm could also be applied to other CPM systems that are subject to a fading multipath channel; two examples of such systems are GPRS (General Packet Radio Service), an evolution of GSM for data transmission, and HIPERLAN (High Performance Radio Local Area Network), a European standard which uses GMSK for data transmission at rates of 24 Mbit/s [1].

Appendix A. Conditions for equivalence of the simplified and complete EM algorithms

We first describe stringent conditions under which the complete (13), (14) and simplified (16), (17) EM algorithms are equivalent. We then show a less restrictive criterion for this equivalence.

In the maximization of $Q(\mathbf{C} \mid \widehat{\mathbf{C}}^p)$ in the M step, we consider the result of minimizing $\lambda(\beta, \mathbf{y}, \mathbf{C})$ with respect to $\mathbf{C}$ for various values of β, and note that a range of β will produce the same sequence that minimizes $\lambda(\beta, \mathbf{y}, \mathbf{C})$. We define S^p to be the set of all β such that the same sequence minimizes $\lambda(\beta, \mathbf{y}, \mathbf{C})$ as minimizes $\lambda(\widehat{\beta}^p, \mathbf{y}, \mathbf{C})$:

$$S^p = \Bigl\{\beta\colon \arg\min_{\mathbf{C}} \lambda(\beta, \mathbf{y}, \mathbf{C}) = \arg\min_{\mathbf{C}} \lambda\bigl(\widehat{\beta}^p, \mathbf{y}, \mathbf{C}\bigr)\Bigr\}. \quad \text{(A.1)}$$

If

$$\operatorname{supp} \rho \subset S^p, \quad \text{(A.2)}$$

where $\operatorname{supp} \rho$ denotes the support of the prior channel density $\rho(\beta)$, then only $\lambda(\widehat{\beta}^p, \mathbf{y}, \mathbf{C})$ needs to be minimized and the integral in (13) need not be evaluated, for the purpose of finding the sequence $\widehat{\mathbf{C}}^{p+1}$ that maximizes $Q(\mathbf{C} \mid \widehat{\mathbf{C}}^p)$. In this case

$$\begin{aligned} &\arg\max_{\mathbf{C}}\Bigl[-\int \lambda(\beta, \mathbf{y}, \mathbf{C}) f\bigl(\beta \bigm| \mathbf{y}, \widehat{\mathbf{C}}^p\bigr)\,\mathrm{d}\beta\Bigr] \\ &\quad = \arg\max_{\mathbf{C}}\bigl[-\lambda\bigl(\widehat{\beta}^p, \mathbf{y}, \mathbf{C}\bigr)\bigr], \end{aligned} \quad \text{(A.3)}$$

and the simplified (16), (17) and complete (13), (14) EM algorithms are equivalent.

As shown below, this equivalence can also be met by satisfaction of the less restrictive condition:

$$\frac{F^p(S^p)}{1 - F^p(S^p)} \geqslant k, \quad \text{(A.4)}$$

where $F^p(S^p)$ denotes the probability at iteration $p+1$ of $\beta \in S^p$, given $\mathbf{y}$ and $\widehat{\mathbf{C}}^p$:

$$F^p\bigl(S^p\bigr) = \int_{S^p} f\bigl(\beta \bigm| \mathbf{y}, \widehat{\mathbf{C}}^p\bigr)\,\mathrm{d}\beta, \quad \text{(A.5)}$$

and $k > 0$ is finite and is defined below. Use of sufficient training data for the lowest desired SNR will allow (A.4) to be met with high probability, as described in section 5. Use of this criterion in practice requires evaluating the multidimensional integral (A.5). The limits of this integral, as defined by (A.1), are determined by minimizing $\lambda(\beta, \mathbf{y}, \mathbf{C})$ with respect to $\mathbf{C}$ for each possible value of the multidimensional parameter β; each such minimization requires the VA to be run, making evaluation of (A.5) computationally onerous.

We first let $Q_R(\mathbf{C} \mid \widehat{\mathbf{C}}^p)$ denote the contribution to $Q(\mathbf{C} \mid \widehat{\mathbf{C}}^p)$ from values of β in the region R:

$$Q_R\bigl(\mathbf{C} \bigm| \widehat{\mathbf{C}}^p\bigr) = -\int_R \lambda(\beta, \mathbf{y}, \mathbf{C}) f\bigl(\beta \bigm| \mathbf{y}, \widehat{\mathbf{C}}^p\bigr)\,\mathrm{d}\beta. \quad \text{(A.6)}$$

We consider the possibility that a sequence estimate $\acute{\mathbf{C}}^{p+1} \neq \widehat{\mathbf{C}}^{p+1}$, where $\widehat{\mathbf{C}}^{p+1}$ is defined by (17), maximizes $Q(\mathbf{C} \mid \widehat{\mathbf{C}}^p)$ at iteration $p+1$, and we define

$$\begin{aligned} &\Delta Q_R\bigl(\acute{\mathbf{C}}^{p+1}, \widehat{\mathbf{C}}^{p+1} \bigm| \widehat{\mathbf{C}}^p\bigr) \\ &\quad = Q_R\bigl(\acute{\mathbf{C}}^{p+1} \bigm| \widehat{\mathbf{C}}^p\bigr) - Q_R\bigl(\widehat{\mathbf{C}}^{p+1} \bigm| \widehat{\mathbf{C}}^p\bigr) \\ &\quad = -\int_R \Delta\lambda\bigl(\beta, \mathbf{y}, \acute{\mathbf{C}}^{p+1}, \widehat{\mathbf{C}}^{p+1}\bigr) f\bigl(\beta \bigm| \mathbf{y}, \widehat{\mathbf{C}}^p\bigr)\,\mathrm{d}\beta, \end{aligned} \quad \text{(A.7)}$$

where we have defined

$$\Delta\lambda\bigl(\beta, \mathbf{y}, \acute{\mathbf{C}}^{p+1}, \widehat{\mathbf{C}}^{p+1}\bigr) = \lambda\bigl(\beta, \mathbf{y}, \acute{\mathbf{C}}^{p+1}\bigr) - \lambda\bigl(\beta, \mathbf{y}, \widehat{\mathbf{C}}^{p+1}\bigr). \quad \text{(A.8)}$$

If

$$\begin{aligned} &\Delta Q_{\operatorname{supp}\rho}\bigl(\acute{\mathbf{C}}^{p+1}, \widehat{\mathbf{C}}^{p+1} \bigm| \widehat{\mathbf{C}}^p\bigr) \\ &\quad = \Delta Q_{S^p}\bigl(\acute{\mathbf{C}}^{p+1}, \widehat{\mathbf{C}}^{p+1} \bigm| \widehat{\mathbf{C}}^p\bigr) \\ &\qquad + \Delta Q_{\operatorname{supp}\rho \setminus S^p}\bigl(\acute{\mathbf{C}}^{p+1}, \widehat{\mathbf{C}}^{p+1} \bigm| \widehat{\mathbf{C}}^p\bigr) \\ &\quad < 0, \end{aligned} \quad \text{(A.9)}$$

then the sequence $\acute{\mathbf{C}}^{p+1}$ cannot maximize $Q(\mathbf{C} \mid \widehat{\mathbf{C}}^p)$. The sequence estimate $\widehat{\mathbf{C}}^{p+1}$ maximizes $Q_{S^p}(\mathbf{C} \mid \widehat{\mathbf{C}}^p)$ by the definitions (A.1) and (17), and thus, $\Delta Q_{S^p}(\acute{\mathbf{C}}^{p+1}, \widehat{\mathbf{C}}^{p+1} \mid \widehat{\mathbf{C}}^p) < 0$. Consequently, if $\acute{\mathbf{C}}^{p+1}$ were to maximize $Q(\mathbf{C} \mid \widehat{\mathbf{C}}^p)$, then $\Delta Q_{\mathrm{supp}\,\rho\setminus S^p}(\acute{\mathbf{C}}^{p+1}, \widehat{\mathbf{C}}^{p+1} \mid \widehat{\mathbf{C}}^p) > 0$ must hold, and thus,

$$\max_{\beta_1 \in \mathrm{supp}\,\rho\setminus S^p}\left(-\Delta\lambda\left(\beta_1, \mathbf{y}, \acute{\mathbf{C}}^{p+1}, \widehat{\mathbf{C}}^{p+1}\right)\right) > 0. \tag{A.10}$$

We let

$$\beta_{\min} = \arg\min_{\beta\in S^p} \Delta\lambda\left(\beta, \mathbf{y}, \acute{\mathbf{C}}^{p+1}, \widehat{\mathbf{C}}^{p+1}\right). \tag{A.11}$$

Since $\beta_{\min} \in S^p$, the definition (A.1) implies that $\Delta\lambda(\beta_{\min}, \mathbf{y}, \acute{\mathbf{C}}^{p+1}, \widehat{\mathbf{C}}^{p+1}) > 0$, if the sequence that minimizes $\lambda(\beta_{\min}, \mathbf{y}, \mathbf{C})$ is unique. If this sequence is not unique and $\Delta\lambda(\beta_{\min}, \mathbf{y}, \acute{\mathbf{C}}^{p+1}, \widehat{\mathbf{C}}^{p+1}) = 0$, then (A.1) implies that the M step (17) of the simplified algorithm will also yield $\acute{\mathbf{C}}^{p+1}$. Thus, hereafter we need only consider $\Delta\lambda(\beta_{\min}, \mathbf{y}, \acute{\mathbf{C}}^{p+1}, \widehat{\mathbf{C}}^{p+1}) > 0$. We thus define a finite number k', which is positive by (A.10),

$$k' = \frac{\max_{\beta_1\in \mathrm{supp}\,\rho\setminus S^p}(-\Delta\lambda(\beta_1, \mathbf{y}, \acute{\mathbf{C}}^{p+1}, \widehat{\mathbf{C}}^{p+1}))}{\Delta\lambda(\beta_{\min}, \mathbf{y}, \acute{\mathbf{C}}^{p+1}, \widehat{\mathbf{C}}^{p+1})}. \tag{A.12}$$

Thus, k' depends on $\acute{\mathbf{C}}^{p+1}$, as well as indirectly on supp ρ and S^p. Then we have $\forall\beta_1 \in \mathrm{supp}\,\rho \setminus S^p$

$$-\Delta\lambda\left(\beta_1, \mathbf{y}, \acute{\mathbf{C}}^{p+1}, \widehat{\mathbf{C}}^{p+1}\right) \leqslant k'\Delta\lambda\left(\beta_{\min}, \mathbf{y}, \acute{\mathbf{C}}^{p+1}, \widehat{\mathbf{C}}^{p+1}\right). \tag{A.13}$$

We use (A.7) and (A.11) to show that

$$\begin{aligned}&\Delta Q_{S^p}\left(\acute{\mathbf{C}}^{p+1}, \widehat{\mathbf{C}}^{p+1} \,\middle|\, \widehat{\mathbf{C}}^p\right)\\ &\quad\leqslant -\Delta\lambda\left(\beta_{\min}, \mathbf{y}, \acute{\mathbf{C}}^{p+1}, \widehat{\mathbf{C}}^{p+1}\right)F^p\left(S^p\right),\end{aligned} \tag{A.14}$$

where $F^p(S^p)$ is given by (A.5), and we also use (A.13) to yield

$$\begin{aligned}&\Delta Q_{\mathrm{supp}\,\rho\setminus S^p}\left(\acute{\mathbf{C}}^{p+1}, \widehat{\mathbf{C}}^{p+1} \,\middle|\, \widehat{\mathbf{C}}^p\right)\\ &\quad\leqslant k'\Delta\lambda\left(\beta_{\min}, \mathbf{y}, \acute{\mathbf{C}}^{p+1}, \widehat{\mathbf{C}}^{p+1}\right)\left[1 - F^p\left(S^p\right)\right].\end{aligned} \tag{A.15}$$

Summation of (A.14) and (A.15) shows that the inequality in (A.9) is satisfied if

$$\frac{F^p(S^p)}{1 - F^p(S^p)} \geqslant k'. \tag{A.16}$$

Hence, if (A.16) is satisfied for all possible sequences $\acute{\mathbf{C}}^{p+1} \neq \widehat{\mathbf{C}}^{p+1}$, then $\widehat{\mathbf{C}}^{p+1}$ as defined by (17) maximizes $Q(\mathbf{C} \mid \widehat{\mathbf{C}}^p)$. Thus, we obtain (A.4) by defining

$$k = \max_{\acute{\mathbf{C}}^{p+1}} k'. \tag{A.17}$$

Appendix B. Convergence of the simplified EM algorithm to the maximum likelihood estimate

We now prove that the simplified EM algorithm converges to a sequence $\mathbf{C}^P$, by the Pth iteration, which maximizes the likelihood function $f(\mathbf{y} \mid \mathbf{C})$. Using Bayes Theorem, we express this likelihood function as

$$L(\mathbf{C} \mid \mathbf{y}) = f(\mathbf{y} \mid \mathbf{C}) = \int f(\mathbf{y} \mid \beta, \mathbf{C})\rho(\beta)\,\mathrm{d}\beta. \tag{B.1}$$

Convergence to a sequence $\mathbf{C}^P$ is guaranteed since $L(\mathbf{C} \mid \mathbf{y})$ increases on each EM iteration, when (A.16) is satisfied, by the same arguments as in [9], and since $L(\mathbf{C} \mid \mathbf{y})$ is clearly bounded.

In order to show $\mathbf{C}^P$ maximizes (B.1), we note that by (12), (16), and (17) $\mathbf{C}^P = \arg\max_{\mathbf{C}} f(\mathbf{y} \mid \widehat{\beta}^P, \mathbf{C})$, and hence, that by (A.1)

$$\mathbf{C}^P = \arg\max_{\mathbf{C}} f(\mathbf{y} \mid \beta, \mathbf{C}), \quad \beta \in S^P. \tag{B.2}$$

We consider the possibility that another sequence $\acute{\mathbf{C}} \neq \mathbf{C}^P$ maximizes $L(\mathbf{C} \mid \mathbf{y})$, and note that

$$\begin{aligned}L&\left(\mathbf{C}^P \,\middle|\, \mathbf{y}\right) - L\left(\acute{\mathbf{C}} \,\middle|\, \mathbf{y}\right)\\ &= \int_{S^P} \Delta f\left(\mathbf{y} \,\middle|\, \beta, \mathbf{C}^P, \acute{\mathbf{C}}\right)\rho(\beta)\,\mathrm{d}\beta\\ &\quad+ \int_{\mathrm{supp}\,\rho\setminus S^P} \Delta f\left(\mathbf{y} \,\middle|\, \beta, \mathbf{C}^P, \acute{\mathbf{C}}\right)\rho(\beta)\,\mathrm{d}\beta,\end{aligned} \tag{B.3}$$

where

$$\Delta f(\mathbf{y} \mid \beta, \mathbf{C}_1, \mathbf{C}_2) = f(\mathbf{y} \mid \beta, \mathbf{C}_1) - f(\mathbf{y} \mid \beta, \mathbf{C}_2). \tag{B.4}$$

Hence, $\mathbf{C}^P$ maximizes $L(\mathbf{C} \mid \mathbf{y})$ if for every sequence $\acute{\mathbf{C}} \neq \mathbf{C}^P$ equation (B.3) is positive. The integrand in the first term in (B.3) is always positive by (B.2), while the second term in (B.3) must be negative if $\acute{\mathbf{C}}$ is to maximize $L(\mathbf{C} \mid \mathbf{y})$. Therefore, $\max_{\beta\in\mathrm{supp}\,\rho\setminus S^P}(-\Delta f(\mathbf{y} \mid \beta, \mathbf{C}^P, \acute{\mathbf{C}}))$ must be positive.

We define the probability, based on the priory density, that $\beta \in S^P$ as

$$F_\rho\left(S^P\right) = \int_{S^P} \rho(\beta)\,\mathrm{d}\beta, \tag{B.5}$$

and also define

$$\kappa' = \frac{\max_{\beta\in\mathrm{supp}\,\rho\setminus S^P}(-\Delta f(\mathbf{y} \mid \beta, \mathbf{C}^P, \acute{\mathbf{C}}))}{\min_{\beta\in S^P} \Delta f(\mathbf{y} \mid \beta, \mathbf{C}^P, \acute{\mathbf{C}})}. \tag{B.6}$$

Then in analogy to the derivation in appendix A of (A.4), it can be seen that a sufficient condition for $\mathbf{C}^P$ to maximize $L(\mathbf{C} \mid \mathbf{y})$ is

$$\frac{F_\rho(S^P)}{1 - F_\rho(S^P)} \geqslant \kappa, \tag{B.7}$$

where

$$\kappa = \max_{\acute{\mathbf{C}}} \kappa'. \tag{B.8}$$

The above proof assumed $\widehat{\beta}^p$ as defined by (15) is unique, which usually occurs in practice. However, even when $\widehat{\beta}^p$ is not uniquely defined by (15), the proof given here, as well as that in appendix A, still holds, as long as $\widehat{\beta}^p$ is redefined. The simplified EM algorithm relies on use of a single value of β, denoted by $\widehat{\beta}^p$ at iteration p, which is used in the E step (16) to replace the integral (13). There are other possible definitions for $\widehat{\beta}^p$ in addition to $\arg\max_\beta f(\beta \mid \mathbf{y}, \widehat{\mathbf{C}}^p)$, which was used in (15). The sufficient conditions (A.4) and (B.7) depend on the definition of $\widehat{\beta}^p$ through the corresponding region S^p defined in (A.1).

We now present definitions that can be used for $\widehat{\beta}^p$ in a few special cases in which (15) does not exist. If $f(\beta \mid \mathbf{y}, \widehat{\mathbf{C}}^p)$ is a uniform density, the point $\widehat{\beta}^p$ could be defined as the mean of $f(\beta \mid \mathbf{y}, \widehat{\mathbf{C}}^p)$, a definition which could also be used even if (15) exists. As a second example, if at iteration p there are multiple values of β that maximize $f(\beta \mid \mathbf{y}, \widehat{\mathbf{C}}^p)$, one value can be chosen at random to be $\widehat{\beta}^p$; the associated region S^p would be used throughout the proof here and in appendix A. If another value of β that also maximizes $f(\beta \mid \mathbf{y}, \widehat{\mathbf{C}}^p)$, which we denote by $\widehat{\beta}^{p,2}$, satisfies $\widehat{\beta}^{p,2} \in S^p$, then it will yield the same final sequence $\mathbf{C}^P$ as $\widehat{\beta}^p$ does. If instead $\widehat{\beta}^{p,2} \in \operatorname{supp}\rho \setminus S^p$, then $\widehat{\beta}^{p,2}$ is treated as any other point in this latter region in the proofs.

Acknowledgements

The authors would like to thank the reviewers for their insightful comments, which were incorporated into the paper. This work has been supported, in part, by the National Science Foundation, the NJ Commission on Science and Technology, and the Ogasawara Foundation for the Promotion of Science and Engineering.

References

[1] N. Benvenuto, P. Bisaglia, A. Salloum and L. Tomba, Worst case equalizer for noncoherent HIPERLAN receivers, IEEE Transactions on Communications 48 (2000) 28–36.

[2] G. Beveridge and R. Schechter, *Optimization: Theory and Practice* (1970).

[3] D. Boss, K. Kammeyer and T. Petermann, Is blind channel estimation feasible in mobile communication systems? A study based on GSM, IEEE Journal on Selected Aareas in Communications 16(8) (1998) 1479–1492.

[4] C. Brutel, J. Boutros and P. Mege, Iterative joint channel estimation and detection of coded CPM, in: *Broadband Communications Proceedings 2000* (2000) pp. 287–292.

[5] J. Chen, A. Paulraj and U. Reddy, Multichannel maximum-likelihood sequence estimation equalizer for GSM using a parametric channel model, IEEE Transactions on Communications 47 (January 1999) 53–63.

[6] J. Cheung and R. Steele, Soft-decision feedback equalizer for continuous phase modulated signals in wideband mobile radio channels, IEEE Transactions on Communications 42 (1994) 1628–1638.

[7] D.R. Cox and D.V. Hinkley, *Theoretical Statistics* (New York, 1974).

[8] D. Dahlhaus, B. Fleury and A. Radović, A sequential algorithm for joint parameter estimation and multiuser detection in DS/CDMA systems with multipath propagation, Wireless Personal Communications 6 (1998) 161–178.

[9] A.P. Dempster, N.M. Laird and D.B. Rubin, Maximum likelihood from incomplete data via the EM algorithm, Journal of the Royal Statistical Society Series B 39(1) (1977) 1–38.

[10] M. Feder and J. Catipovic, Algorithms for joint channel estimation and data recovery – Application to equalization in underwater communications, IEEE Journal of Oceanic Engineering 16 (1991) 42–55.

[11] C.N. Georghiades and J.C. Han, Sequence estimation in the presence of random parameters via the EM algorithm, IEEE Transactions on Communications 45 (March 1997) 300–308.

[12] C.N. Georghiades and D.L. Snyder, The expectation–maximization algorithm for symbol unsynchronized sequence detection, IEEE Transactions on Communications 39 (January 1991) 54–61.

[13] H. Kobayashi, Simultaneous adaptive estimation and decision algorithm for carrier modulated data transmission systems, IEEE Transactions on Communication Technologies (June 1971) 268–280.

[14] J.W. Modestino, Reduced-complexity iterative maximum-likelihood sequence estimation on channels with memory, in: *Proceedings of the International Symposium on Information Theory*, San Antonio, TX (January 1993).

[15] T.K. Moon, The expectation–maximization algorithm, IEEE Signal Processing Magazine (November 1996) 47–60.

[16] L.B. Nelson and H.V. Poor, Iterative multiuser receivers for CDMA channels: An EM-based approach, IEEE Transactions on Communications 44 (December 1996) 1700–1710.

[17] Z. Yang and X. Wang, Turbo equalization for GMSK signaling over multipath channels based on the Gibbs sampler, IEEE Journal on Selected Areas in Communications 19 (2001) 1753–1763.

[18] S. Zabin and H.V. Poor, Efficient estimation of class a noise parameters via the EM algorithm, IEEE Transactions on Information Theory 37 (January 1991) 60–72.

[19] L.M. Zeger and H. Kobayashi, MLSE for CPM signals in a fading multipath channel, in: *Proceedings of IEEE Pacific Rim Conference on Communications, Computers, and Signal Processing*, Victoria, Canada (1999) pp. 511–515.

Linda M. Zeger received the B.A. degree in physics from Princeton University in 1989, and the M.A. and Ph.D. degrees in physics from Harvard University in 1991 and 1994, respectively. She worked at Educational Testing Service, Princeton, NJ, performing research in statistical estimation from 1994 to 1996. Since 1996, she has been with Lucent Technologies, and she is a senior member of IEEE. Her interests are in wireless communications, including CDMA cellular systems, scheduling of packet data transmission, and detection and estimation.
E-mail: lzeger@lucent.com

Hisashi Kobayashi is the Sherman Fairchild University Professor of Electrical Engineering and Computer Science at Princeton University, where he served as the Dean of Engineering and Applied Science (1986–1991). From 1967 to 1982 he was with the IBM Research Center in Yorktown Heights, and held many managerial positions, including the founding Director of the IBM Tokyo Research Laboratory (1982–1986). His current research activities include performance modeling/analysis of communication networks, wireless geolocation algorithms, network reliability/security, and teletraffic/queueing theory. He is a Fellow of IEEE (1977), a recipient of the Humboldt Prize, Germany (1979), the IFIPS Silver Core Award (1981). He is a member of Engineering Academy of Japan (1992). He received Ph.D. from Princeton (1967), BE and ME from University of Tokyo (1961 and 1963). He was a radar designer at Toshiba Corp., Japan (1963–1965).
E-mail: hisashi@ee.princeton.edu